

The Emissions Modeling Framework:

A System for Managing and Running Complex Environmental Simulations and Data

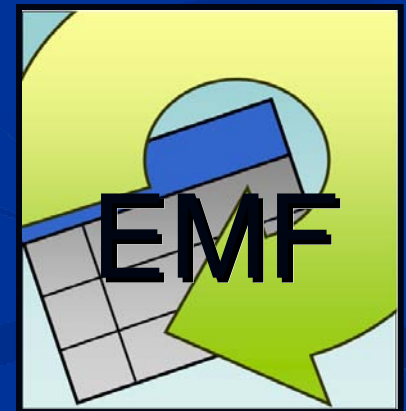
Alison Eyth

U.S. EPA

Office of Air Quality

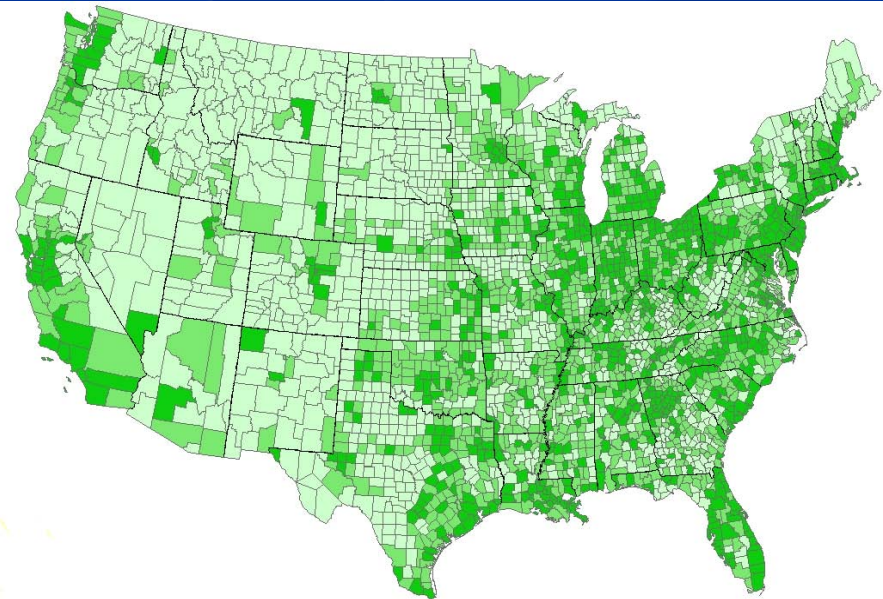
Planning and Standards

eyth.alison@epa.gov



Emissions Inventories

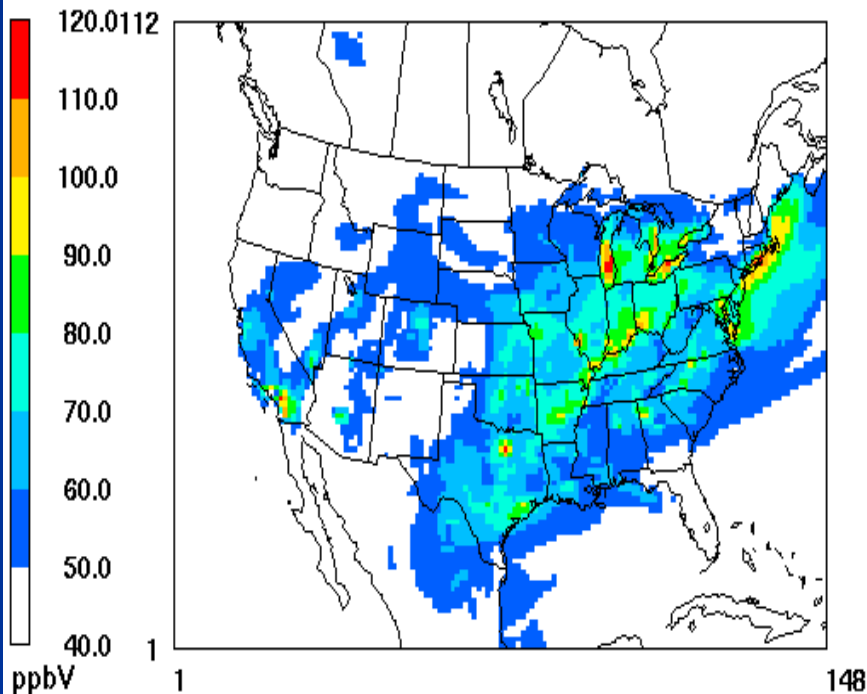
- **Emission sources** release pollution into the atmosphere (e.g., factories, motor vehicles, gas stations, home heating)
- Depending on the type of source, amounts of emissions are estimated for **stacks** or **counties**
- **Emission inventories** are compiled lists of emissions source data for multiple pollutants and source categories
- Data are frequently given as **annual values** (e.g. tons/yr)
- Inventories provide input data for **air quality models** (AQMs), the outputs of which are used to make decisions about rules



Air Quality Model Results

Max. 8hr-Average Ozone Conc.

