Exploring Synergies in Air Quality Data Systems:

The Visibility Information Exchange Web System (VIEWS) and the Community Modeling and Analysis System (CMAS)

Uma Shankar, Shawn McClure and Alexis Zubrow

Air Quality Community Summer 2009 Meeting
Santa Barbara, CA, July 6-9, 2009

A collaboration between UNC-Institute for the Environment, the Cooperative Institute for Research in the Atmosphere and the University of Maryland Baltimore County
NASA Grant # NNX08AL28G

NASA ROSES Project Overview

Title:

Improving an Air Quality Decision Support System through the Integration of Satellite Data with Ground-based Observations, Modeling Results, and Emissions Estimates *(Co-PIs: Shankar, McClure; awarded in 2008; work ongoing through 2011)*

Goals:

•Provide for the analysis and visualization of satellite data in combination with monitoring, modeling, and emissions data

•Develop routine capture, analysis, and processing algorithms with high temporal and spatial resolution in order to provide better land use/land cover model inputs

•Obtain finer temporal and spatial resolution of emissions data in remote areas and from individual sources and source clusters

•Improve boundary and other modeling inputs with 3- to 4-D pollutant data

•Evaluate gridded chemistry-transport models, such as CMAQ through the synergistic use of ground-based and satellite data, and A&V tools

What user needs are we focused on?

- Simplifying data discovery, access, and analysis to encourage non-expert use, improve efficiency in AQ decision support
- Consolidating relevant datasets with an integrated set of analysis tools
- Providing useful, core metadata while minimizing metadata "overload"
- Suggesting appropriate applications for data
- Facilitating intercomparisons of model and observational datasets
- Providing expert interpretations of data
  - Example: RoMANS 2006 study simulations with CMAQ to diagnose modeled ammonia deposition in Rocky Mountain National Park (for NPS)
  - Use of CMAS Atmospheric Model Evaluation Tool (AMET) for routine model evaluation against ground-based and satellite data
  - Use of python-based Process Analysis (py-PA) for diagnostic evaluation
  - Development of advanced visualization/animation capabilities in VIEWS to analyze aerosol size distributions (observed and modeled)

Collaborations and Community Activities

EPA Data Summit (RTP - 2008):

- Convened groups with key roles in managing, analyzing, and disseminating earth observation data in order to explore efficient means of leveraging existing efforts
- Explored mechanisms and potential opportunities for “interoperability” between systems
- Assisted EPA OAQPS in honing its role in the larger earth observation community
- Established a community-wide strategy for responding to user-defined needs

