

Perspectives on the Use and Presentation of International Environmental Data and Implications for US Data Collection and Integration

US EPA TRI Training Conference

November 3, 2010

John Chelen

Hampshire Research Institute

Questions We Need to Think About

1. What are we trying to accomplish and why is it hard?
2. How might we better correlate international environmental data?
3. What are the implications for US data collection?
4. What Are the implications for US data presentation?
5. Can we find consensus on possible next steps?

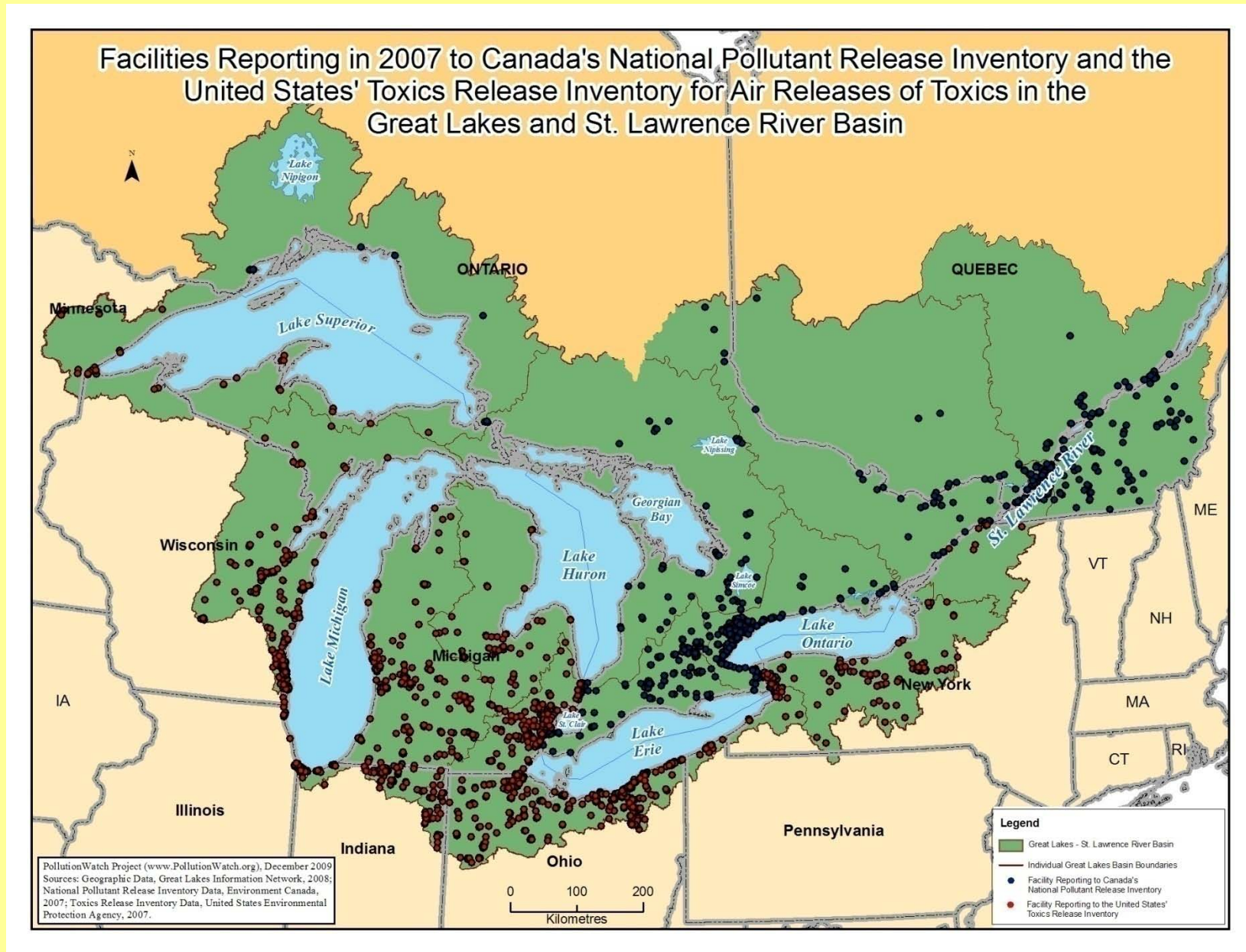
1. What are we trying to accomplish and why is it hard

Environmental Issues Don't Respect Borders

Example:

Along its northern border region, US facilities are not the only source, and sometimes not the primary source, of pollution in a specific area. Canadian and US data are comparable and can be used in policy analysis.