w=2002ac_tox_v4.61_L3th_us36b.aconc.o3_8hrmax_LST.ioapi

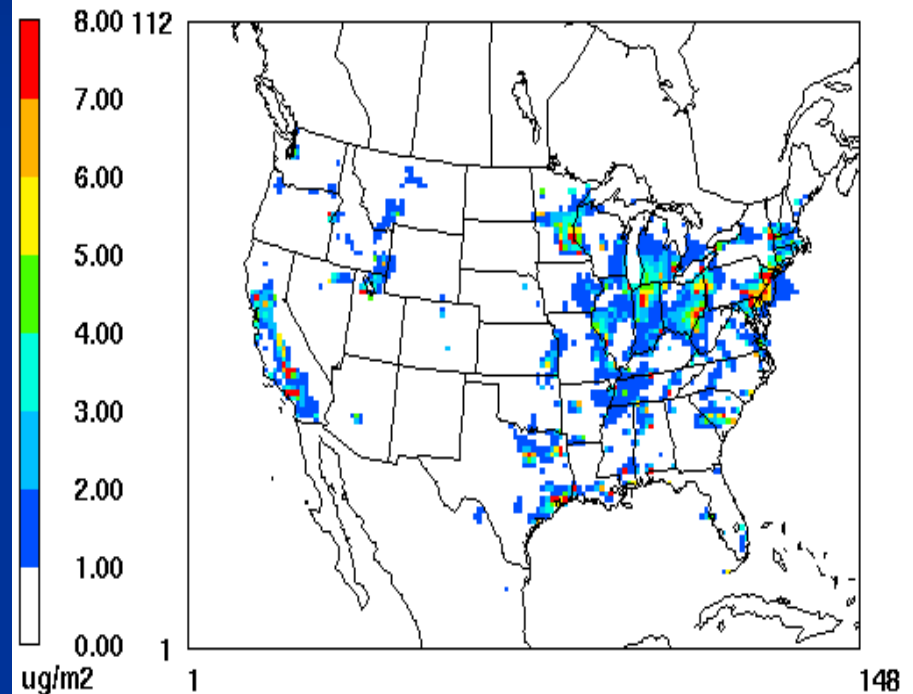


July 8, 2002 0:00:00
Min= 17.4 at (86,112), Max= 118.1 at (25,45)

Total Mercury Deposition

Daily Total All Species

z=2002ac_tox_v4.61_L3th_us36b.dailysum.dep.07

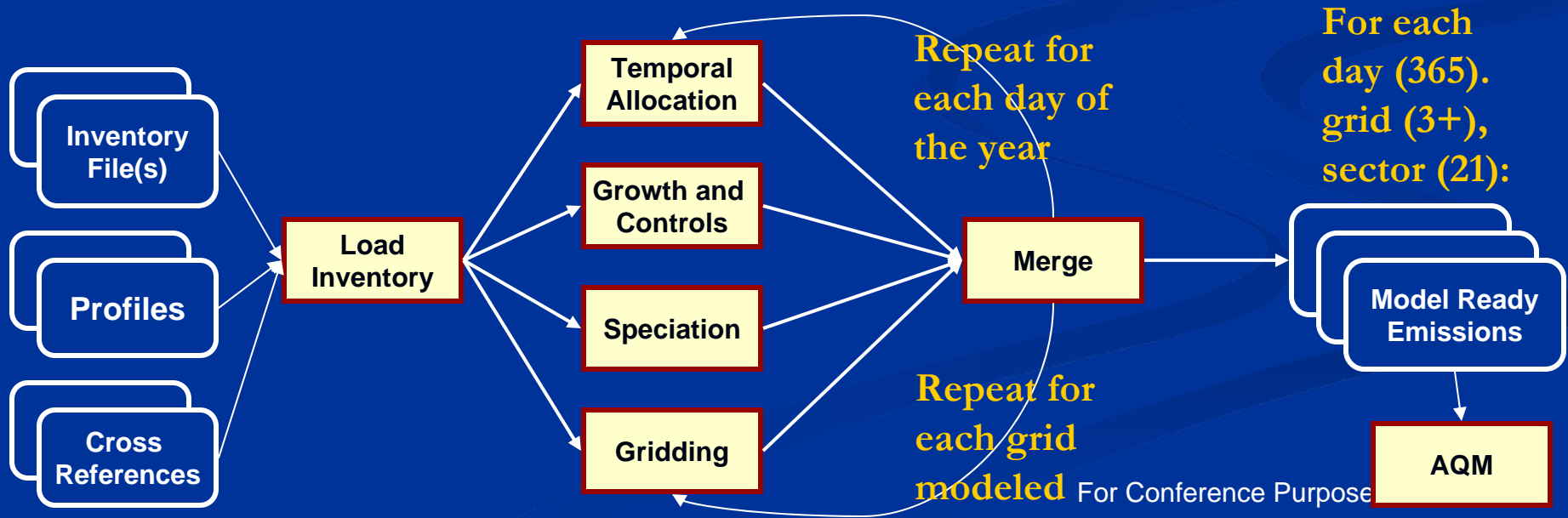


July 8, 2002 1:00:00
Min= 0.00 at (67,112), Max= 55.88 at (127,67)

- AQMs estimate the **concentration of pollutants** in the atmosphere (e.g., CO, NO, NO₂, O₃, Pb, Hg, SO₂) and their **deposition**
- Regional-scale AQMs are run for **grids** with rectangular grid cells

Emissions Modeling

- Processes emissions inventories to match the resolution needed by the air quality model (hourly, by grid cell, for specific pollutant species)
 - At EPA, inventory data are split into **21 sectors**
 - Data for each sector goes through **multiple processing steps**
 - Processing is usually done for **each region (grid)** and for **each day** of the year being modeled



Why Build the EMF?

- For each one year AQM run:
 - *Hundreds* of program executions are required
 - *Hundreds* of input files are used
 - *Thousands* of AQM-ready files and reports are created
- Input files get **modified** over time into many versions
- Lots of different **things can go wrong** during a run
 - If an error is discovered, re-runs can require weeks/months
- Inventories can have *millions* of records (need QA)
- There are **multiple simultaneous model runs** (i.e., cases) to manage to support different projects
- **Multiple people** need to access the same model input and output data (e.g., EPA staff, contractors)

