Health & Ecoinformatics

William Sonntag
Office of Environmental Information
U.S. Environmental Protection Agency
March 24, 2004

Presentation Overview

- EPA Mission and Needs for Health Data
- EPA Health & Ecoinformatics
  - Enterprise Architecture
  - Environmental Data Standards Council
  - EPA-State Environmental Exchange Network
  - EPA 2003 Draft Report on the Environment
  - Selected Partnerships
    - Department of Health and Human Services MOU
    - Department of Energy MOU
    - Group on Earth Observations

EPA Mission

- The mission of the U.S. Environmental Protection Agency is to protect human health and safeguard the natural environment--air, water, and land--upon which life depends.
- EPA's purpose is to ensure that:
  - All Americans are protected from significant risks to human health and the environment where they live, learn and work.
  - National efforts to reduce environmental risk are based on the best available scientific information.
  - Etc.

Why Does EPA Need Health Data?

- EPA has a multitude of responsibilities such as:
  - Developing and enforcing standards to ensure
    - clean air,
    - clean water, and
    - protection from toxic chemicals and hazardous materials
  - Protecting land resources, and
  - Coordinating emergency responses
- These responsibilities are carried out to protect the health of the general public/susceptible populations and the environment

What does EPA do with health data?

- Identify health risks associated with contaminants
- Conduct dose response assessments
  - Develop tolerance levels for exposures
  - Set standards for pollutants
  - Establish water quality criteria and drinking water standards
- Set priorities for policy decision-making
- Evaluate public health impacts of environmental decisions
- Educate the public on environmental risks
- Report on health effects of environmental conditions
- Describe health trends in response to environmental conditions

What health data does EPA need?

- Data of known quality and available:
  - At multiple scales – national, state, site specific
  - Over multiple years
  - For specific susceptible and general populations

What health data does EPA need?

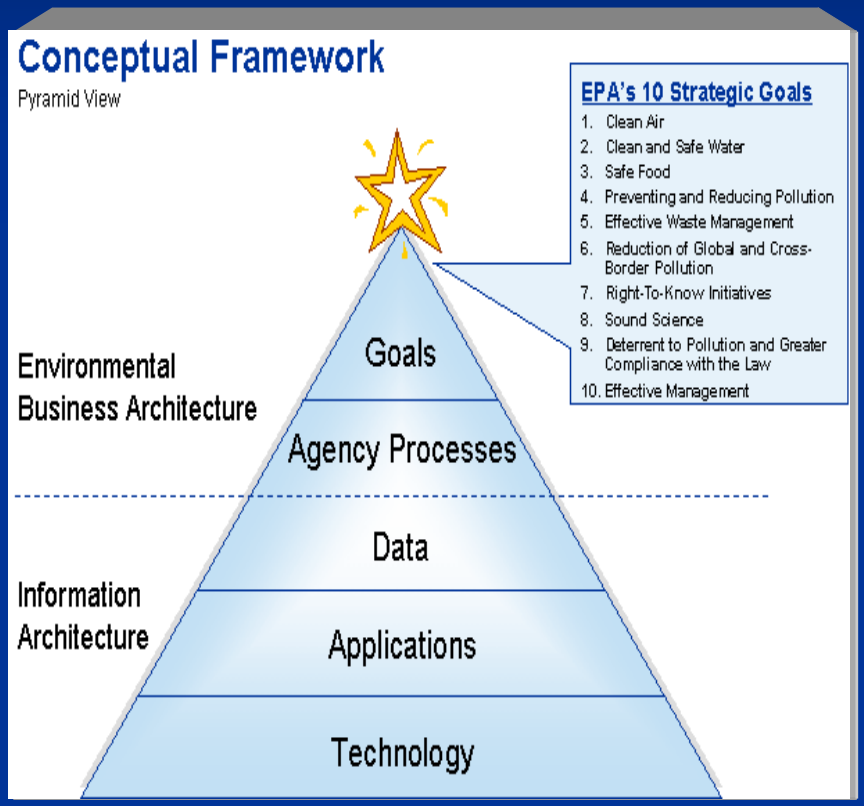
- Specific types of health data
  - Morbidity data for prevalence and incidence of specific diseases
  - Mortality data
  - Exposure data
  - Biomarkers of exposure and effects

What are the issues with current health data?

