RSIG
(Remote Sensing Information Gateway)

RSIG is a free software system for quickly and easily combining, visualizing and saving subsets of selected remote geospatial environmental data on the Web.

Its primary audience is environmental modelers, interested in historic and contemporary air–quality analysis, trends and events.

RSIG was developed by Lockheed Martin for the US EPA in RTP, NC.

http://badger.epa.gov/rsig/

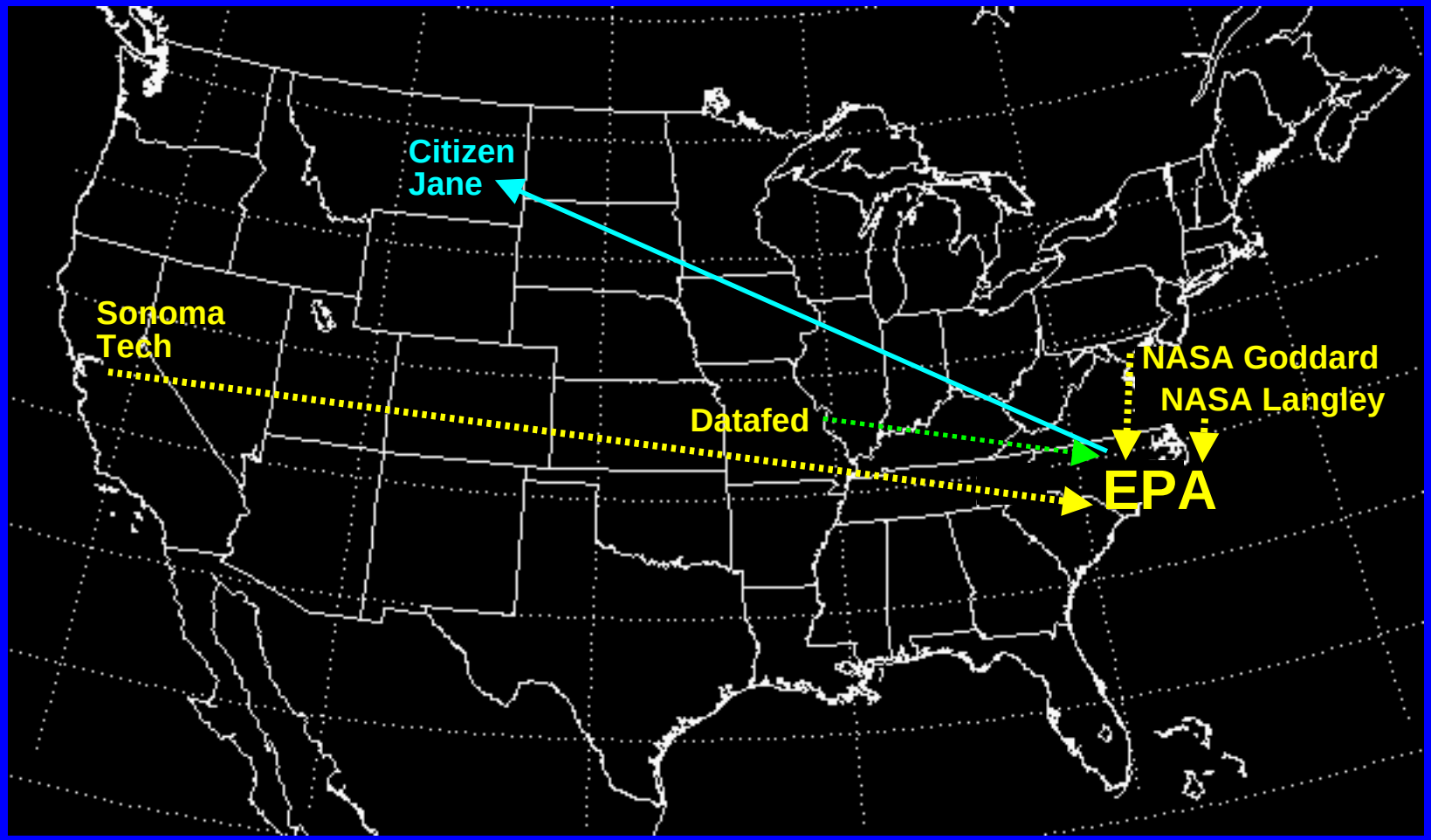
plessel.todd@epa.gov 1–919–541–5500
freeman.matt@epa.gov 1–919–541–4293

RSIG Unique/Key Features:

1. Single website virtually serving many kinds of data
2. Subsetting
3. Aggregation
4. Visualization
5. Save subset to a local file
6. Fast!

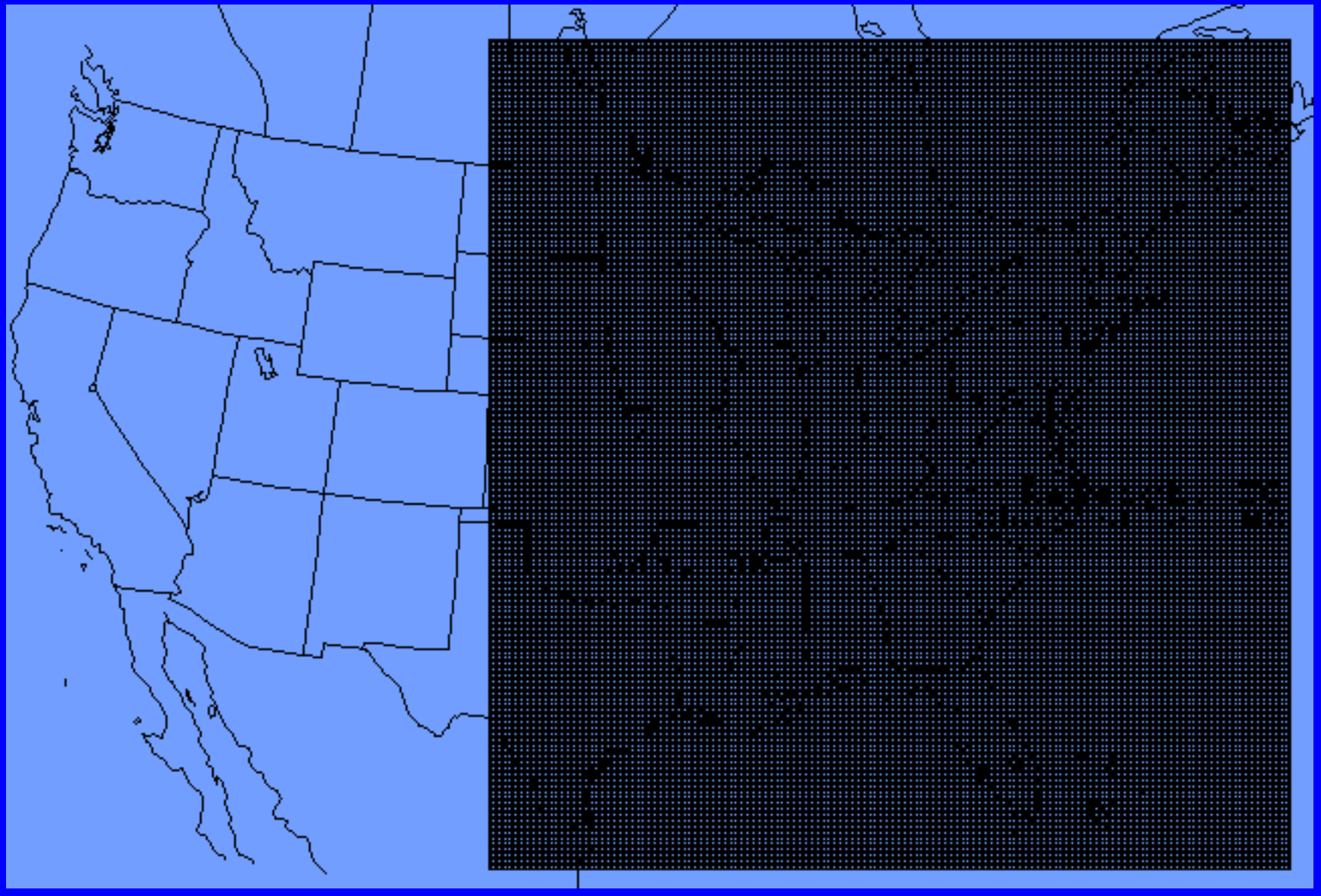
RSIG Unique/Key Features:

1. Single website virtually serving many kinds of data



RSIG Data

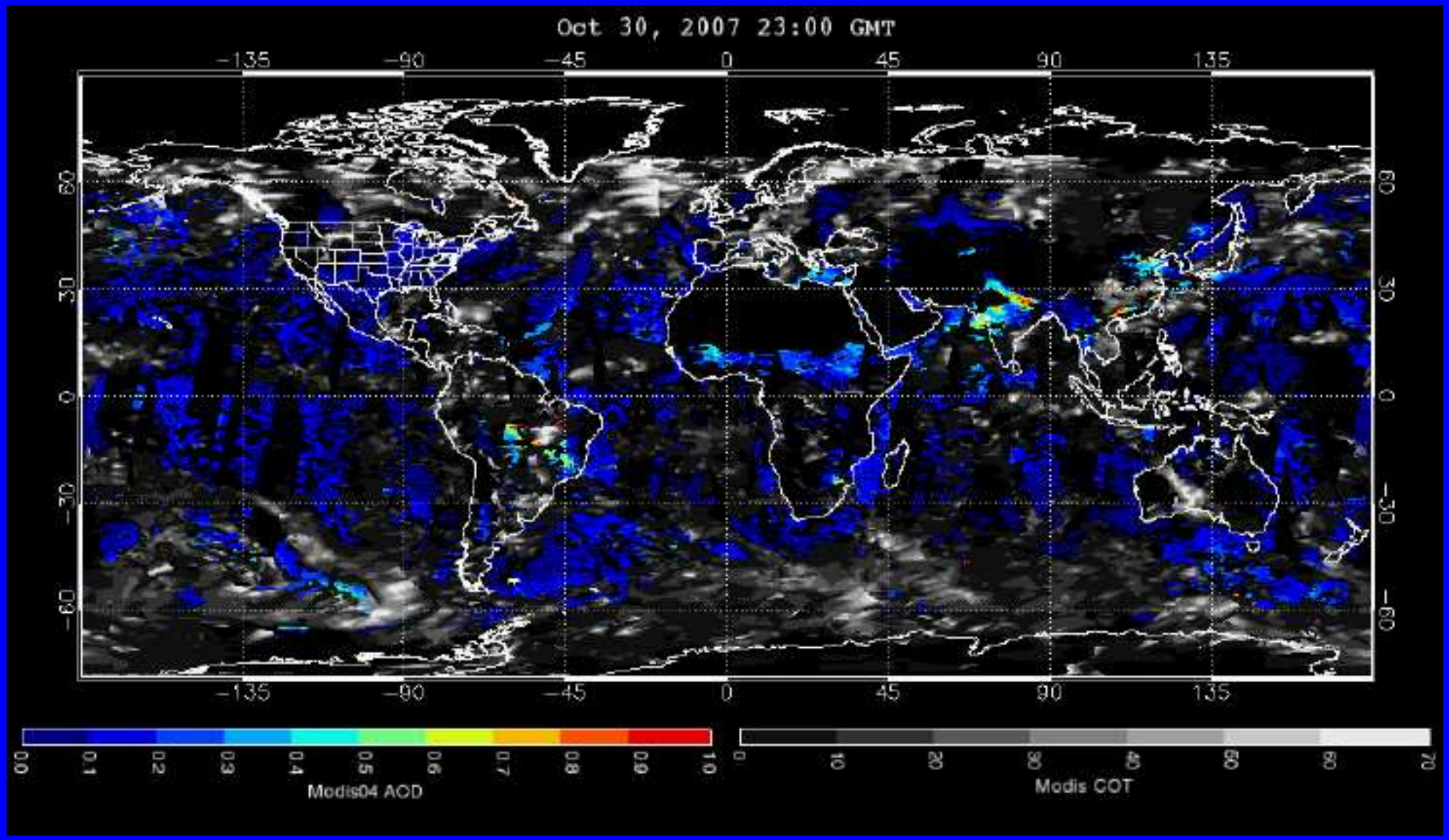
CMAQ model 12km 3D gridded air–quality and meteorology (Ozone, PM 2.5, Temp, Wind)



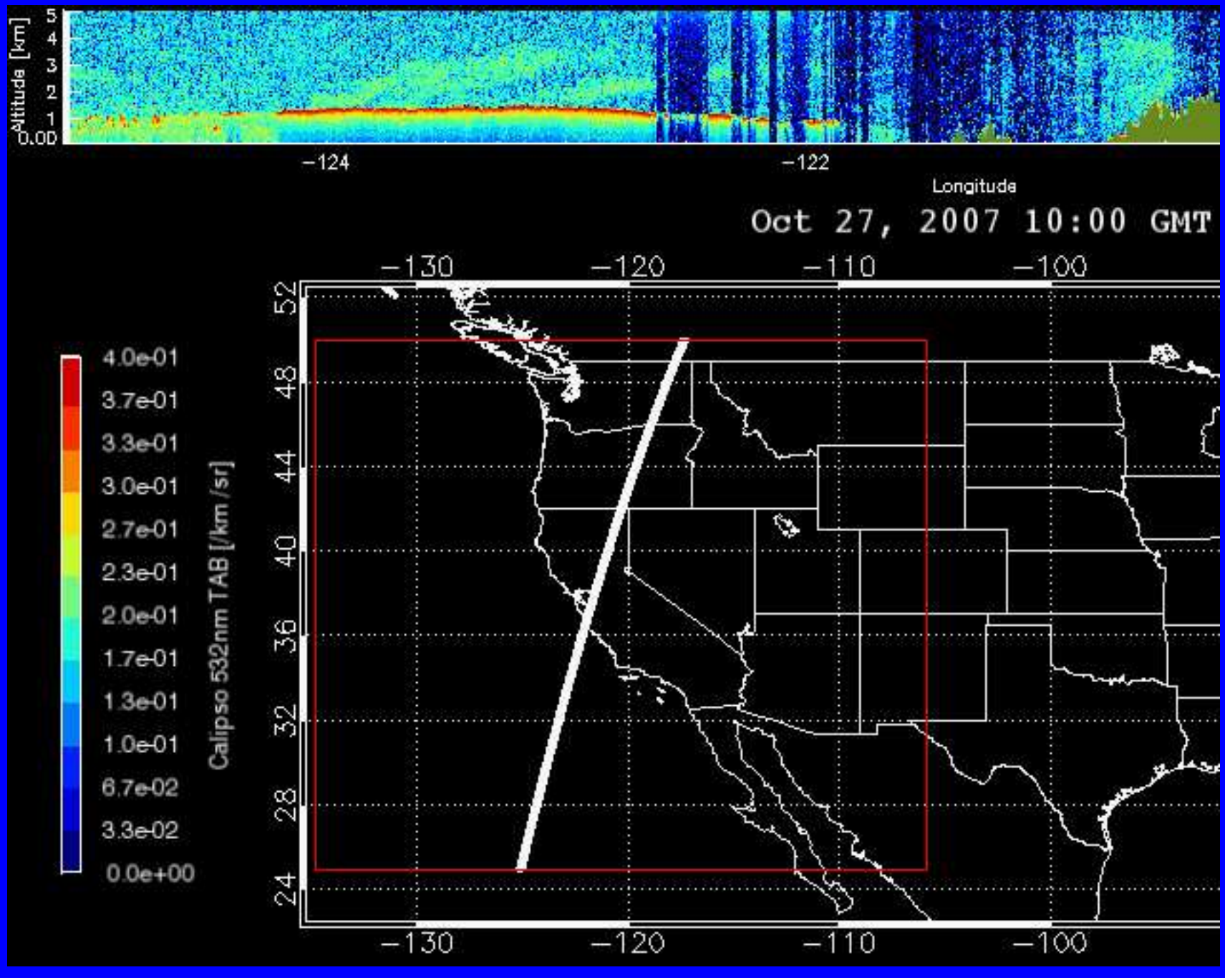
RSIG Data

MODIS global satellite surface–integrated data

(Ozone, Aerosol Optical Depth, Clouds, etc.)