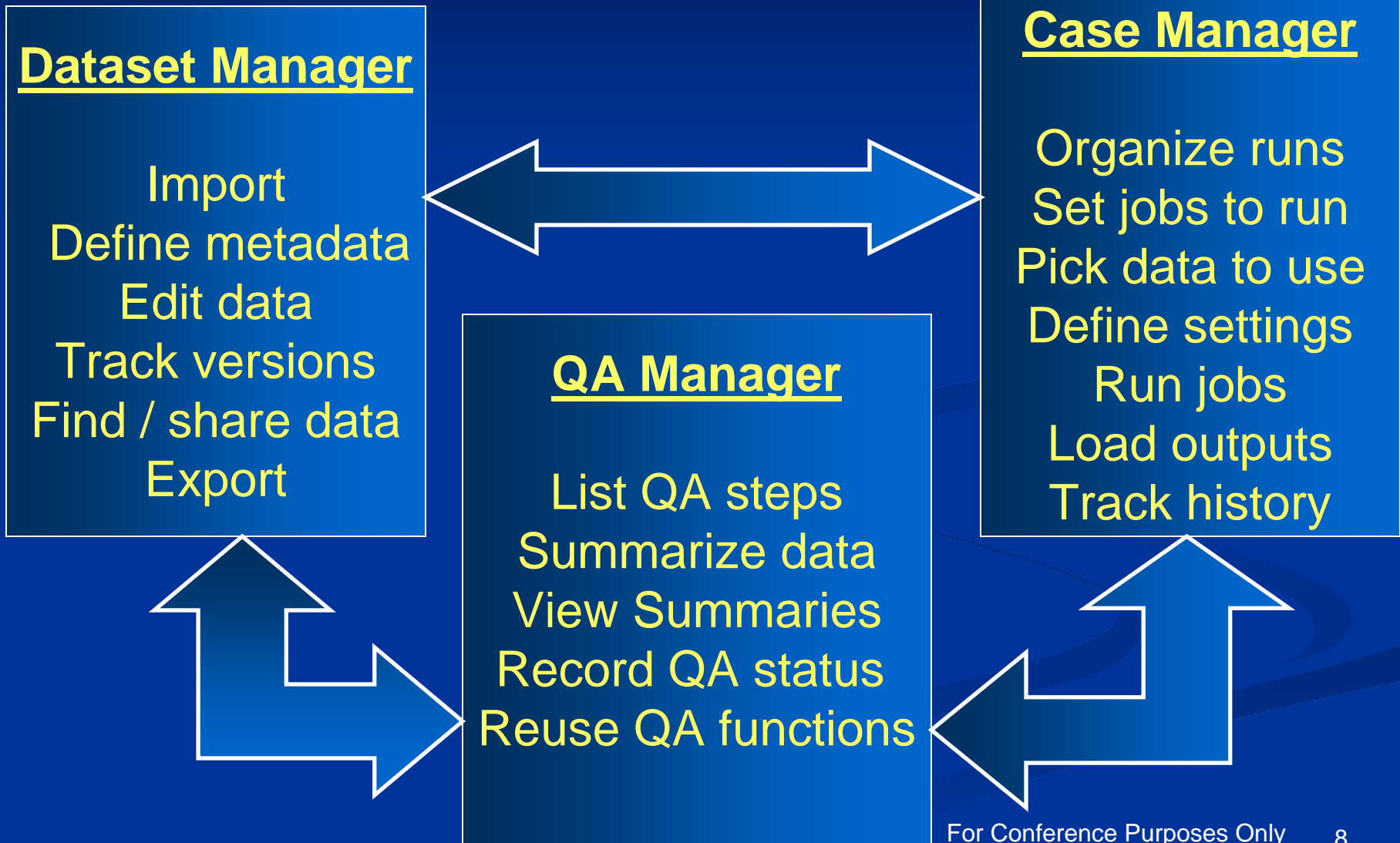
Prior Attempts at Modeling Frameworks

- EPA has previously tried to solve this data and run management problem for air quality modeling
 - **Models-3 (1990s)**: only the visualization tool and I/O interfaces were successful, not the data/run management
 - **Multimedia Integrated Modeling System (early 2000s)**: an object-oriented view of integrated modeling but was not widely adopted
- Frameworks to support other types of models (e.g., water quality, sediment) have been built by EPA and others
 - **Community of Practice for Integrated Environmental Modeling** has been organized
- Most prior frameworks were *desktop* systems that supported the individual model user, not workgroups