- Challenges finding and accessing data
- Limited availability of geographically referenced health data for integration
- Limited availability of morbidity data
- Limited availability of exposure data

4 Components of Ecoinformatics

- Knowledge Management Tools – Applications Web Services & Data Arrays (Draft EPA ROE)
- Metadata Management & Repositories (SRS & ISO 11179 Activities)
- Data Standards (EDSC)
- Exchange Formats – (Environmental Information Exchange Network)



EPA Enterprise Architecture

Target Architecture Framework

EXCHANGE NETWORKS

Public

Industry

Non-government Partners

Government Partners

EPA Users

Intranet

Security and Access Controls

Extranet

USE

System of Access

Program Support
Public Access
Decision Support

STORE for USE

Enterprise Repository

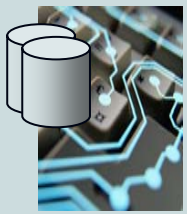
Metadata Holdings
Catalog
Shared Geospatial Data
Central Registries
Data Warehouse

CONNECT and EXCHANGE

CDX Services

PROCESS and STAGE

Operational Databases & Applications



Management Practices
(Architecture, Policies, Standards, Security)

Three Architectures

- Environmental protection
- Research (R&D)
- Administrative

Environmental Data Standards Council (EDSC)

EDSC Approved Standards

- The EDSC has approved standard data elements for the following:
  - Biological Taxonomy
  - Chemical Identification
  - Contact Information
  - Date
  - Enforcement/Compliance
  - Facility Identification
  - Federal Facility Identification

EDSC Approved Standards (con't)

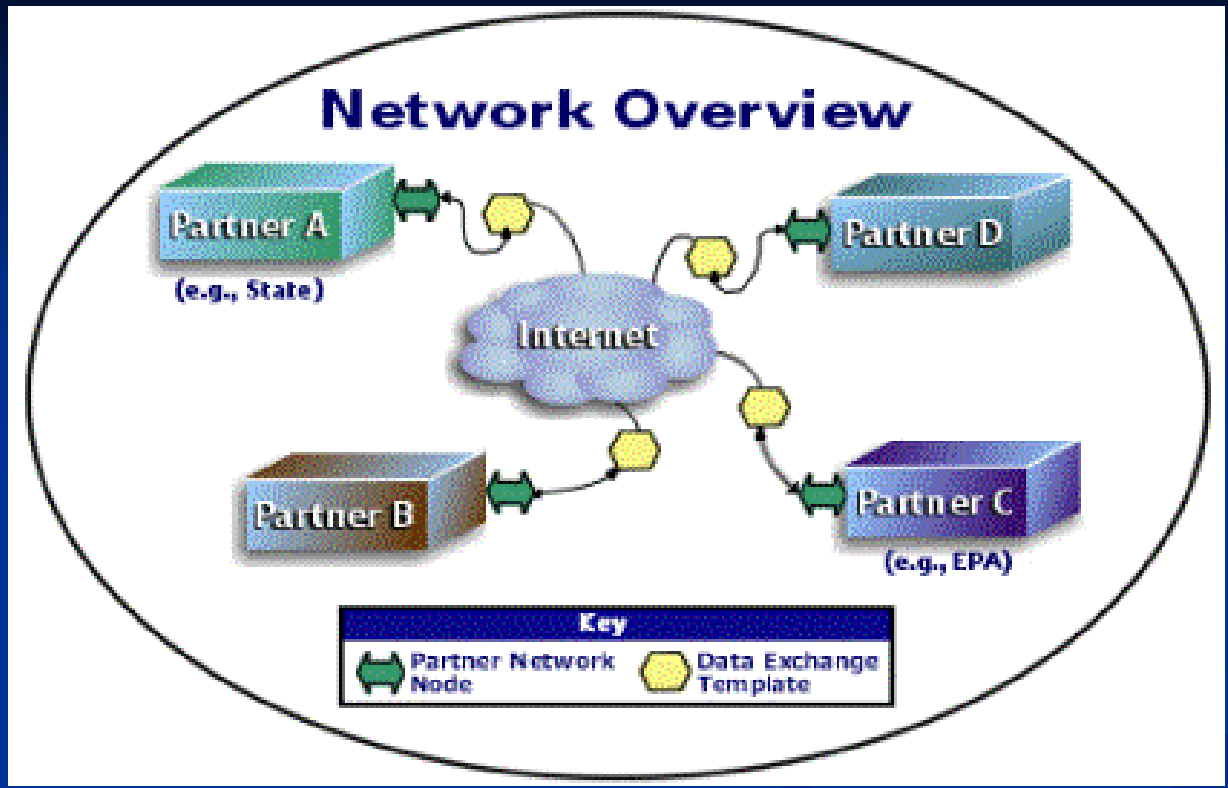
- Latitude/Longitude
- Permitting Information
- Reporting Water Quality Results for Chemical and Microbiological Analytes
- SIC/NAICS
- Tribal Identifier

- The EDSC has developed final XML tags for some of the EDSC approved data standards. Draft XML tags are available for review.