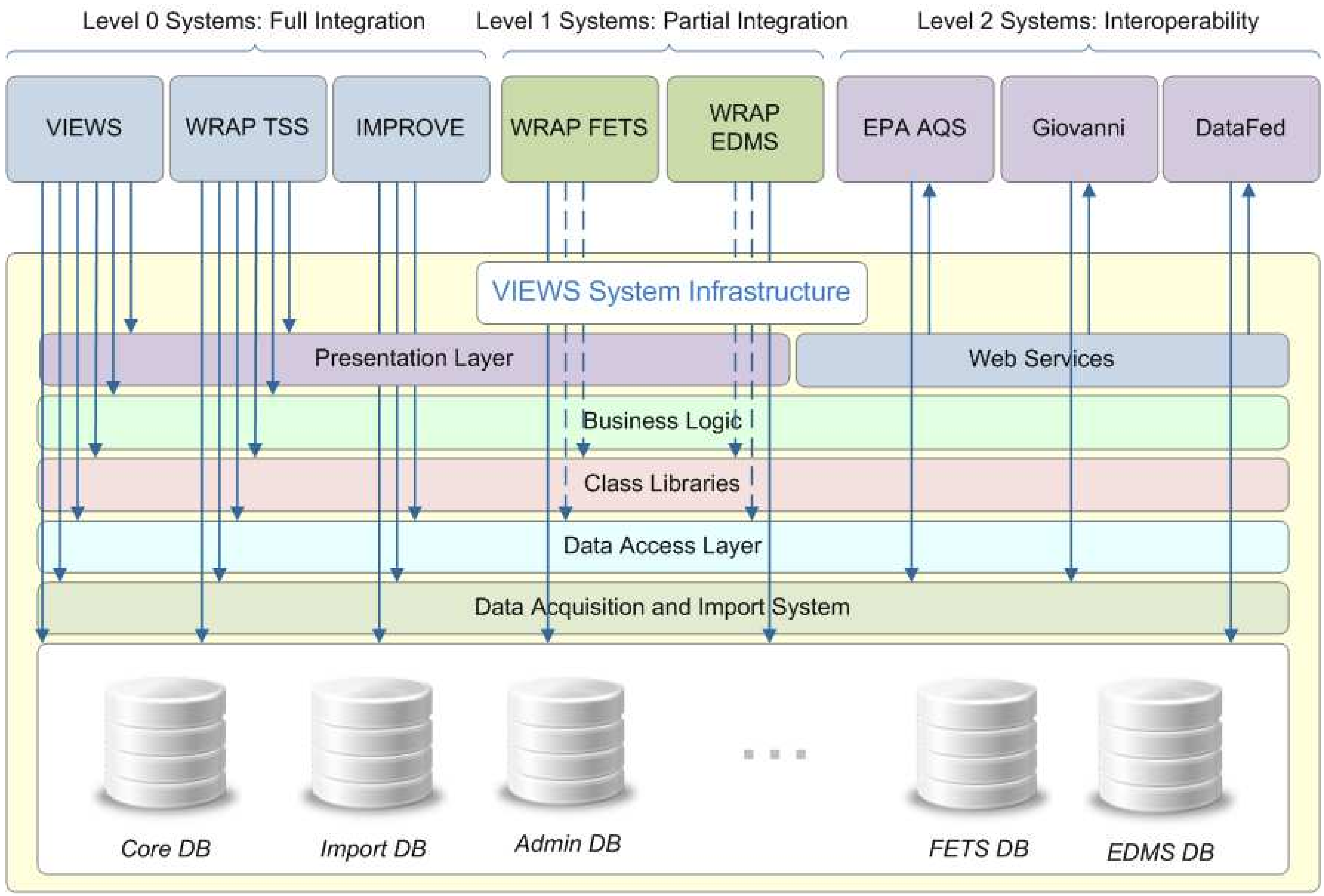
Earth Science Information Partners (ESIP) Federation

- Collaborate on interoperability efforts for data/metadata discovery and exchange
- Contribute data and services to the Earth Information Exchange

GEO/GEOSS Architecture Implementation Pilot (AIP) 2

- Contribute components and services to the GEOSS registry
- Participate in efforts to standardize and streamline data discovery and exchange
- Develop VIEWS/TSS as an “persistent operational exemplar” of the GEOSS architecture

Interoperability Levels



Leveraging the VIEWS Architecture

First question we ask ourselves when designing a new feature:

"Can its design be generalized for VIEWS while providing for customization and domain-specific use by connected systems?"

If "yes", then the benefits are:

- Project and configuration management are simplified
- Dependent systems can be more "lightweight"
- The system as a whole can be extended more easily
- Past and current investments are leveraged and maximized
- Interoperability and data sharing with external systems is made easier
- Developers can "implement once, reuse often"
- Collaboration is facilitated and expedited

Project activities so far

- Conducted a design workshop May 2008 to engage end users and refine scope
- Acquired the following datasets:
  - OMI (Aura) AOD: L2G, L3
  - MODIS (Aqua and Terra) AOD: L3
  - MODIS (Aqua and Terra) AOD on 12 km x12 km CMAQ ConUS NCEP grid
  - CALIPSO L2 extinction on the above grid forthcoming
- Explored available data structure and organization; identified parameters to extract
- Determined bandwidth and speed limits for data download operations
- Estimated storage requirements and data download frequencies
- Created a prototype visualization tool: http://vista.cira.colostate.edu/nasa
- Created a project website and wiki: http://vista.cira.colostate.edu/airdatawiki
- Formulated ideas for portal and access to new analysis tools (AMET, pyPA, vis)
- Began processing CMAQ input data for RoMANS simulations for 2006
- Conducted a pilot training course on satellite data use in AQ applications (Prados)