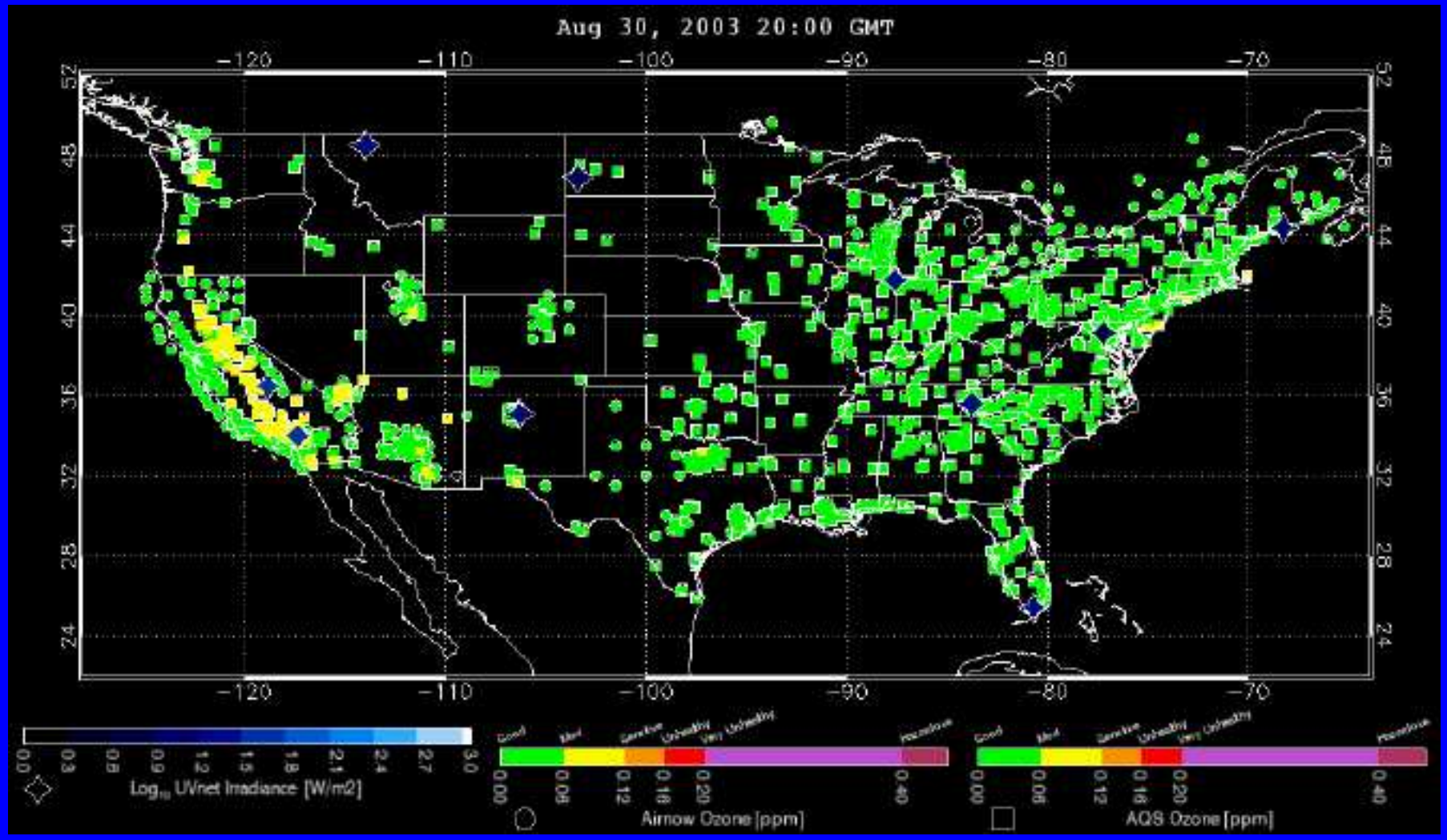


RSIG Data
CALIPSO global satellite LIDAR data (Backscatter).



RSIG Data

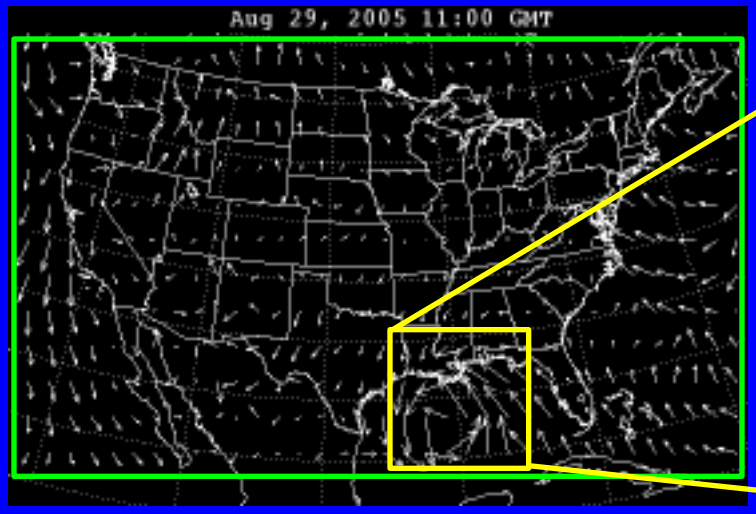
US Ground Stations:
Airnow and *AQS* (Ozone, PM2.5),
UVNET (Irradiance),
NESDIS/GOES–Biomass Burning (PM2.5, CO, etc.)

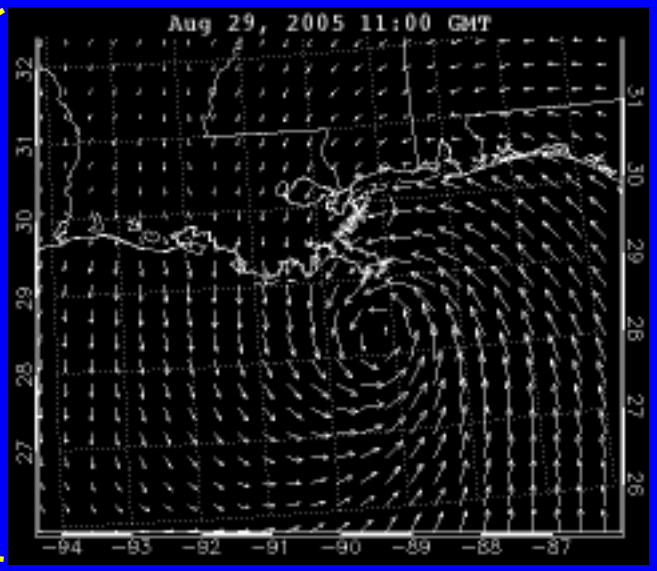


RSIG Unique/Key Features:

2. Subsetting

Time: extract & transfer only data within an hourly range (over many days).
Variable: extract & transfer only variables of interest.
Domain: extract & transfer only data within a chosen longitude–latitude box.





E.g., each (daily met) CMAQ file is 442 rows x 265 columns x 22 layers x 10 variables = 5GB!

Subset is only 2.5 MB.
1 / 2000th the source data size!
Convenient, fast and efficient to transmit across the network.

RSIG Unique/Key Features:

3. Aggregation

Combine data *from thousands of files* covering subset of interest *into a single data stream*.

E.g., Just 10 days worth of unsubsetted data occupies 200GB of disk space! MODIS data alone counts for about 3,500 files!

```
sapphire:/project/rsig/demo/data:52 >du -sk .
186730490       .
sapphire:/project/rsig/demo/data:53 >ls -1 MODIS/* | wc -l
3465
```

