Earth Science Data: Why So Difficult??!@#!!

Chris Lynnes

The Way Things Should Be

Alice wants to determine a volcanic eruption's vertical and horizontal extent, at the highest possible spatial resolution

1. In her analysis tool, she types "aerosols" in a search box
2. The tool presents a table of aerosol choices, showing key charcteristics in a comparison table:

- MISR aerosols from LaRC
- MODIS aerosols from MRDC
- CALIPSO vertical classifications from LaRC
- OMI UV aerosols from GES DISC
- Multi-sensor climatology computed by Giovanni

- GOCART model from DataFed
- CALIPSO SO2/aerosol experimental discrimination from Dr. Bob
- HIRDLS profile from GES DISC
- POLDER aerosols from CNES
- GOES-R aerosols from NOAA
- and more (ACE, GEO-CAPE, NPOESS

3. Alice checks desired boxes, draws a box around the volcano, se date/time and clicks "Go"
4. Within a minute or so, all data are in her analysis tool, pre-subsetted, reformatted, with quality screening applied

The Way Things Are

- Look through a bunch of directories to find the right datasets
  - Probably miss some good ones
- Learn and use several different search tools to get data files
- Use a special interactive interface to get the climatology
- Wait for email for some data, then go pick up
- Get data in at least 5 different formats, myriad data structures
  - Got to reformat to get them into analysis tool
  - Got to reproject somehow
  - Got to integrate vertical profile with horizontal data
- Spend days reading documentation on quality
  - Spend more days writing code to screen data
- Hear about Dr. Bob's experimental data through good luck
  - JPEG picture is nice but can't be integrated with data
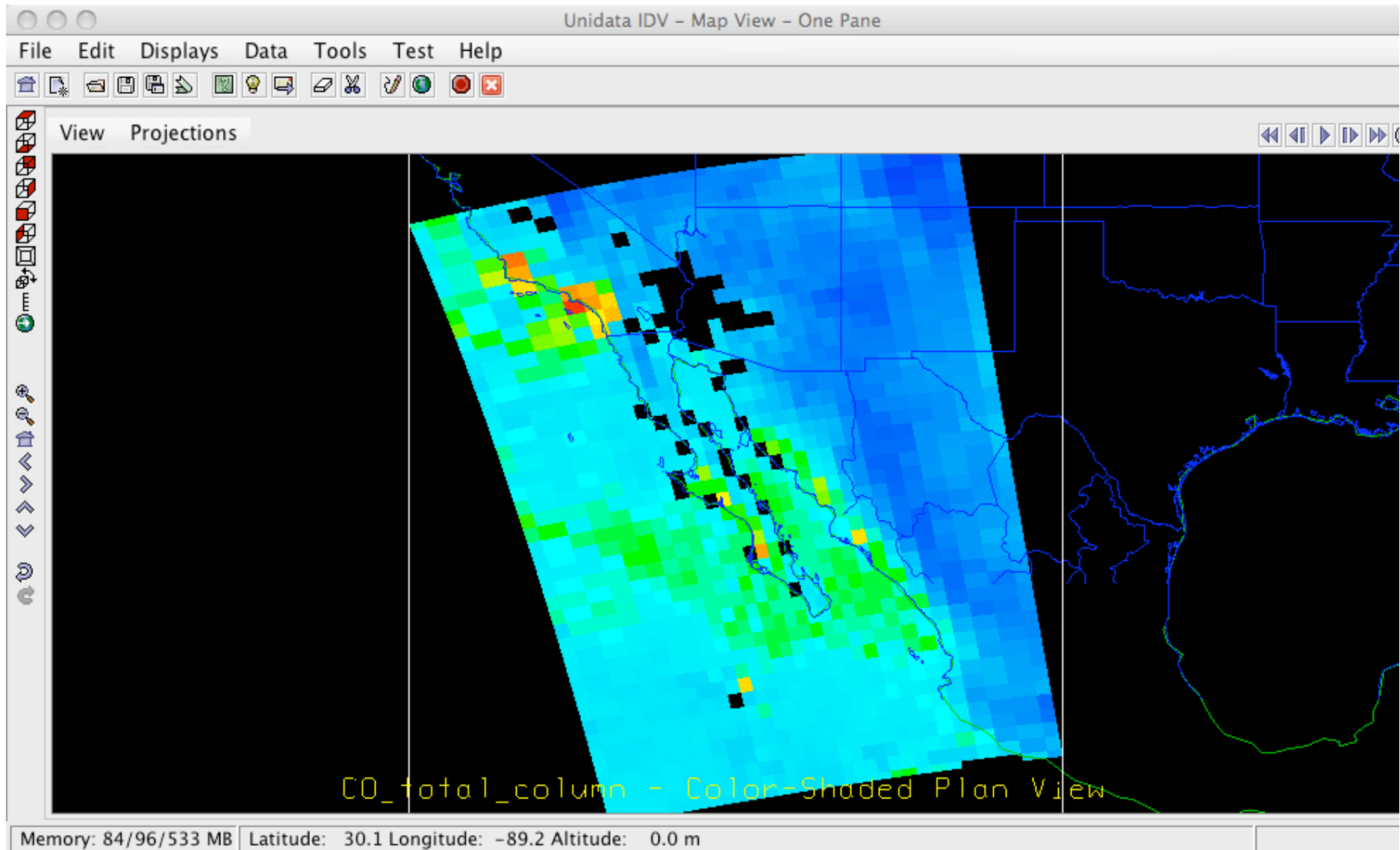
Why So Difficult???!@#$%!!