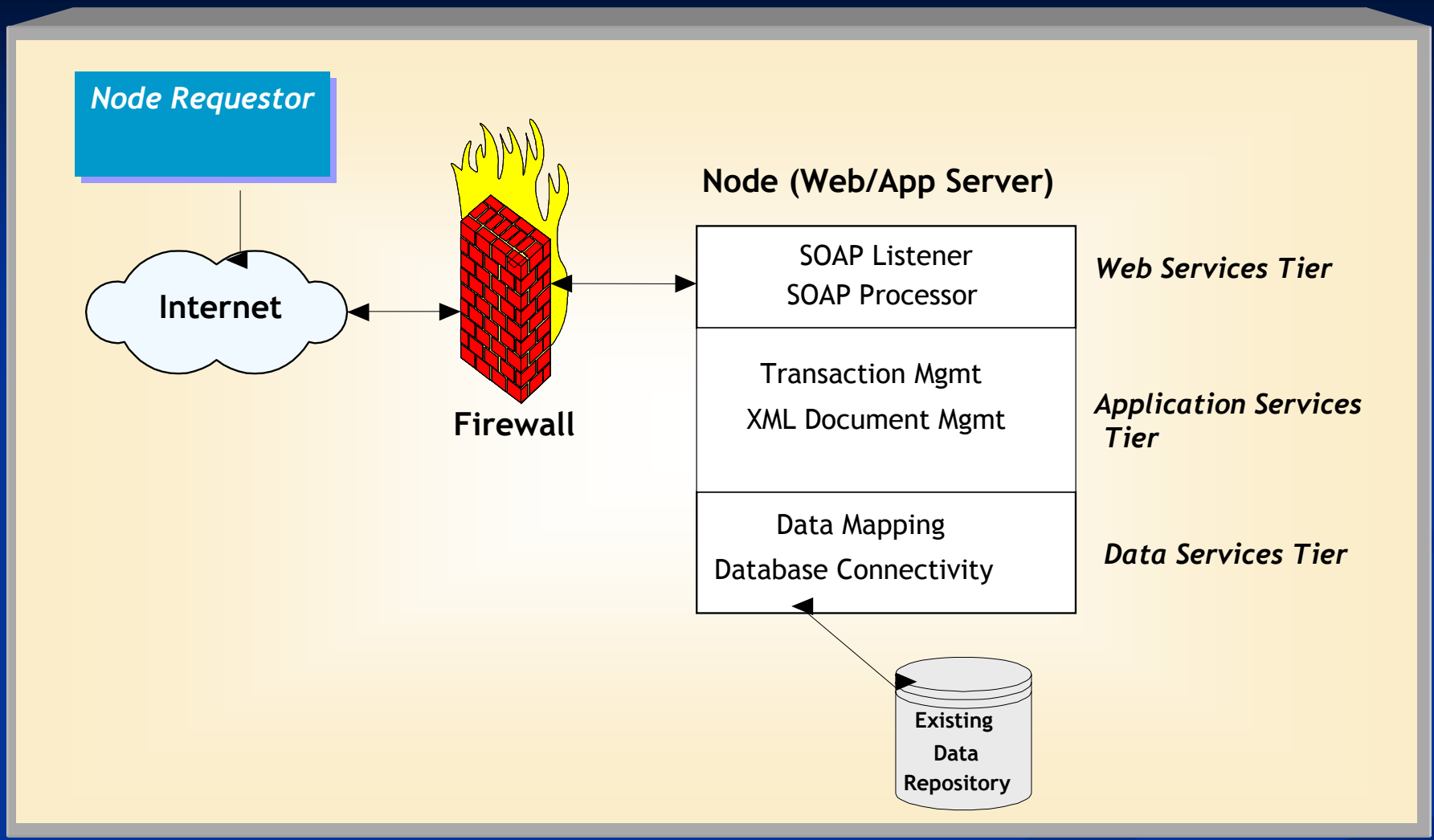
EDSC Standards Under Development

- EDSC Action Teams are developing standard data elements for the following:
  - Attached Binary Object
  - Environmental Sampling, Analysis, and Results
  - Measure
  - Method
  - Representation of Date and Time
  - Sample Treatment

The Exchange Network



The Exchange Network is an Internet and standards based method for exchanging environmental information among partners



Exchange Network Node Components

Progress to Date

- Data Transport
  - Revised Node Specification and Protocol to Version 1.1
  - Node code available for wide range of platforms
  - Standardized test bed established for conformance
- Information Structure/Semantics
  - Ongoing work to establish semantic standards through Environmental Data Standards Council (e.g., permits/ licenses, environmental sampling and results)
  - Revising Core Reference Model (CRM) of common data blocks
  - Establishing "XML Architecture" to support re-use and harmonization of XML payloads, using the CRM

Performance Measures

- Focus on building infrastructure and exchanging data
- Goal is to create critical mass by end of 2004
- Measures with Specific Targets
  - #1: Data flowing to National Systems
  - #2: States capable of multiple flows
  - #3: Technical architecture with fully functional nodes

Progress on Performance Measures

- **Performance Measure #1 – data flow to EPA National Systems with no diminution of services**

Type	Schema Available	Goal - # of States	**Progress**
FRS	Yes	20	
Beaches	Yes	7	
NEI	Yes	12	
PCS/IDEF	Yes	10	
RCRAInfo	No	10	
SDWIS	No	10	