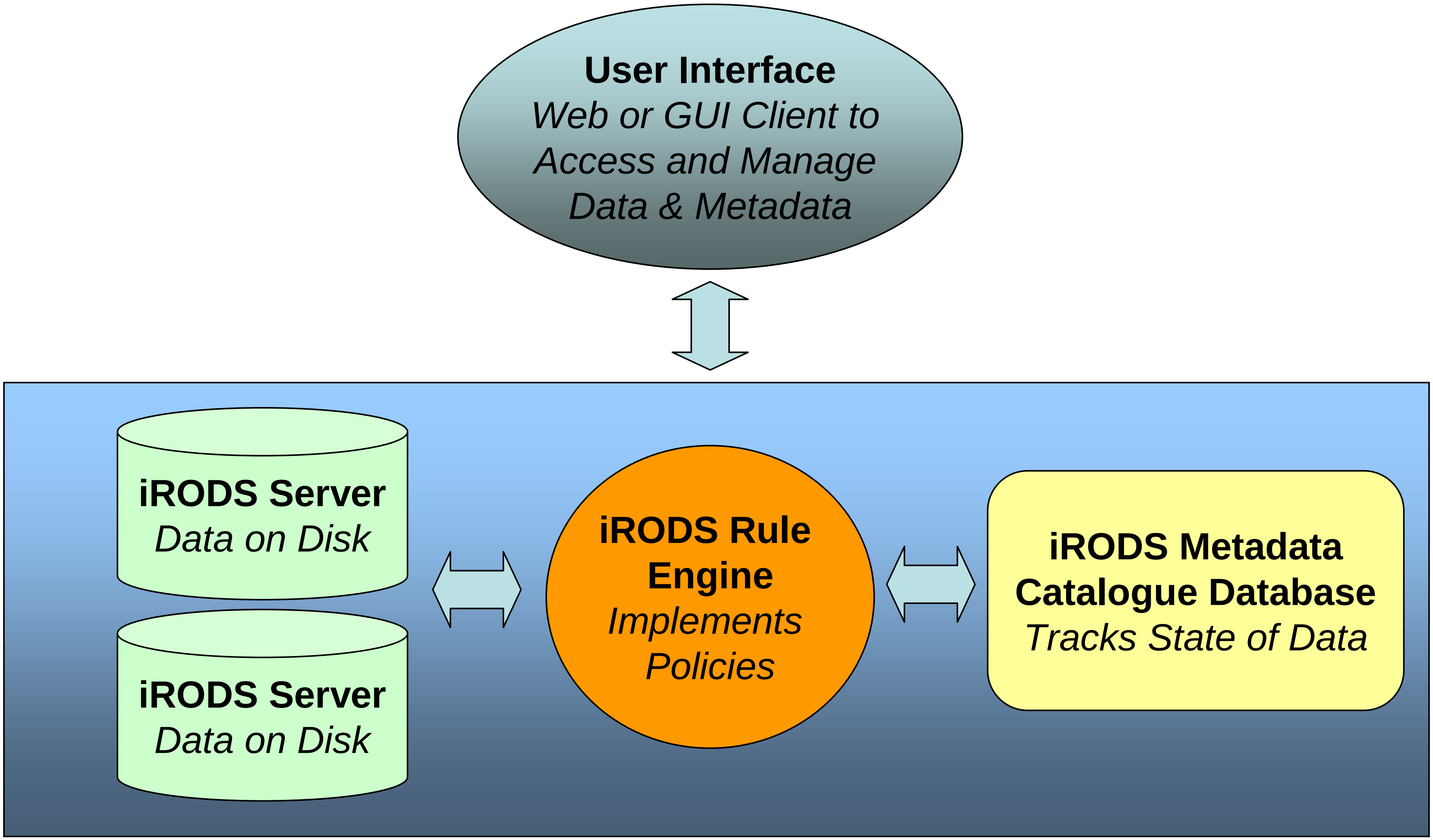
What services will we offer?