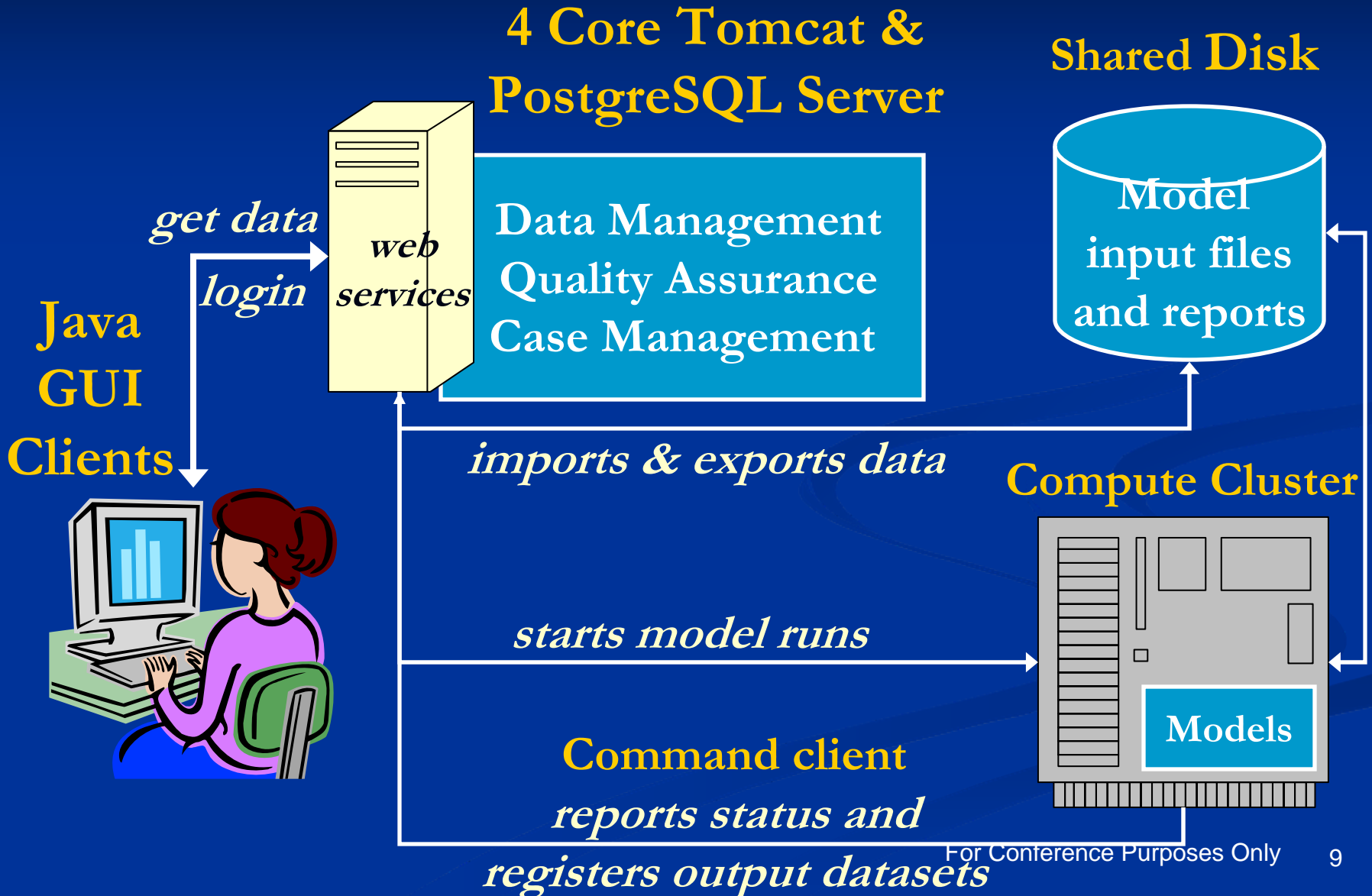
Goals of the EMF

- Improve the *timeliness and quality* of the emissions data used by air quality models
 - Reduce the need for reruns due to mistakes
- Provide *transparency & tracking* of data and model settings
 - For datasets, track versions, metadata, and provide integrated quality assurance (QA) steps
 - Improve communication within the workgroup regarding versions of data and settings used for modeling cases
- Make it *easier* for people to run emissions and related models
- For regulatory purposes, make it possible to *reproduce data* being created today for years into the future

Major Components of the EMF



EMF Architecture at EPA



EMF Main GUI

All other EMF windows appear within this Main window

Messages about the status of long-running processes (e.g., imports, QA summaries, model runs) appear in the Status window

| Message Type | Message | Timestamp |
|--------------|---------|-----------|
| | | |

- Stored information (datasets, cases) is accessed via the Manage menu
- Choosing an item from the Manage menu opens a 'Manager' window

Dataset Manager

Dataset Manager

Refresh

Show Datasets of Type: ORL Point Inventory (PTINV) Name Contains: Advanced

| # | Select | Name | Last Modified Date | Type | Status | Creator |
|----|--------------------------|---|--------------------|-----------------------------|-----------------------------|----------|
| 1 | <input type="checkbox"/> | ptnonipm_xportfrac_cap2020ck1_14_35 | 2009/12/01 13:23 | ORL Point Inventory (PTINV) | Created by control strategy | rhc |
| 2 | <input type="checkbox"/> | seca_c3_caps2020pf31_14_35 | 2009/11/30 15:13 | ORL Point Inventory (PTINV) | Created by control strategy | rhc |
| 3 | <input type="checkbox"/> | ptnonipm_hap2015ck_CoST | 2009/11/27 10:22 | ORL Point Inventory (PTINV) | Created by control strategy | rhc |
| 4 | <input type="checkbox"/> | ptnonipm_xportfrac_cap2015ck_CoST | 2009/11/27 10:20 | ORL Point Inventory (PTINV) | Created by control strategy | rhc |
| 5 | <input type="checkbox"/> | ptipm_EPA302ARRA_CP_7_summer_2020_14_35 | 2009/11/24 17:58 | ORL Point Inventory (PTINV) | Created by control strategy | rhc |
| 6 | <input type="checkbox"/> | ptipm_EPA302ARRA_CP_7_winter_2020_14_35 | 2009/11/24 17:58 | ORL Point Inventory (PTINV) | Created by control strategy | rhc |
| 7 | <input type="checkbox"/> | 2020ck1_15_35_ptipm | 2009/11/23 15:04 | ORL Point Inventory (PTINV) | Created by control strategy | rhc |
| 8 | <input type="checkbox"/> | ptnonipm_notkilns_hap2005ci_tox_ckilns1 | 2009/11/12 14:04 | ORL Point Inventory (PTINV) | Imported | bte |
| 9 | <input type="checkbox"/> | ptnonipm_neikilns_hap2005ci_tox_ckilns1 | 2009/11/12 14:03 | ORL Point Inventory (PTINV) | Imported | bte |
| 10 | <input type="checkbox"/> | ptnonipm_notkilns_xportfrac_cap2005ci_tox_ckilns1 | 2009/11/12 14:02 | ORL Point Inventory (PTINV) | Imported | bte |
| 11 | <input type="checkbox"/> | ptnonipm_neikilns_xportfrac_cap2005ci_tox_ckilns1 | 2009/11/12 14:01 | ORL Point Inventory (PTINV) | Imported | bte |
| 12 | <input type="checkbox"/> | Cntld_ptnonipm_xportfrac_cap2020ck1 | 2009/11/11 20:19 | ORL Point Inventory (PTINV) | Created by control strategy | delvecch |
| 13 | <input type="checkbox"/> | seca_c3_BAF_HAPs2015pf31_26oct2009 | 2009/10/26 09:19 | ORL Point Inventory (PTINV) | Imported | rhc |