Progress on Performance Measures

- **Performance Measure #1 – data flow to EPA National Systems with no diminution of services**

Type	Schema Available	Goal - # of States	Progress
FRS	Yes	20	6
Beaches	Yes	7	2
NEI	Yes	12	Due in Spring
PCS/IDEF	Yes	10	1
RCRAInfo	No	10	Not available yet
SDWIS	No	10	Not available yet

Performance Measure - #2

- States are capable of multiple flows (two or more) to partners
  - Target: 10-14 States doing multiple flows by end of 2004
  - Progress: None to date, but many close

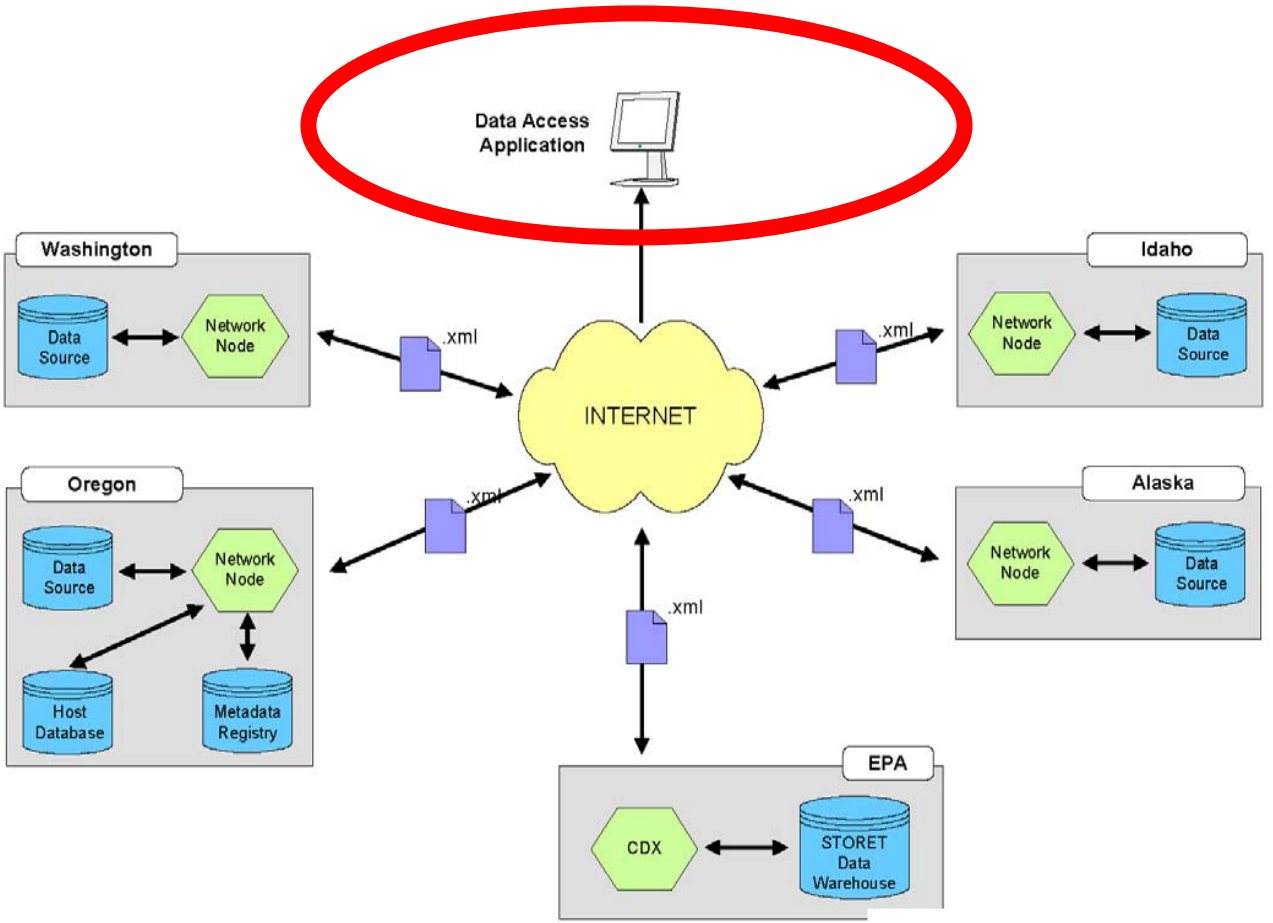
Performance Measures – #3

- The technical architecture is in place with partners having established fully functional nodes
  - Target: 35 Nodes by the end of 2004
  - Progress
    - 7 Operational and exchanging data
    - 9-11 Testing
    - 10 Building
    - 14 Planning

It's Web Services Just Like Expedia.com

- Data flows from decentralized sources
- Customized requests by users based on needs
- Infinite combinations of data and data arrays as long as its exposed to web services queries
- After all the airlines still hold their flight data in two main systems