- Raw data and metadata retrieval (e.g., download all IMPROVE data for 2007)
- Geospatial data retrieval (e.g., retrieve a map of interpolated NO_2 from OMI)
- Data coincidence discovery (e.g., show me all the fires within 100 km of a site)
- On-the-fly transformation and formatting of data and metadata
- Upload and management of user-supplied data and metadata
- Generation of clickable geospatial visualizations and analysis products from observed as well as modeled data
- Developer access to “embeddable” components, data feeds, and tools

CMAS is enabling CMAQ model data service via iRODS

- Through a CMAS Work Assignment UNC–IE is evaluating the Integrated Rule-Oriented Data System (iRODS) for serving data
  - Means for VIEWS to track and access a diverse range of data products from both local and remote archived locations (CMAQ model, satellite, monitor, etc.)
    - Currently experimenting with model data only
  - http://www.irods.org
  - Middleware for distributed data access
  - Masks interactions with a diverse data serving infrastructure
  - Users see a single virtual archive
  - Data grid / data cloud

iRODS Paradigm



iRODS features

Pros

- User can ignore location of data
- Data replication/archiving
- Weighting of data servers - giving greater precedence to one over another
- User metadata can be associated with each file
- Scriptable icommands (command-line functions)
- Specific rules can be created to customize upload/download of (meta)data
- Can chain together iRODS functions and external programs
- Very flexible

Cons

- Metadata search capabilities are immature
- Metadata hierarchy not supported
- iRODS API's not mature; missing some functionality
- No current functions (microservices) for our specific applications; would have to develop functions to subset, regrid, etc.
- Possibly too flexible

What services could we use?

- Data "re-gridding" (two-way; on-the-fly or asynchronous transformation from one grid resolution to another)
  - Want flexibility in specification of domain and grid resolution
  - Adapting an existing GIS-based tool to rasterize and project model output to the satellite data grid
  - Comparing with existing approaches, e.g., Remote Sensing Information Gateway (RSIG); 3D-AQS (12-km res, ConUS)

- "Retrieve" metrics from AQ model outputs for one-to-one comparison with satellite-derived parameters, e.g., AOD, tropospheric NO_2 column
  - Will post-process model output in the near term, but would likely transition to a web service in VIEWS

- Selection/ subsetting of individual parameters from satellite data products for the desired resolution and domain for *all* levels of data

- Easier access to metadata regarding caveats for AQ applications, or development of such guidance
  - CMAS training course (NASA sponsored) will help provide some of this

What common infrastructure is most needed?

- A consolidated, uniform understanding and description of the many levels of the metadata hierarchy (see “Barriers” question)
  - One person’s metadata may be another person’s *data,* or vice versa.
  - Need a common infrastructure for identifying, describing, and structuring these many-tiered metadata more coherently
  - Need this especially for data caveats in comparisons w/ model output
- A more thorough, community-wide understanding of data exchange standards (such as the OGC standards):
  - The mere fact that a system supports a certain set of standards is not a guarantee that it supports them *usefully*
- A minimum level of maintenance funding to support ongoing participation in efforts like GEOSS (see “Barriers” question)
- Archiving observational, model, and derived products

What are the barriers to this common infrastructure?

- Participant organizations in GEOSS still seem to be working out a common vocabulary and requirements for GEOSS
  - Widely-varying needs and opinions about what constitutes appropriate “metadata” complicate standardization efforts
  - Model metadata lack a standard for succinctly describing the configuration used, e.g., transport schemes, chemical mechanisms, convective cloud parameterizations - not typically found in output file headers!
- Little strategic oversight: unclear who would provide this, or how effort should be divided among the participant organizations in GEOSS to avoid duplication
  - Well-intentioned projects often turn into “one-off” efforts that fail to leverage or build upon previous work
- Developer’s Dilemma: Deliverables to sponsors of funded projects take precedence over *pro bono* collaborations such as in GEOSS, so momentum is difficult to maintain even when interest is high