247 rows : 12 columns: 0 Selected [Filter: None, Sort: Last Modified Date(-)]

View Edit Properties Edit Data Remove Import Export Purge Close

- Select a Dataset Type to see the available Datasets of that type
- The Manager table can be sorted and filtered based on contents

Dataset Properties Editor

The screenshot shows the 'Dataset Properties Editor' window for the dataset 'ptnonipm_xportfrac_cap2020ck1_14_35'. The window has several tabs: Summary, Data, Keywords, Notes, Revisions, History, Sources, and QA. The 'Summary' tab is active, displaying the following information:

- Name:** ptnonipm_xportfrac_cap2020ck1_14_35
- Description:**
 - #ORL POINT
 - #TYPE Point Inventory for CAPS
 - #COUNTRY US
 - #YEAR 2020
 - #DESC ANNUAL ptnonipm NOT including offshore oil (FIPS=85000 in original PQA file)
 - #DESC US (including AK and HI), PR, VI, and Tribal
- Project:** PM NAAQS RIA (/pm_naaqs_ria)
- Creator:** rhc
- Dataset Type:** ORL Point Inventory (PTINV)

Below this, there are two columns of metadata fields:

| | |
|--|---|
| Time Period Start: 01/01/2020 00:00 | Status: Created by control strategy |
| Time Period End: 12/31/2020 23:59 | Last Modified Date: 12/01/2009 13:23 |
| Temporal Resolution: Annual | Last Accessed Date: 12/01/2009 13:23 |
| Sector: ptnonipm | Creation Date: 12/01/2009 13:23 |
| Region: US | Intended Use: public |
| Country: US | Default Version: 0 (Initial Version) |

At the bottom of the window, there are four buttons: Refresh, Save, Export, and Close.

- Allows users to see and change metadata

Data Editor

Data Editor: ptnonipm_xportfrac_cap2020ck1_rerunCoST, Version: Test version, Table: DS_ptnonipm_xportfrac_cap2020ck1_rerunCoST_118... □ ↗ ✕

Saved any changes and applied Sort 'ANN_EMIS desc' and Filter 'POLL='CO' and ANN_EMIS>1000' Lock expires at 2010/04/27 14:21

Sort Order: Current: 1 - 300 Filtered: 392 of 1259262

Row Filter: Apply ⏪ ⏩ ⏪ ⏩

| SELECT | FIPS String(6) | PLANTID String(15) | POINTID String(15) | STACKID String(15) | SEGMENT String(15) | PLANT String(40) | SCC String(10) | POLL String(1) | ANN_EMIS Double | STKHGT Double | EF S |
|--------------------------|-------------------|-----------------------|-----------------------|-----------------------|-----------------------|----------------------------|-------------------|-------------------|--------------------|------------------|---------|
| <input type="checkbox"/> | 18127 | 00001 | 014 | 25 | 30 | ISGBURNSHARBORLLC | 30300813 | CO | 79993.69 | 79.0 | 01 |
| <input type="checkbox"/> | 39007 | 0204010193 | P002 | R9 | P002P1 | MILLENMIUMINORGANICCHEM... | 30107002 | CO | 49458.05 | 200.0 | 02 |
| <input type="checkbox"/> | 49023 | 10303 | 19593 | 1616 | 1 | LeamingtonCementPlant | 31499999 | CO | 34039.616 | 98.0 | 02 |
| <input type="checkbox"/> | 18089 | 00121 | 101 | 0 | 02 | USSGARYWORKS | 30300813 | CO | 33212.83 | 100.74 | 02 |
| <input type="checkbox"/> | 24005 | 005-0147 | 31 | S31 | 02 | ISGSPARROWSPPOINTLLC | 30300819 | CO | 29971.323 | 100.0 | 02 |
| <input type="checkbox"/> | 18127 | 00001 | 015 | 62 | 11 | ISGBURNSHARBORLLC | 30300915 | CO | 29422.98 | 82.0 | 01 |
| <input type="checkbox"/> | 39155 | 0278000463 | P902 | R22 | P902P1 | WCISTEEL, INC. | 30300913 | CO | 27174.94 | 160.0 | 02 |

Revision Information (auto-saved when window is closed)

| What was changed | Why it was changed | References |
|----------------------|----------------------|----------------------|
| <input type="text"/> | <input type="text"/> | <input type="text"/> |

- Allows users to sort, filter, and edit data

- When saving, must say what was changed and why

QA Management

- Summaries of data can be created and saved using enhanced Structured Query Language
 - By county, by plant, by emissions category, ...
- Reference database adds descriptive columns
 - County names, state abbreviations, code descriptions
- Complex QA functions are coded in Java and can be chosen from a list of 'QA programs'
 - Comparing values of two datasets
 - Summing monthly inventory data to annual
- Results can be viewed in a sort-filter table
- In some cases results can be viewed using Google Earth or exported to GIS Shapefiles

Adding a QA Step from a Template

The screenshot shows the 'Dataset Properties Editor' window for the dataset 'ptnonipm_xportfrac_cap2020ck1_rerunCoST'. The 'QA' tab is active, displaying a table of QA steps. An 'Add QA Steps' dialog box is open, allowing the user to select a template for a new QA step.

| # | Select | Name | Version | Required | Order | QA Status | Run |
|---|--------------------------|--|---------|--------------------------|-------|-------------|-----|
| 1 | <input type="checkbox"/> | Summarize by Pollutant with Descriptions | 0 | <input type="checkbox"/> | 1.0 | Not Started | |
| 2 | <input type="checkbox"/> | Summarize by State and Pollutant | 0 | <input type="checkbox"/> | 1.0 | Not Started | |

Add QA Steps: ptnonipm_xportfrac_cap2020ck1_rerunCoST

Version: Initial Version (0) ▼

Required:

Optional:

- Summarize by Mact Code, U.S. State and Pollutant with Descriptions
- Summarize by Plant and Pollutant
- Summarize by Pollutant
- Summarize by Pollutant with Descriptions