Pacific Northwest Water Quality Data Exchange



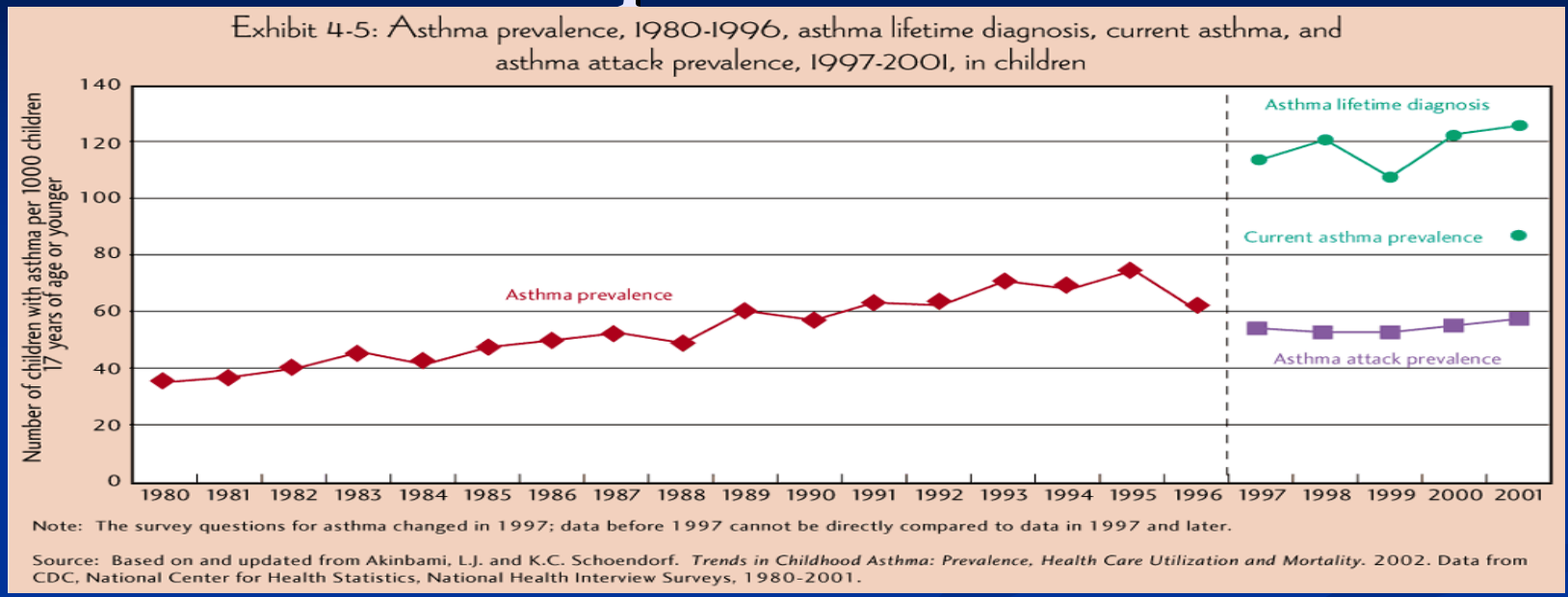
2003 Draft Report on the Environment

Purpose for the Draft Report on the Environment

- Identify indicators – measures of environmental results – to describe status and trends in the environment and human health

- Describe what EPA knows – and doesn‘t know – about the current state of the environment at the national level and how it is changing

Example Indicator



- Although air quality has improved at a national level, areas such as inner cities, continue to experience intermittent exposure to poorer air quality, which may contribute to asthma prevalence.
- Researchers believe that air pollutants may increase the severity or frequency of asthma attacks in children who have the disease.

Next Steps for the Report on the Environment

- Examine questions and indicators for improvements
- Examine and fill gaps in indicators and data
- Engage partners
- Generate additional products

Partnership: Department of Health and Human Services

EPA-HHS MOU

- Collaboration between CDC and EPA on Environmental Health Tracking (network coordination)
- Cooperative Projects
  - HELIX – Atlanta Area EPHT
  - Air Quality Data Improvement and EPHT
  - State of WI and EPA Region 5 – Technical Assistance
  - NY State Project – Integrating networks