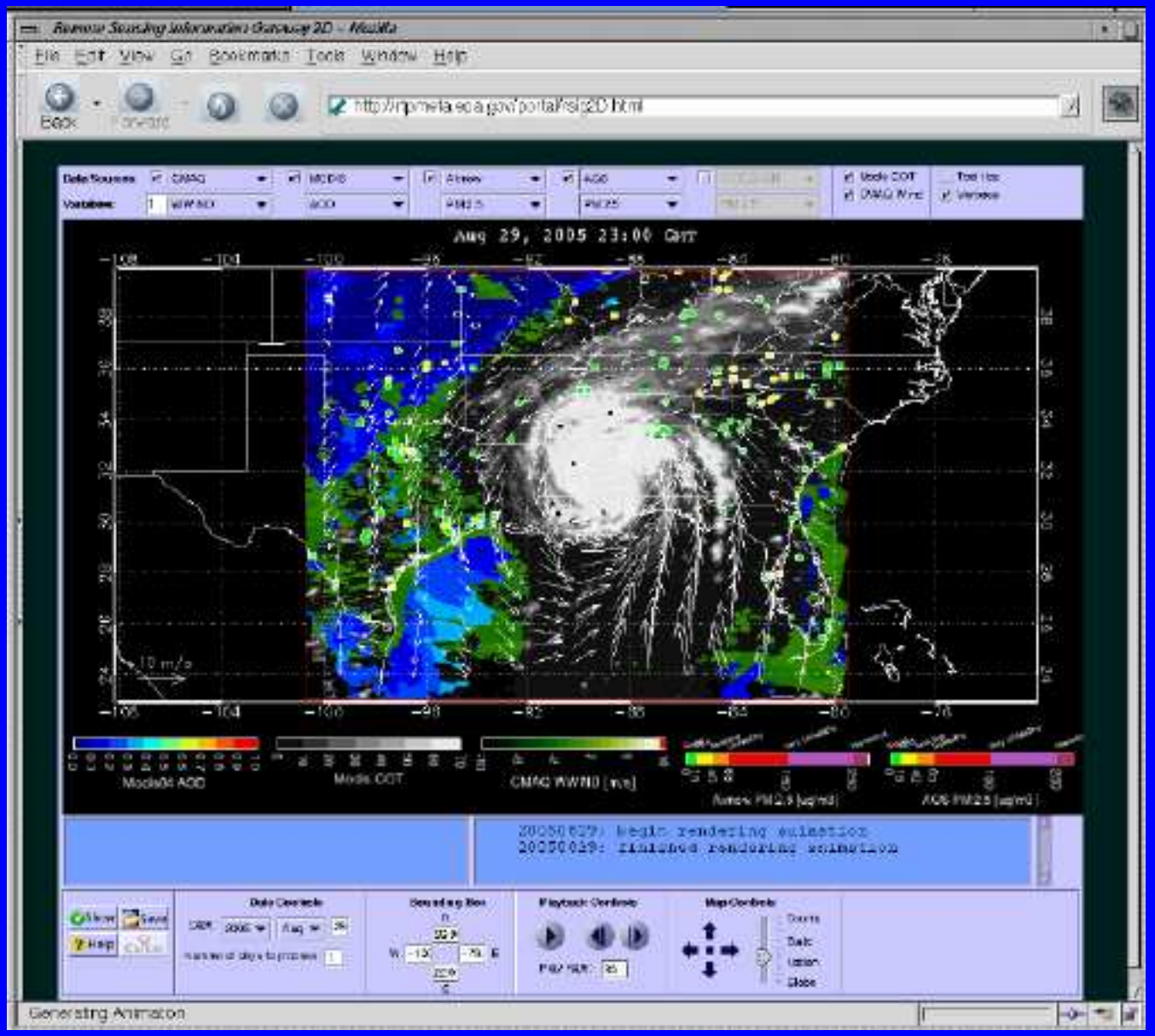
RSIG quickly determines which few files might cover the region of interest:

```
sapphire:/project/rsig/demo/data/MODIS/2005:59 >head bounds.mod04 ; echo "..." ; tail bounds.mod04
MOD04_L2.A2005237.0000.004.2005237130801.hdf  147.668506  175.317569   15.309453   36.463558
MOD04_L2.A2005237.0005.004.2005237125733.hdf  144.281166  169.226692   -2.459210   18.520328
MOD04_L2.A2005237.0010.004.2005237130934.hdf  139.782254  164.963577  -20.355334    0.630548
MOD04_L2.A2005237.0015.004.2005237130807.hdf  133.320187  161.766645  -38.259256  -17.076903
MOD04_L2.A2005237.0020.004.2005237125713.hdf  122.526061  159.414321  -56.189316  -34.394028
MOD04_L2.A2005237.0025.004.2005237130252.hdf  100.268832  160.528160  -74.027835  -50.942405
MOD04_L2.A2005237.0120.004.2005237130416.hdf -180.000000  180.000000   66.539196   89.997721
MOD04_L2.A2005237.0125.004.2005237130431.hdf -180.000000  180.000000   61.180824   86.201825
MOD04_L2.A2005237.0130.004.2005237130901.hdf  126.647362  174.841691   45.827809   68.277003
MOD04_L2.A2005237.0135.004.2005237130344.hdf  125.000007  157.798333   28.891350   50.370785
...
MOD04_L2.A2005242.2210.004.2005243200032.hdf -180.000000  180.000000  -77.625459  -54.067024
MOD04_L2.A2005242.2305.004.2005243203413.hdf -180.000000  180.000000   68.547156   89.999129
MOD04_L2.A2005242.2310.004.2005243193532.hdf -180.000000  180.000000   58.248390   82.582472
MOD04_L2.A2005242.2315.004.2005243201844.hdf -180.000000  180.000000   42.403632   64.570514
MOD04_L2.A2005242.2320.004.2005243191556.hdf -180.000000  180.000000   25.296542   46.667416
MOD04_L2.A2005242.2325.004.2005243191703.hdf -180.000000  180.000000    7.679840   28.732815
MOD04_L2.A2005242.2330.004.2005243191706.hdf  151.791562  176.508657  -10.162290   10.800181
MOD04_L2.A2005242.2335.004.2005243191600.hdf  146.612822  172.757702  -28.072702   -7.029580
MOD04_L2.A2005242.2340.004.2005243195829.hdf  138.743117  169.929589  -45.955020  -24.605614
MOD04_L2.A2005242.2345.004.2005243193213.hdf  124.505491  167.971551  -63.784922  -41.637905
```

RSIG Unique/Key Features:

4. Visualization

Interactive **animation (at 30 frames / second!) of all data on a map displayed in a web browser. Can save animation file in MPEG or KMZ format.**



RSIG Unique/Key Features:

5. Save subset to a local file in various formats

XDR binary – Simple human–readable ASCII header followed by simple portable binary data arrays (easy to read into model applications):

```
AIRNOW 1.0
Full US Domain
2004-07-22T00:00:00-0000
# data dimensions: timesteps stations
120 482
# Variable names:
PM25
# Variable units:
ug/m3
# IEEE-754 32-bit reals sites[stations][2=<longitude,latitude>] and
# IEEE-754 32-bit reals data[timesteps][stations]:
```

ASCII spreadsheet – Simple to import into a spreadsheet application:

Timestamp(UTC)	STATION(none)	LONGITUDE(deg)	LATITUDE(deg)	pm25(ug/m3)
2005-08-26T20:00:00-0000	280470008	-89.0497	30.3900	15.0000
2005-08-26T21:00:00-0000	280470008	-89.0497	30.3900	16.0000
2005-08-26T22:00:00-0000	280470008	-89.0497	30.3900	18.0000

...

RSIG Unique/Key Features:

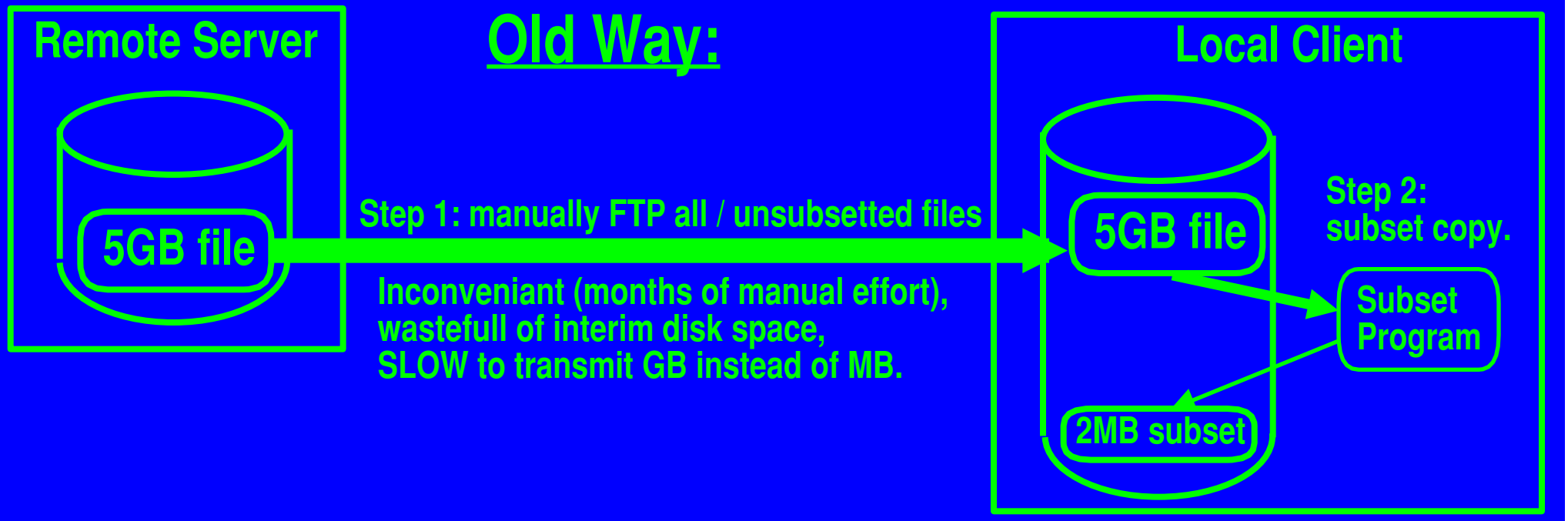
NetCDF – CMAQ/IOAPI or COARDS Conventions

```
netcdf modis {
dimensions:
        points = 169222 ;
variables:
        float longitude(points) ;
                longitude:units = "degrees_east" ;
        float latitude(points) ;
                latitude:units = "degrees_north" ;
        double Scan_Start_Time(points) ;
                Scan_Start_Time:units = "seconds since 1993-1-1 00:00:00.0 0" ;
        float Cloud_Optical_Thickness(points) ;
                Cloud_Optical_Thickness:units = "none" ;
                Cloud_Optical_Thickness:missing_value = -9999.f ;
        int yyyyddd(points) ;
                yyyyddd:units = "date" ;
        int hhmmss(points) ;
                hhmmss:units = "time" ;
        float time(points) ;
                time:units = "hours since 2005-08-26 03:25:00.0 -00:00" ;

// global attributes:
                :west_bound = -90.f ;
                :east_bound = -85.f ;
                :south_bound = 30.f ;
                :north_bound = 35.f ;
                :Conventions = "COARDS" ;
                :history = "http://daac.gsfc.nasa.gov/MODIS/Terra/
                product_descriptions_modis.shtml/,MODISSubset,XDRConvert" ;
}
```

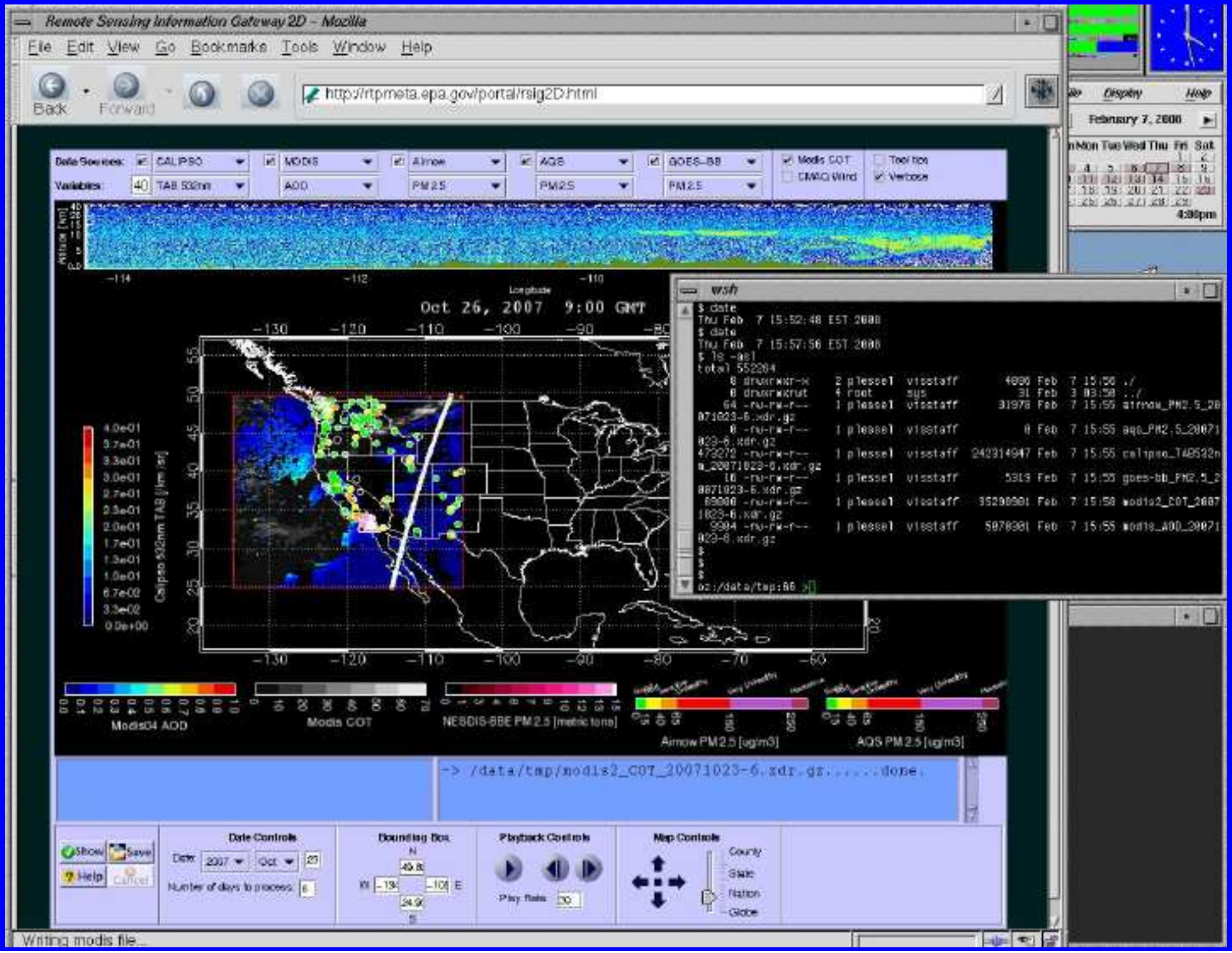
HDF – Original satellite files (huge & slow) for those that want them ...
PNG, MPEG, KMZ (GoogleEarth) formats for images/movies.
GeoTiff, *ARC GRID* and other requested output formats will be added.

RSIG Unique/Key Features:

6. Fast: *Months* of manual effort to get data – now done in *seconds!*



Scenario: Examine California Wildfires in 5 minutes



Data Retrieval and Vis: The RSIG Way – Quick & Easy

Visualization "about a minute per day":

Just 5 minutes to stream and visualize 6 days worth of data!
+ CALIPSO LIDAR Backscatter
+ MODIS Aerosol Optical Depth
+ MODIS Cloud Optical Thickness
+ Airnow PM 2.5
+ NESDIS Biomass Burning PM 2.5

Saving the data subset to local disk:

Just 5 minutes to stream & save full−resolution subsetted data and yielded under 250MB in compressed simple−format files.

```
$ date
Thu Feb  7 15:52:48 EST 2008
$ date
Thu Feb  7 15:57:56 EST 2008
$ ls −asl
    64 −rw−rw−r−−    1 plessel  visstaff      31978 Feb  7 15:55 airnow_PM2.5_20071023−6.xdr.gz
     0 −rw−rw−r−−    1 plessel  visstaff          0 Feb  7 15:55 aqs_PM2.5_20071023−6.xdr.gz
473272 −rw−rw−r−−    1 plessel  visstaff  242314947 Feb  7 15:55 calipso_TAB532nm_20071023−6.xdr.gz
    16 −rw−rw−r−−    1 plessel  visstaff       5319 Feb  7 15:55 goes−bb_PM2.5_20071023−6.xdr.gz
 69000 −rw−rw−r−−    1 plessel  visstaff   35298901 Feb  7 15:58 modis2_COT_20071023−6.xdr.gz
  9904 −rw−rw−r−−    1 plessel  visstaff    5070301 Feb  7 15:55 modis_AOD_20071023−6.xdr.gz
```

− How long would this have taken using web form+email+ftp?
− How many files would need to be downloaded?
− How many GB of disk space would be needed to store the files?

Data Retrieval: The Old Way

To get 6 days of CALIPSO data for the region of interest:

1. Register with the NASA Goddard DAAC to create an account.
2. Log in to the DAAC web site.
3. Select the appropriate "project" and "parameters" – after you've figured–out the NASA nomenclature.
4. Wait 3 days for an email telling you how to download the files.
5. FTP the HDF files.
   In this case, 144 files @ ~0.5GB / file = 72GB and takes 50 hours to download.
6. Process the HDF files into useful, subsetted formats.
   (You'll discover that only 23 of the 144 files actually contain a track that intersect the continental US.)