Buttons: OK, Cancel

Bottom Bar: Add from Template (circled), Add Custom, Edit, Copy, Set Status, Run, View Results

Footer: Refresh, Save, Export, Close

Viewing QA Step Results

View QA Step Results: Summarize by U.S. State and Pollutant with Descriptions2

File

QA_DSID80955_V0_20100430135839e_and_Poll

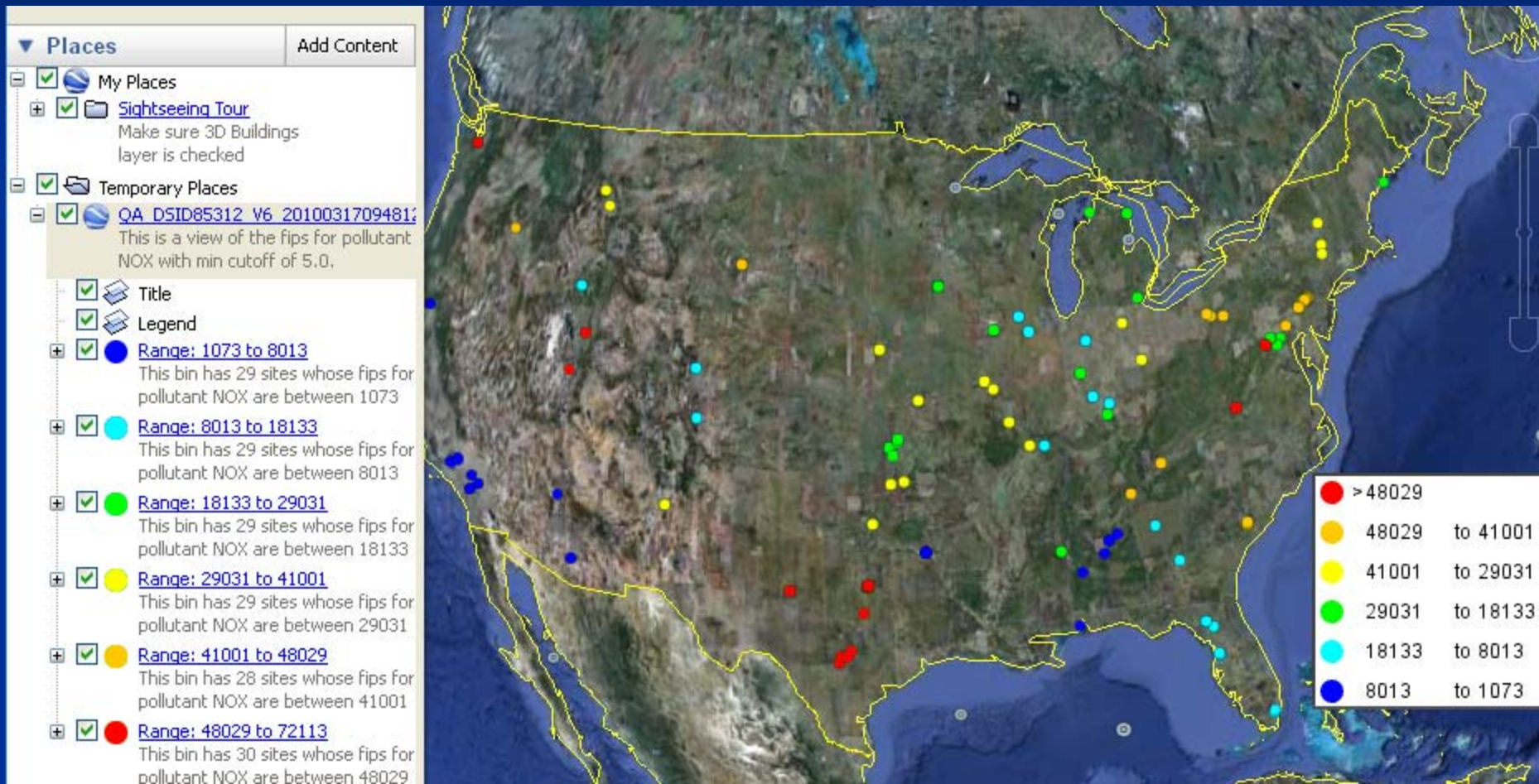
File Name C:\DOCUME~1\aeyth\LOCALS~1\Temp\QA_DSID80955_V0_20100430135839e_and_Pollutant_with_Descriptions.csv

Header #DATASET_NAME=ptnonipm_xportfrac_cap2020ck1_rerunCoST #DATASET_VERSION_NUM= 0 #CREAT Full Description

| | fipsst | state_abbr | state_name | poll | longitude | latitude | ann_emis | pollutant_cod |
|----|--------|------------|------------|------|-----------|----------|------------|-------------------------|
| 1 | 18 | IN | Indiana | CO | -86.44 | 39.76 | 351,101.56 | Carbon Monoxide |
| 2 | 48 | TX | Texas | NOX | -100.08 | 31.16 | 283,203.44 | Nitrogen oxides |
| 3 | 48 | TX | Texas | CO | -100.08 | 31.16 | 260,160.33 | Carbon Monoxide |
| 4 | 39 | OH | Ohio | CO | -82.66 | 40.19 | 230,244.25 | Carbon Monoxide |
| 5 | 01 | AL | Alabama | CO | -86.68 | 32.60 | 166,268.18 | Carbon Monoxide |
| 6 | 22 | LA | Louisiana | NOX | -91.42 | 30.98 | 161,836.15 | Nitrogen oxides |
| 7 | 48 | TX | Texas | SO2 | -100.08 | 31.16 | 155,792.19 | Sulfur Dioxide |
| 8 | 22 | LA | Louisiana | SO2 | -91.42 | 30.98 | 151,223.49 | Sulfur Dioxide |
| 9 | 22 | LA | Louisiana | CO | -91.42 | 30.98 | 144,596.34 | Carbon Monoxide |
| 10 | 06 | CA | California | CO | -119.26 | 37.26 | 136,023.82 | Carbon Monoxide |
| 11 | 48 | TX | Texas | VOC | -100.08 | 31.16 | 134,793.46 | Volatile organic compou |
| 12 | 17 | IL | Illinois | SO2 | -89.50 | 39.74 | 133,181.30 | Sulfur Dioxide |
| 13 | 24 | MD | Maryland | CO | -77.26 | 38.81 | 108,217.51 | Carbon Monoxide |
| 14 | 12 | FL | Florida | CO | -83.82 | 27.77 | 107,277.89 | Carbon Monoxide |
| 15 | 47 | TN | Tennessee | CO | -85.98 | 35.82 | 106,934.89 | Carbon Monoxide |
| 16 | 39 | OH | Ohio | SO2 | -82.66 | 40.19 | 105,138.78 | Sulfur Dioxide |