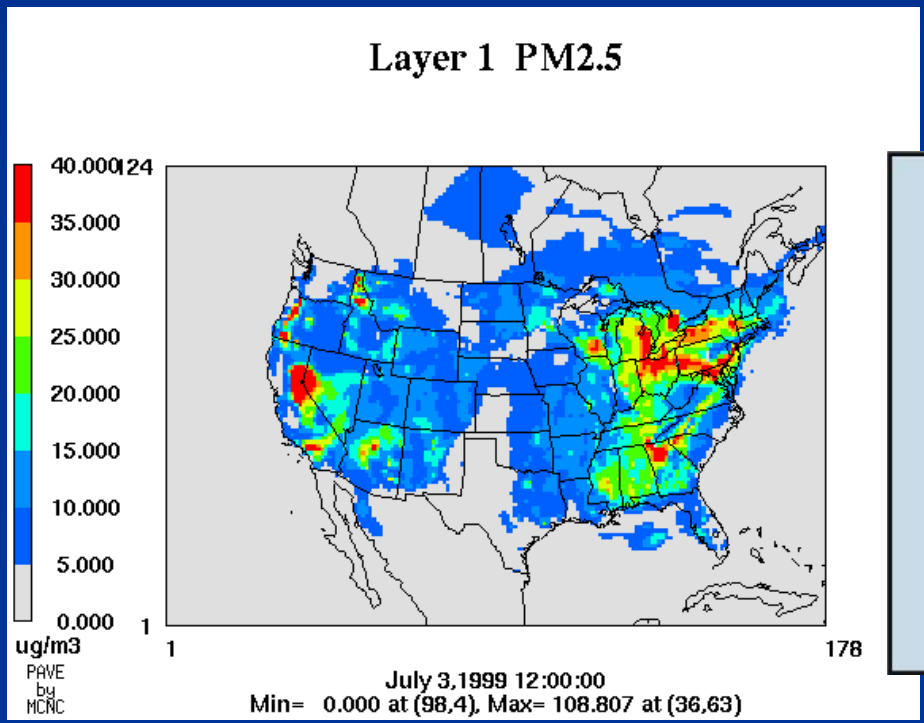
Future Work?

- Examine needs for environmental health data standards?
- Privacy and confidentiality?
- Outreach to environmental and health professionals about what data is available?
- Early lessons from EPHTN implementation?