Most so–called web–accessible data sources use similarly slow, cumbersome and limited approaches.

Back Forward

http://eosweb.larc.nasa.gov/HORDE?BSNMITWL_Start.cgi

SEARCH FILES

Login
New User
Update Info
Retrieve Login & Password
CD-ROM & Videocassette
Tool Unavailability

Please check out the Tool Unavailability page to see schedule down times for this tool.

Projects

CALIPSO

Total: 1

Parameters

AEROSOL OPTICAL THICKNESS
AEROSOL PARTICLE PROPERTIES
BACKSCATTER COEFFICIENT
BAROMETRIC PRESSURE

Total: 29

Refine List | Reset Lists

Files | Data Sets | Get Info

0020 CAL_IIR_L1-Prov-V1-10
1782 CAL_LID_L1-Prov-V1-19
1804 CAL_LID_L1-Prov-V1-11
4553 CAL_LID_L1-Prov-V1-20
2440 CAL_LID_L1-Prov-V1-22
13898 CAL_LID_L1-Prov-V2-01
3258 CAL_LID_L2_01kmCLay-Prov-V1-10
3857 CAL_LID_L2_01kmCLay-Prov-V1-20

Total: 25

ADVANCED SEARCH

Time Ranges	Start Date	Stop Date
Optional (YYYY-MM-DD)	1993-01-01	2008-00-05
Geographic Search	Top: 90.00	
	Left: -180.00	180.00 Right
	-90.00 Bottom	

DayNight: All Day Night

Clear Form | Get Results

Data Retrieval and Subsetting: The RSIG Way

Uncompressed XDR files total just 1GB.

```
$ uncompress *.gz
$ ls −asl
   224 −rw−rw−r−−   1 plessel  visstaff    114367 Feb  7 15:55 airnow_PM2.5_20071023−6.xdr
1615560 −rw−rw−r−−   1 plessel  visstaff  827165389 Feb  7 15:55 calipso_TAB532nm_20071023−6.xdr
   128 −rw−rw−r−−   1 plessel  visstaff     61444 Feb  7 15:55 goes−bb_PM2.5_20071023−6.xdr
 292400 −rw−rw−r−−   1 plessel  visstaff  149704919 Feb  7 15:58 modis2_COT_20071023−6.xdr
  36232 −rw−rw−r−−   1 plessel  visstaff   18548851 Feb  7 15:55 modis_AOD_20071023−6.xdr
```

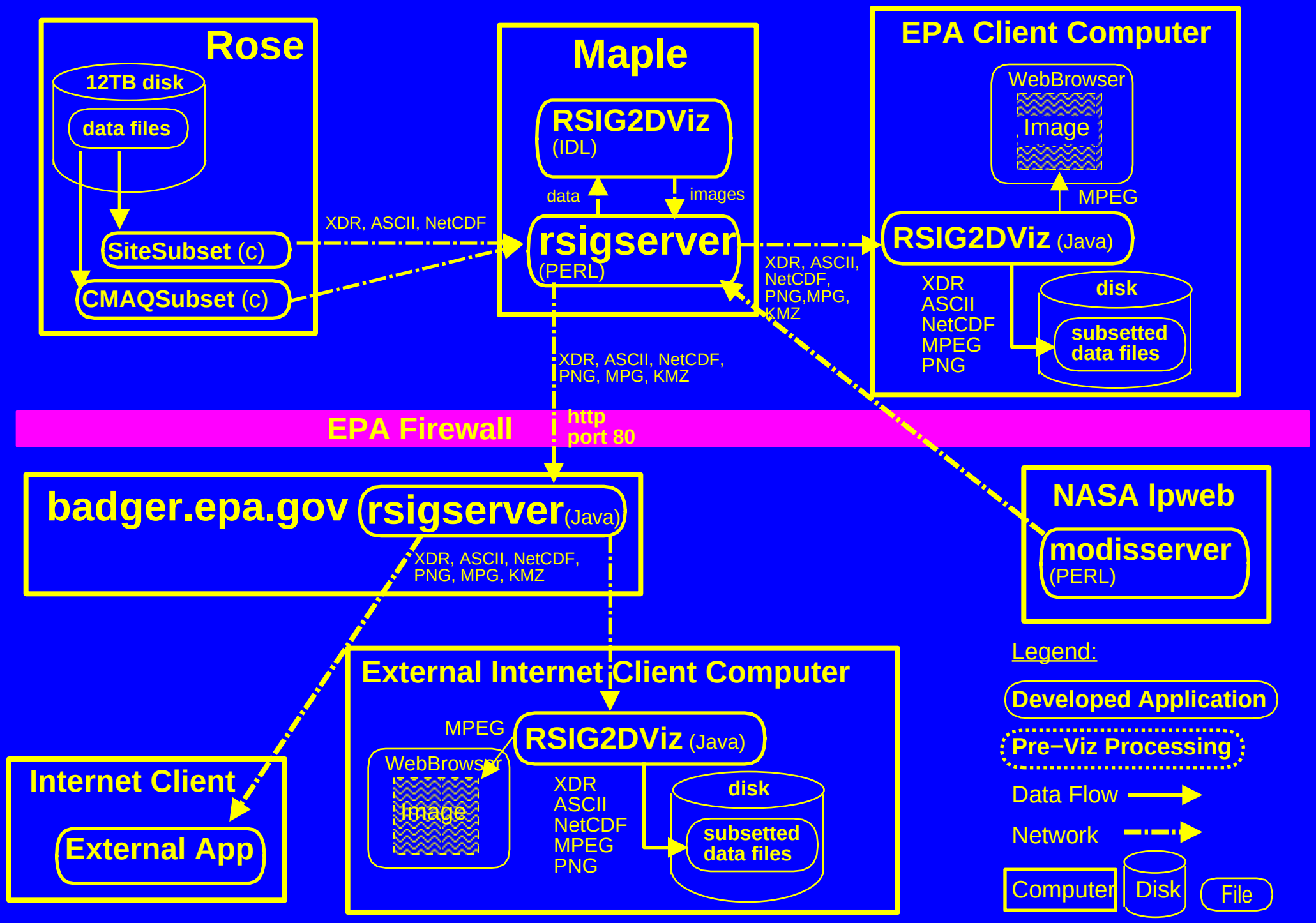
ASCII headers and XDR binary arrays are efficient and easy to read/parse:

```
$ head −12 airnow_PM2.5_20071023−6.xdr
AIRNOW 1.0
subset
2007−10−23T00:00:00−0000
# data dimensions: timesteps stations
144 194
# Variable names:
pm25
# Variable units:
ug/m3
# MSB 32−bit integers ids[stations] and
# IEEE−754 32−bit reals sites[stations][2=<longitude,latitude>] and
# IEEE−754 32−bit reals data[timesteps][stations]:
```