383 rows : 13 columns [Filter: None, Sort: ann_emis(-)]

Plant Summary of NO_x (NO+NO₂) Emissions

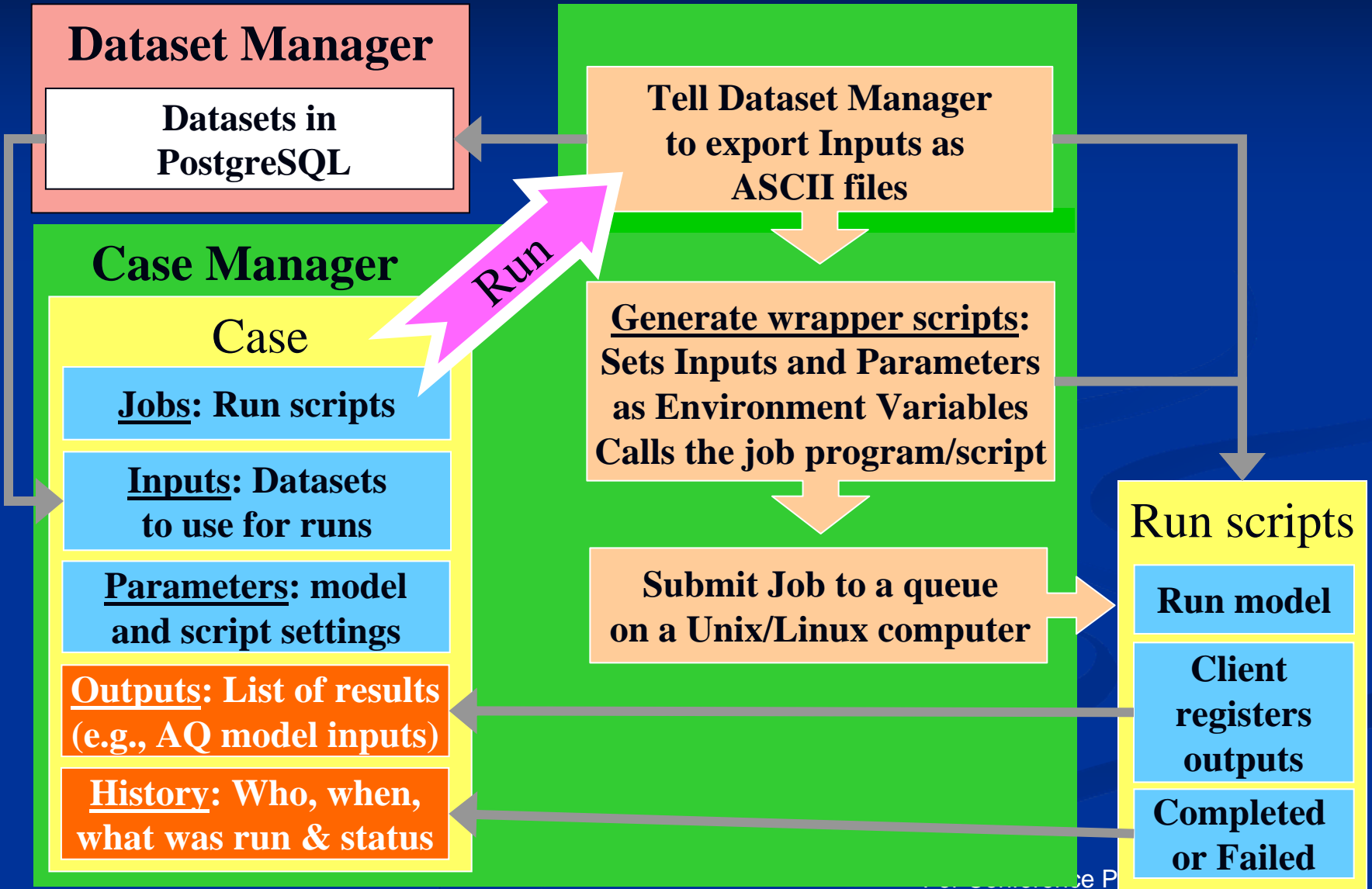


- Results with latitude/longitude can be exported to a KMZ file and viewed using Google Earth

Case Management

- A Case stores information about model runs
 - Summary attributes (grid, time period, sectors)
 - Jobs to run (individual programs or scripts)
 - Datasets used as inputs by the jobs
 - Parameters needed to control the jobs
 - Outputs from jobs
 - History and results of the jobs
- Has all information needed to run models and track their results
- When a job is run, EMF writes a C-shell script that:
 - Specifies inputs and parameters as environment variables
 - Runs the job's program or script (using any script language)
 - Reports back to the server with run status information
- Job scripts can use a Java command-line client to communicate to the server about status and outputs

Case Management Overview



Case Manager

Case Manager

Refresh

Show Cases of Category: Run SMOKE base year

Name Contains:

Icons:

| # | Select | Name | Project | Model to ... | Modeling Re... | Last Modified ... | Run SMOKE |
|----|-------------------------------------|--|-------------------------------|--------------|----------------|-------------------|-----------|
| 36 | <input type="checkbox"/> | 2006af: CAP-HAP_SAPRC07 | ORD_B10_2006_SAPRC07_C... | SMOKE | National | Allan Beidler | Run SMOKE |
| 37 | <input type="checkbox"/> | 2006af: CAP-HAP_SAPRC07 PAH | ORD_B10_2006_SAPRC07_C... | SMOKE | National | James Beidler | Run SMOKE |
| 38 | <input type="checkbox"/> | 2006af: CAP-HAP_SAPRC07 with PAH/Arsenic | ORD_B10_2006_SAPRC07_C... | SMOKE | National | Chris Allen | Run SMOKE |
| 39 | <input checked="" type="checkbox"/> | 2006am: CAP-CI, CB05_SOA INLINE AQMEII | ORD AQMEII | SMOKE | National | Ryan Cleary | Run SMOKE |
| 40 | <input type="checkbox"/> | 2006am: CAP-CI, CB05_SOA INLINE AQMEII - ... | | SMOKE | National | Ryan Cleary | Run SMOKE |
| 41 | <input type="checkbox"/> | 2006am: CAP-CI, RACM2 INLINE AQMEII | ORD AQMEII | SMOKE | National | James Beidler | Run SMOKE |
| 42 | <input type="checkbox"/> | 2006am: CAP-CI, SAPRC07 INLINE AQMEII | ORD AQMEII | SMOKE | National | Ryan Cleary | Run SMOKE |
| 43 | <input type="checkbox"/> | 2006cm_06b | Ozone Designations (/o3desig) | SMOKE | National | Chris Allen | Run SMOKE |
| 44 | <input type="checkbox"/> | Copy of 2006af: CAP-HAP_SAPRC07 | | SMOKE | National | William Benjey | Run SMOKE |
| 45 | <input type="checkbox"/> | Testing for B73 AQMEII | ORD AQMEII | SMOKE | National | Charles Chang | Run SMOKE |

45 rows : 19 columns: 1 Selected [Filter: None, Sort: Name(+)]

Buttons: View, Edit, New, Remove, Copy, Sensitivity, Import, Close

- Cases are organized by Category
- The Case Manager shows available cases

Case Editor

Case Editor: 2006am: CAP-CI, CB05_SOA INLINE AQMEII

Summary Jobs Inputs Parameters Outputs History

Name: 2006am: CAP-CI, CB05_SOA INLINE AQMEII Abbreviation: B73_2006am

Category: Run SMOKE base year Is Final: Is Template:

Description: Annual 2006 Emissions Processing for AQME11. 12US1 domain with surrogates developed specifically for this project. CB05SOA speciation mechanism. 2006 CEM and 2006 offsite inventory included. Sectors: afdust ag

Project: ORD AQMEII Copied From: 2005cm_05b

Run Status: Complete Last Modified By: Ryan Cleary on 04/13/2010 12:30

Model & Version: SMOKE 2.4 Downstream Model: CMAQ v4.7 N1c

Modeling Region: National Speciation: cmaq_cb05_soa

Regions: Continental US 12km (12US1) Meteorological Year: 2006

Met/Emis Layers: 34 1 Base Year: 2006

Start Date & Time: 01/01/2006 00:00 Future Year: 2006

End Date & Time: 12/31/2006 23:59

Refresh Load Export Save View Parent View Related Close

Other EMF Components

- Control Program Manager
 - Stores information how existing regulations affect sources
- Control Strategy Manager
 - Uses emissions control measures and control programs to generate controlled and future year emission inventories
- Sector Manager
 - Subsets and summarizes inventories by industrial sector
 - Splits broader inventories to sector-specific ones
- Speciation Tool
 - Creates profiles to split inventory pollutants to model species
- Surrogate Tool
 - Creates spatial surrogates to grid county-level emissions

Summary of EMF Benefits

- Successfully **used in production** for several years: thousands of datasets, hundreds of cases
- Data and run specifications are **easily shared**
- **Multiple versions** of datasets are stored
- **Quality assurance** of the data is easier
- **Cases (runs) are stored perpetually**
 - Settings can be reviewed at any time by anyone
 - Data can be recreated at a later date
- **Easier to set up a new case** by copying and adjusting an existing one
- Facilitates **collaboration and reuse** of data and run settings (e.g., control strategy inventories)
- Many **parts of EMF are generic** so that it could be extended to run other models

EMF Availability

- Designed to be open source
- Available documentation is limited
- Complexity of system and DB set-up limits our ability to distribute it widely
- Funding for long-term user support has not yet been identified
- A beta testing program to test feasibility of a true public release is just wrapping up
- Development will continue next year

Possibility for Applications other than Emissions Modeling

- EMF already used for some AQM-related tools other than the core emissions model (SMOKE)
- Interested in extensions to related topics: benefits analysis, economic analysis, ...
- Comma-separated value (.csv) dataset types allow for non-emissions data to be imported
- Recently implemented **flexible file formats** allow most kinds of tabular data to be managed by EMF
- QA templates could be developed for new formats
- Cases could be created to run other models
- Supporting other models will require some changes:
 - Filtering available lists so they are specific each model
 - Generalizing metadata on some editors

Remaining Challenges

- Current implementation uses web-services, but it is slow to transfer a lot of objects at once
 - Are there other tools we might consider using for client-server communication? (Currently using Axis)
 - Considering a web-based GUI in addition to the fat client
- A lot of data is spatial in nature (e.g., points, counties)
 - More mapping tools would be helpful (e.g., filled counties)
 - PostGIS add-on to PostgreSQL provides spatial functions, but not visualization
 - Consider whether to visualize data on server or bring to the client machine (.kmz is created on client machine)
- Broader user-base requires more sophisticated treatment of permissions (e.g., user/group/world; user roles)
- Identifying potential collaborators for extensions