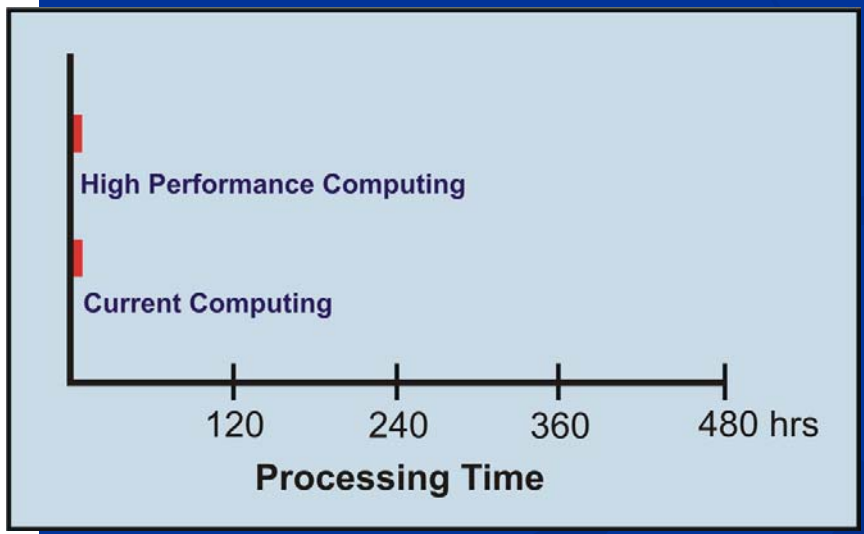
Partnership: Department of Energy MOU

Collaborative Effort: EPA & DOE

High Performance Computing



CMAQ Model Run Time



Collaborative Effort: EPA & DOE

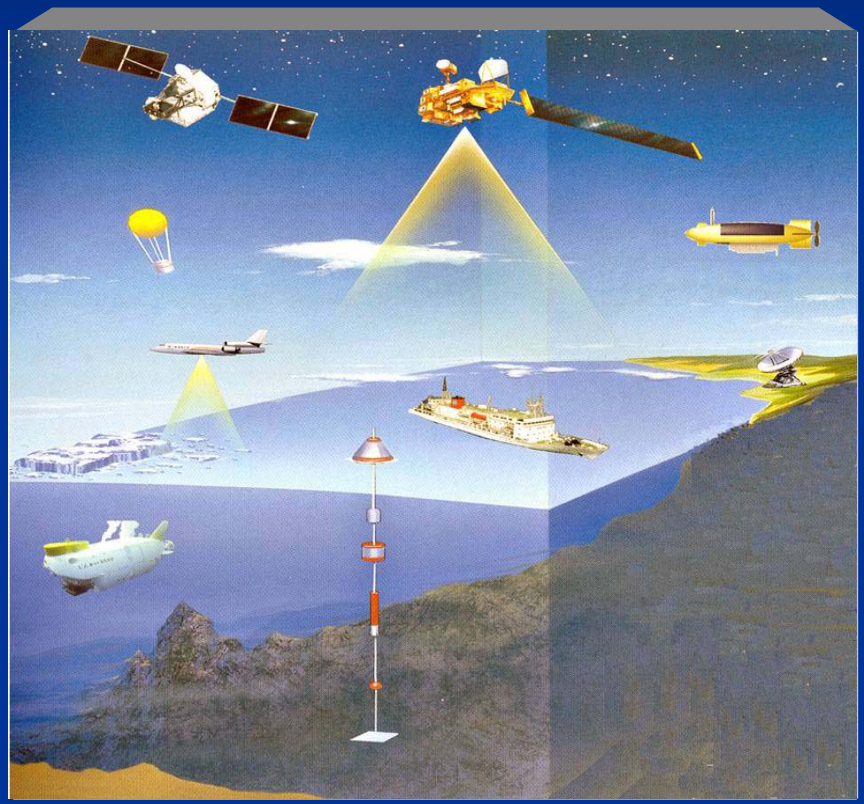
Computational Toxicology

- Computational Toxicology is the application of mathematical and computer models for predicting effects and improving the understanding of mechanisms of action during exposures.
- The overall goal of Computational Toxicology at EPA is to make use of emerging technologies to improve quantitative risk assessment by reducing uncertainties in the source-to-adverse outcome continuum.

Collaborative Effort: EPA & DOE

Computational Toxicology

- The three strategic objectives of the Computational Toxicology Program at EPA are to develop:
  - improved linkages across the source-to-outcome continuum,
  - approaches for prioritizing chemicals for subsequent screening and testing, and
  - better methods and predictive models for quantitative risk assessment.

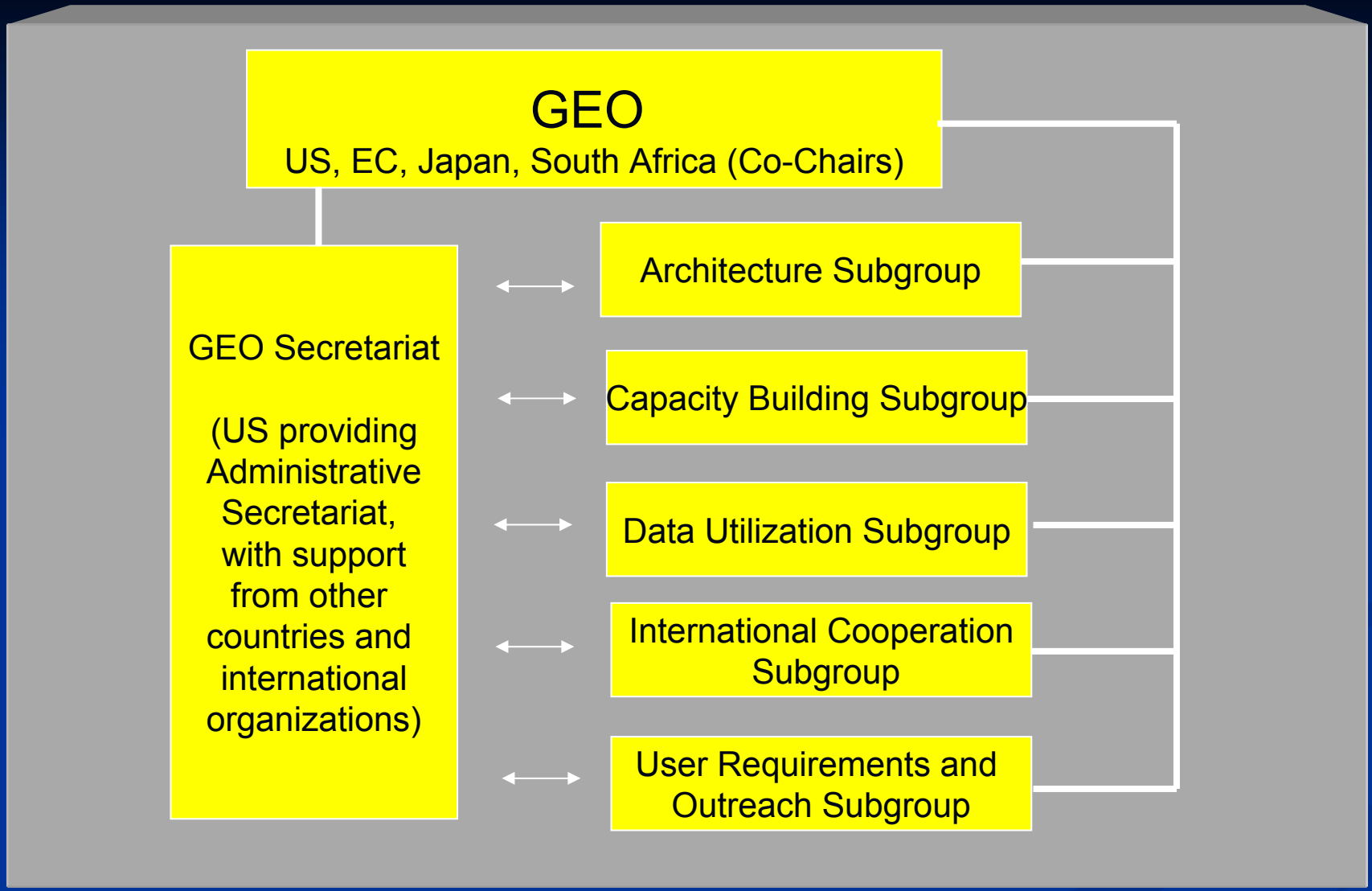


Partnership: Group on Earth Observations (GEO)