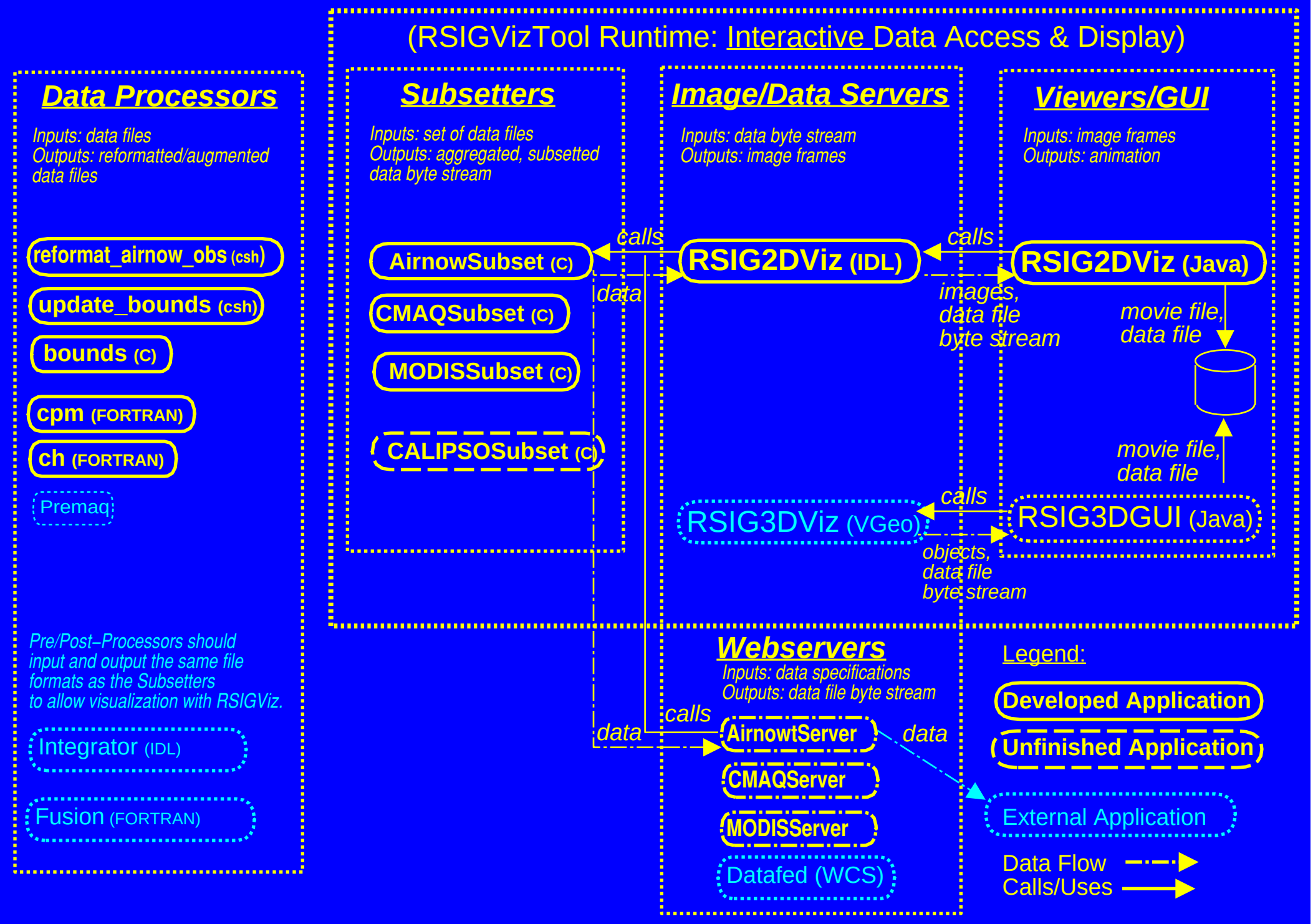
Example IDL code to Read Airnow XDR stream

```
command = 'wget -q -c -T 0 -O - ' + "'" + $
          'http://badger.epa.gov/rsig/rsigserver?' + $
          'SERVICE=wcs&VERSION=1.0.0&REQUEST=GetCoverage&' + $
          'COVERAGE=airnow.pm25&' + $
          'TIME=2005-08-26T20:00:00Z/2005-08-27T01:59:59Z&' + $
          'BBOX=-90,30,-88,35,0,0&FORMAT=xdr' + "'"

spawn, command, unit = pipe

; Read ASCII header of XDR-format data stream:

header = strarr( 12 )
line = ''

for index = 0, 11 do begin
  readf, pipe, line
  header[ index ] = line
endfor

timestamp = header[ 2 ] ; E.g., 2007-10-23T00:00:00-0000.
variable  = header[ 6 ] ; E.g., pm25.
units     = header[ 8 ] ; E.g., ug/m3.

; Allocate data arrays:

dimensions = strsplit( header[ 4 ], ' ', /extract )
timesteps  = long( dimensions[ 0 ] ) ; E.g., 144.
stations   = long( dimensions[ 1 ] ) ; E.g., 194.
ids        = lonarr( stations )
lonlats    = fltarr( 2, stations ) ; IDL multi-dimensional arrays are column-major like FORTRAN...
data       = fltarr( stations, timesteps )

; Read XDR binary data arrays:

readu, pipe, ids
readu, pipe, lonlats
readu, pipe, data
swap_endian_inplace, ids,     /swap_if_little_endian
swap_endian_inplace, lonlats, /swap_if_little_endian
swap_endian_inplace, data,    /swap_if_little_endian
close, pipe
free_lun, pipe
result = { timestamp: timestamp, variable: variable, units: units, $
           timesteps: timesteps, stations: stations, ids: ids, lonlats: lonlats, data: data }
print, result
```