We already know how to do this stuff!

We know how to do:
- Distributed search (EOSDIS V0)
- Subsetting (HEW, OPeNDAP)
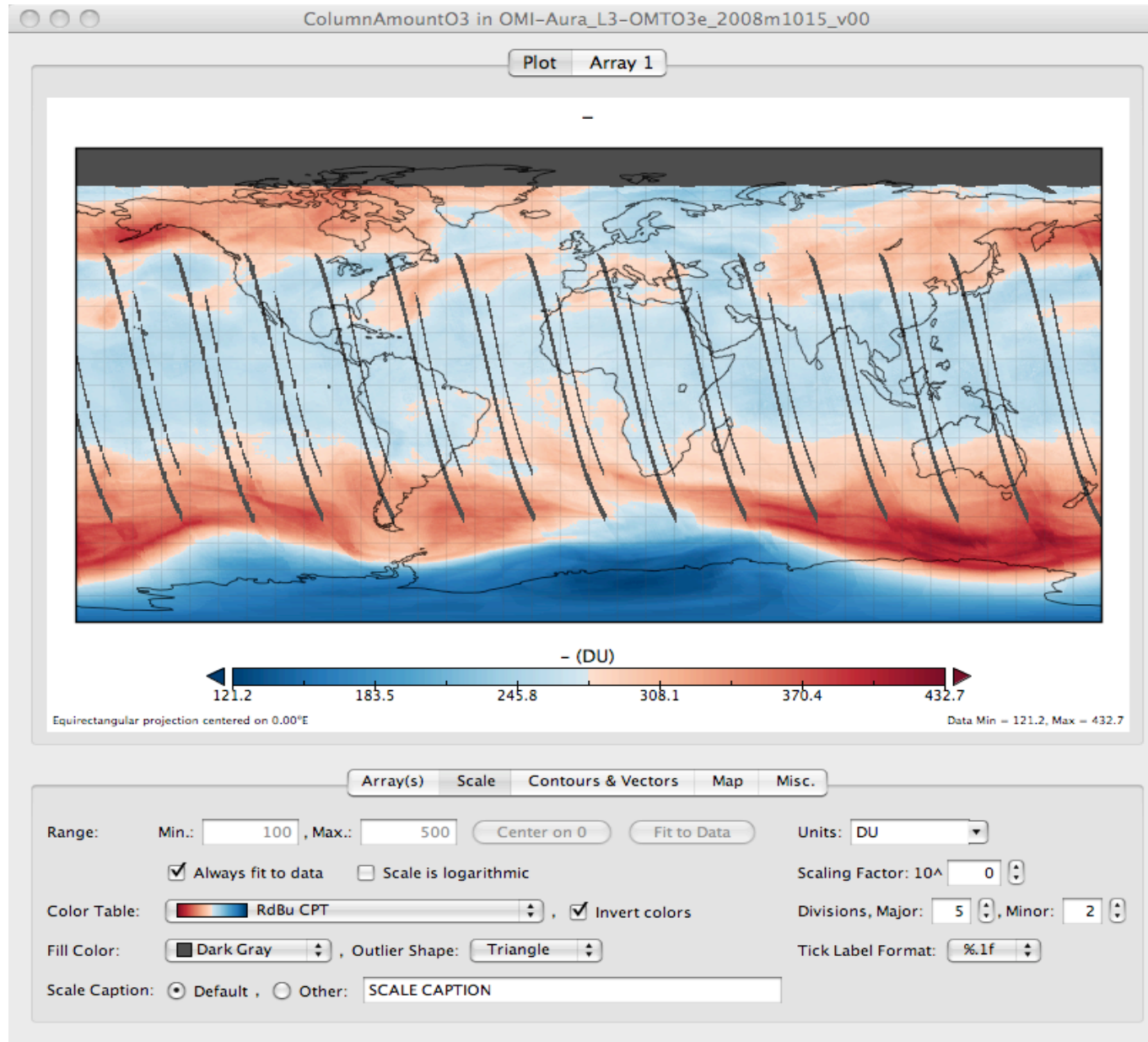- Reformatting (NetCDF/CF-1)
- Quality Screening

This should be a matter of tweaking, deploying, SMOC (simple matter of coding

Enabling Technology: CF-1

*AIRS Swath data in NetCDF-CF1, displayed in IDV
WITHOUT any pre-processing*

Enabling Technology: OPeNDAP

OMI Level 3 data remotely acquired and displayed in Panoply by providing just a URL (no preprocessing)

Other Demonstrated Capabilities

- Server-side quality screening
  - Giovanni
  - CEOP Satellite Data Server
  - Data Quality Screening Service
- Multi-center search
  - Federated: EOSDIS Version 0, OpenSearch
  - Centralized: ECHO
- Self-publication (Dr. Bob's case)
  - Datacasting
- Plug-in search modules
  - Firefox / OpenSearch

What Do We Need?

- More use of enabling technologies
  - OPeNDAP, OpenSearch, WCS, CF-1
  - A little tweaking to make them more powerful and attractive
- Easy-to-use, turnkey tools for deployir technologies
- Gateways, proxies and agents for thos who cannot deploy the technologies