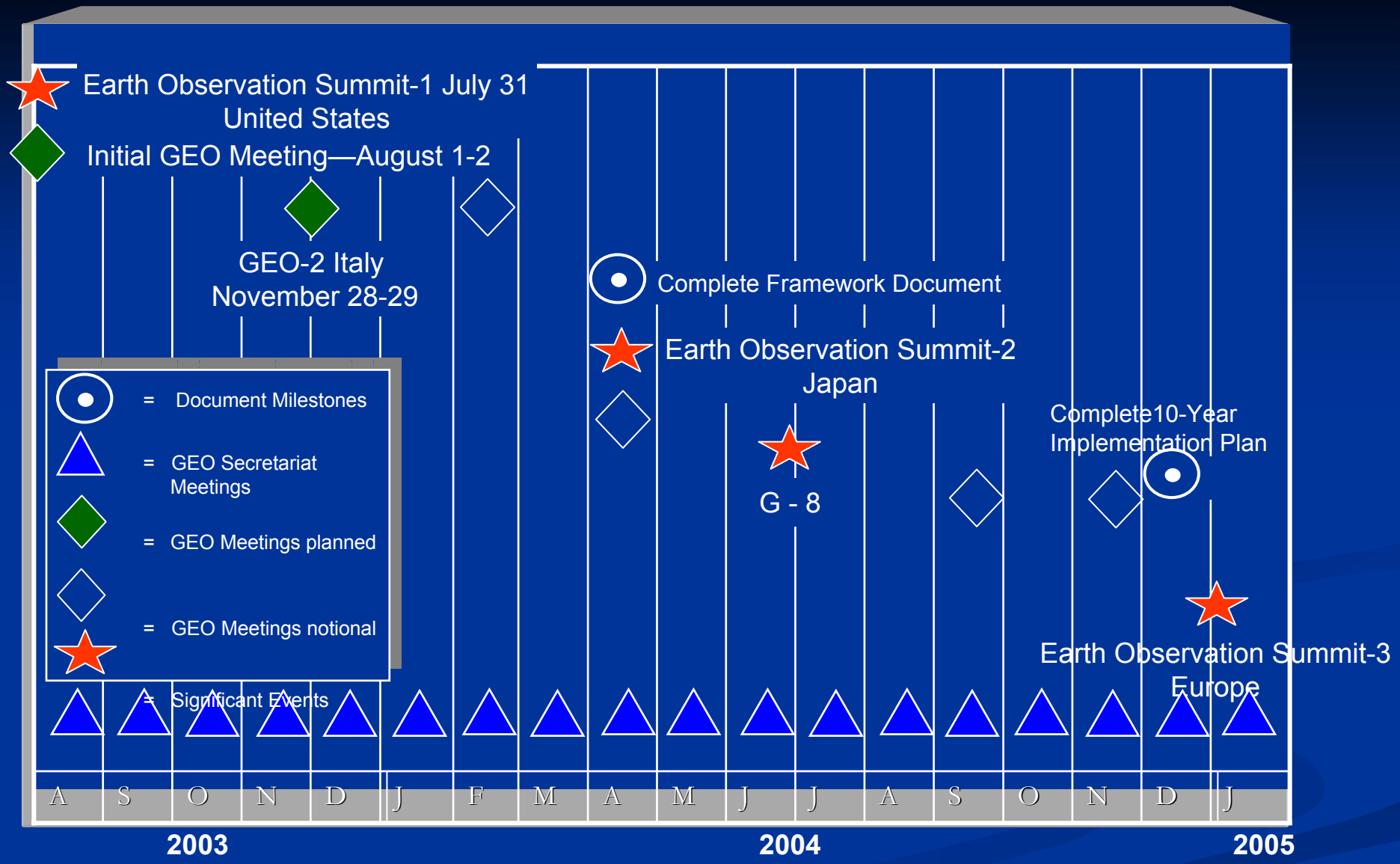
GEO Objectives

- Support the creation of a comprehensive, coordinated, and sustained Earth observing system or systems to provide timely, quality, long-term, global information; to monitor continuously the state of the Earth; to increase understanding of dynamic Earth processes; to enhance prediction of the Earth system; and to further implement environmental treaty obligations.

- Develop a 10-Year Implementation Plan with a conceptual framework to be presented at a ministerial meeting in Tokyo in the Spring of 2004. The plan will be presented at a GEO ministerial summit in late 2004.



Structure for GEO



Schedule for GEO

And So What Does All this Mean?

- Collaboration & partnerships are key
- Networks of all kinds will be linked
- Our efforts will be in a global context
- Expectations need to be realistic
- Make the data move!