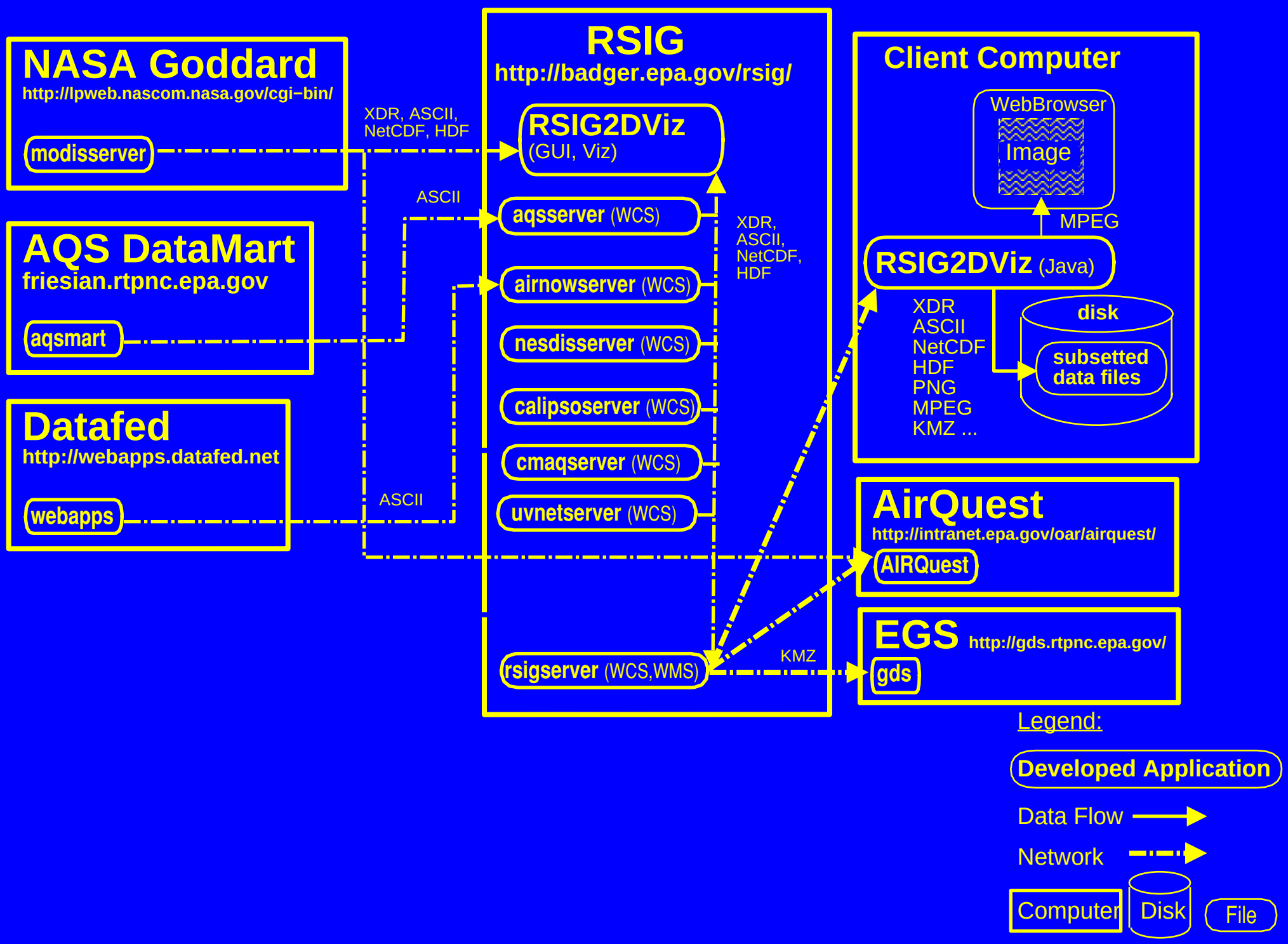
Current RSIG Runtime Data Flow Diagram



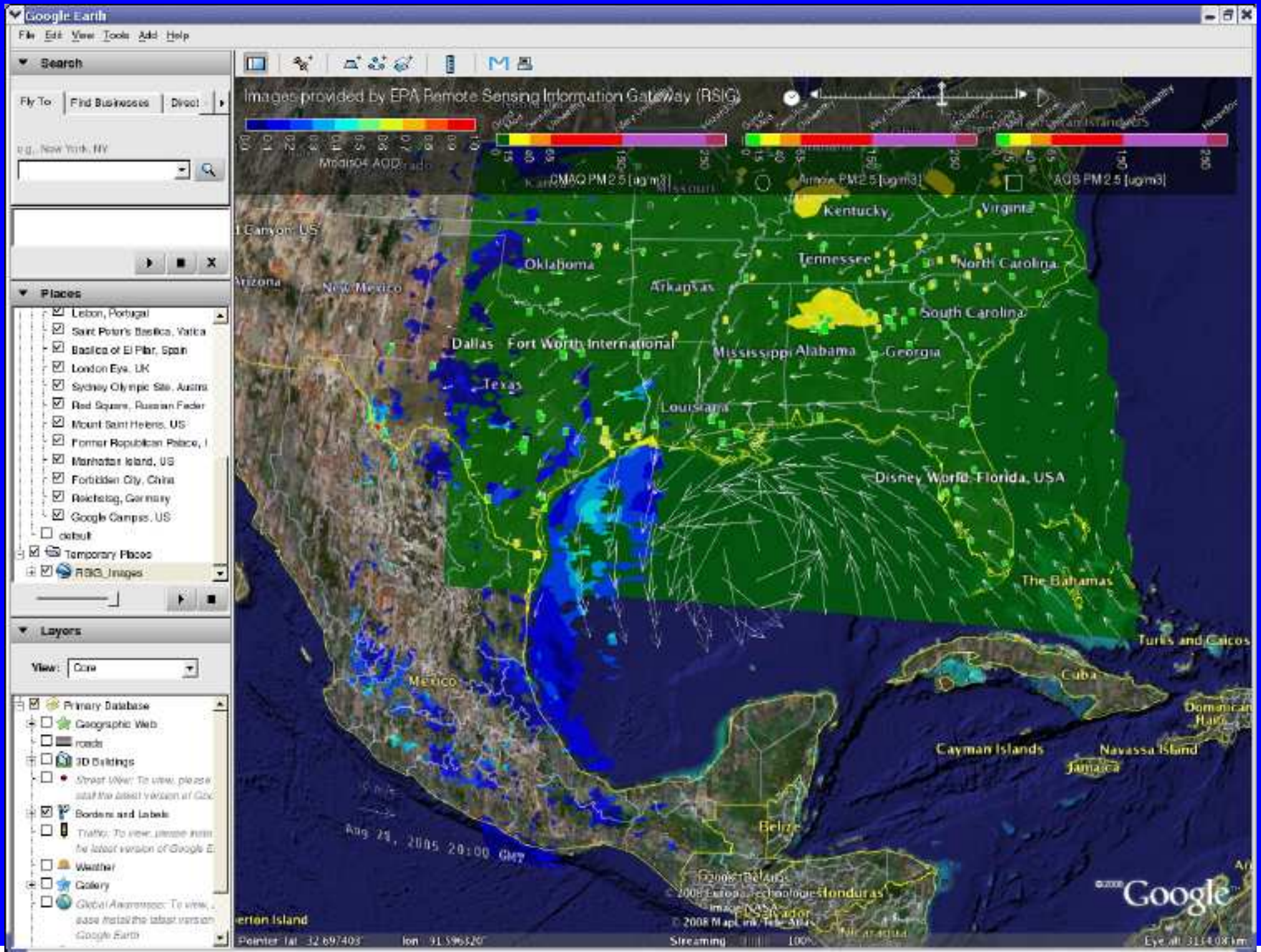
RSIG System Component Diagram



RSIG System Connections



RSIG –> GoogleEarth



Interoperability Data Standards Used By RSIG

OGC–WCS/XML	Describes available data (variables, time–range,...)
OGC–WCS	Specify data subset by variable, time, lon–lat box.
OGC–WMS	Specify and receive images and movies.
IEEE–754	Computer–readable format for real–valued data streamed from WCS.
XDR	Efficient streamable computer–readable format for real and integer data.
NetCDF	Widely–used file format for arrays of scientific data.
NetCDF/COARDS	NetCDF conventions for geo–spatial data.
NetCDF/IOAPI	NetCDF conventions for gridded model met and AQ data.
HDF	Widely–used file format for arrays of scientific data.
HDF–EOS	HDF conventions used by NASA.
ASCII	Human–readable format for small amounts of data streamed from WCS.
PNG	Most portable image file format.
MPEG	Most portable movie file format.
KML/KMZ	Geo–referenced image/movie format used by GoogleEarth and similar applications.
HTML	Used for web–based documentation.
PDF	Portable Document Format for fancy–formatted content.
HTTP/80	Communication protocol for web–based applications.
SI	Mother of all standards... RSIG converts data to SI units. (F to C, mb to Pa, western longitudes are negative, UTC, etc.)

Why Standards Are Not Enough

In practice, interoperability requires more than just compliance with existing standards. The devil is in the details. Here are some gotchas we had to overcome:

– OGC–WCS DescribeCoverage and GetCoverage omits data units (!)
  The RSIG WCS servers include them in the label so users don't have to guess:

```
<label>ozone(ppb)</label>
<label>pm25(ug/m3)</label>
<label>pres(Pa)</label>
<label>co2(metric tons)</label>
```

– WCS does not nail–down the specifics of the format of the resulting data. E.g., NetCDF–COARDS vs NetCDF–IOAPI, CSV, etc. (No universal Data Models.)

– HDF/4/5/EOS: NASA satellite data files use inconsistent conventions for describing/encoding/structuring data. E.g., Bounding box, variable names, dimensionality, data order, units (non–SI, omitted!), SDS, VData.

– Performance:
  – Too slow if data files are on tape or
  – if DB has no appropriate index file.
  – Need formats that are fast, efficient, simple and streamable (i.e., no disk I/O) like RSIG's custom ASCII/XDR format vs NetCDF.
  – Custom Extensions to WCS: STRIDE=2, SPARSE=800, COMPRESS=1

– Units (SI), time (UTC), missing data value sentinels, etc.

Additional Value–Added by RSIG

To facilitate combining data from various sources, RSIG adds value as follows:

+ Units:
  + Always included.
  + Conversion: to SI, ppm, negative Western longitudes, GMT.
+ GMT–timestamp, Longitude, Latitude are always included, and where possible, Elevation (m above MSL).
+ Missing data convention.
+ Data word size (64–bit).
+ Data order, e.g., ground–up to match CMAQ model.
+ On–the–fly compression/decompression to speed–up network transfers.
+ Regridding to CMAQ grid.
  1. Project onto CMAQ grid.
  2. Aggregate: nearest, mean, weighted (1/r^2).
  3. Save as NetCDF–IOAPI format to facilitate reading by existing CMAQ apps.
+ Metadata: RSIG's approach to metadata:
  1. Provide a link (URL) to the data source web site.
  2. Where possible, include this link and append subsequent processing in output files (depending on format, e.g., NetCDF–COARDS: history)

```
:history = "http://daac.gsfc.nasa.gov/MODIS/Terra/product_descriptions_modis.shtml/,MODISSubset,XDRConvert"
```

Planned Enhancements (as of May 2008)

+ Install calipsoserver WCS (and CALIPSOSubset, XDRConvert) on a publicly−accessible computer at NASA Langley. This will provide on−demand streaming subsets of the entire range (2005 – present) of CALIPSO data.

+ Add new data sources including:
  + MOPITT CO, AURA−OMI NO2, SCIAMACHY NO2, GASP, several more data products from DATAFED.net, etc.
  + What external data would you like to access via RSIG?
  + What EPA data would you like to share via RSIG?

+ Add new output formats: GeoTiff, ESRI/ARC Grid.

+ Launch HB/MCMC and read and display resulting data from HB and LARCTM.

+ Assist others in connecting RSIG to their own applications and projects. (E.g., using rsigserver WCS and WMS to stream data and images into other analysis and visualization applications.)

Please send RSIG suggestions, questions and comments to

plessel.todd@epa.gov 1−919−541−5500 or
freeman.matt@epa.gov 1−919−541−4293