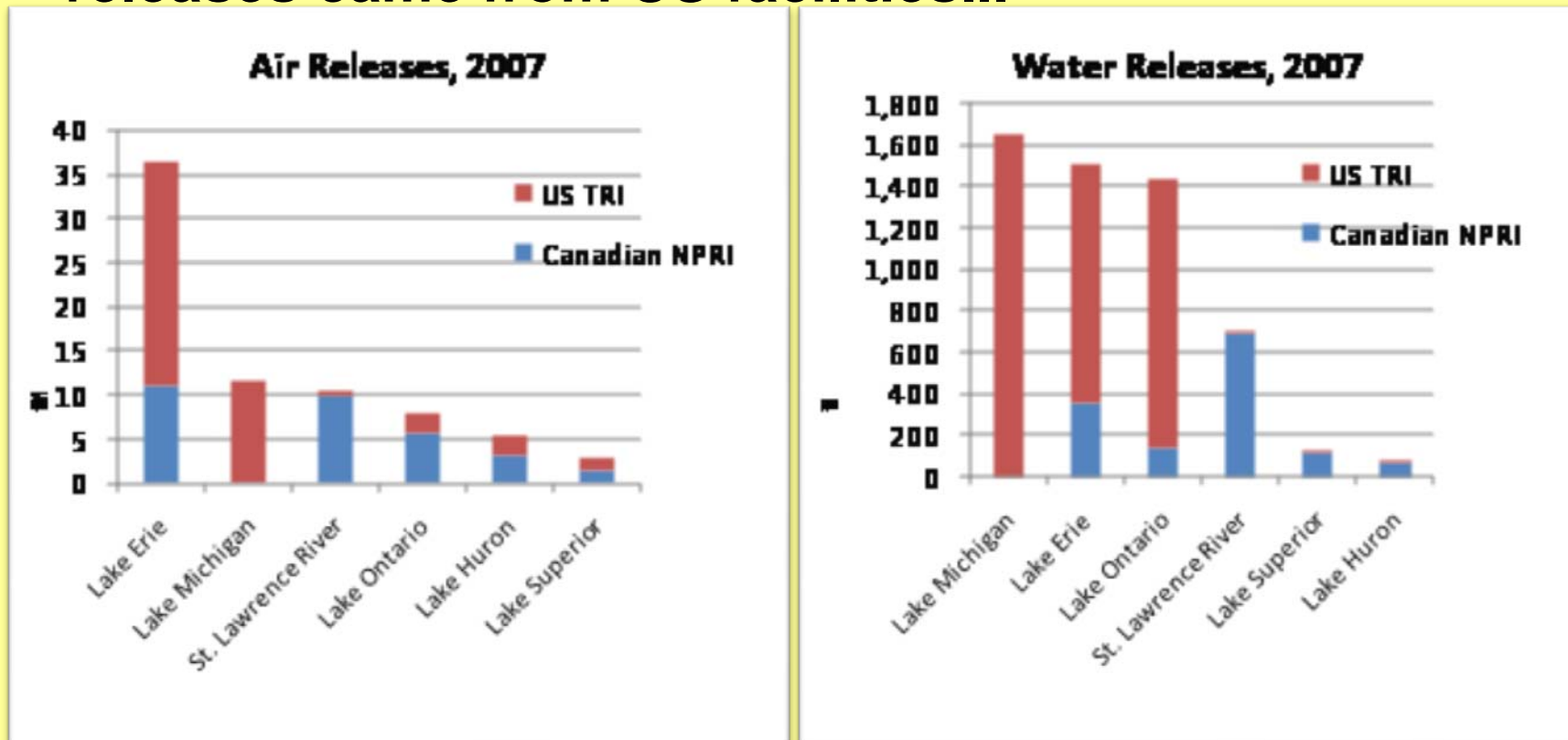
Great Lakes Facilities Reporting to Canada's NPRI and US TRI



Canadian and US facilities in Great Lakes, from PollutionWatch report, "Partners in Pollution 2: An Update on the Continuing Canadian and United States Contributions to Great Lakes-St. Lawrence River Ecosystem Pollution" December 2009, www.pollutionwatch.org

For water releases, 75% was from US facilities and 25% was from Canadian facilities.

In the Lake Ontario basin, air releases came predominantly from Canadian facilities but water releases came from US facilities...



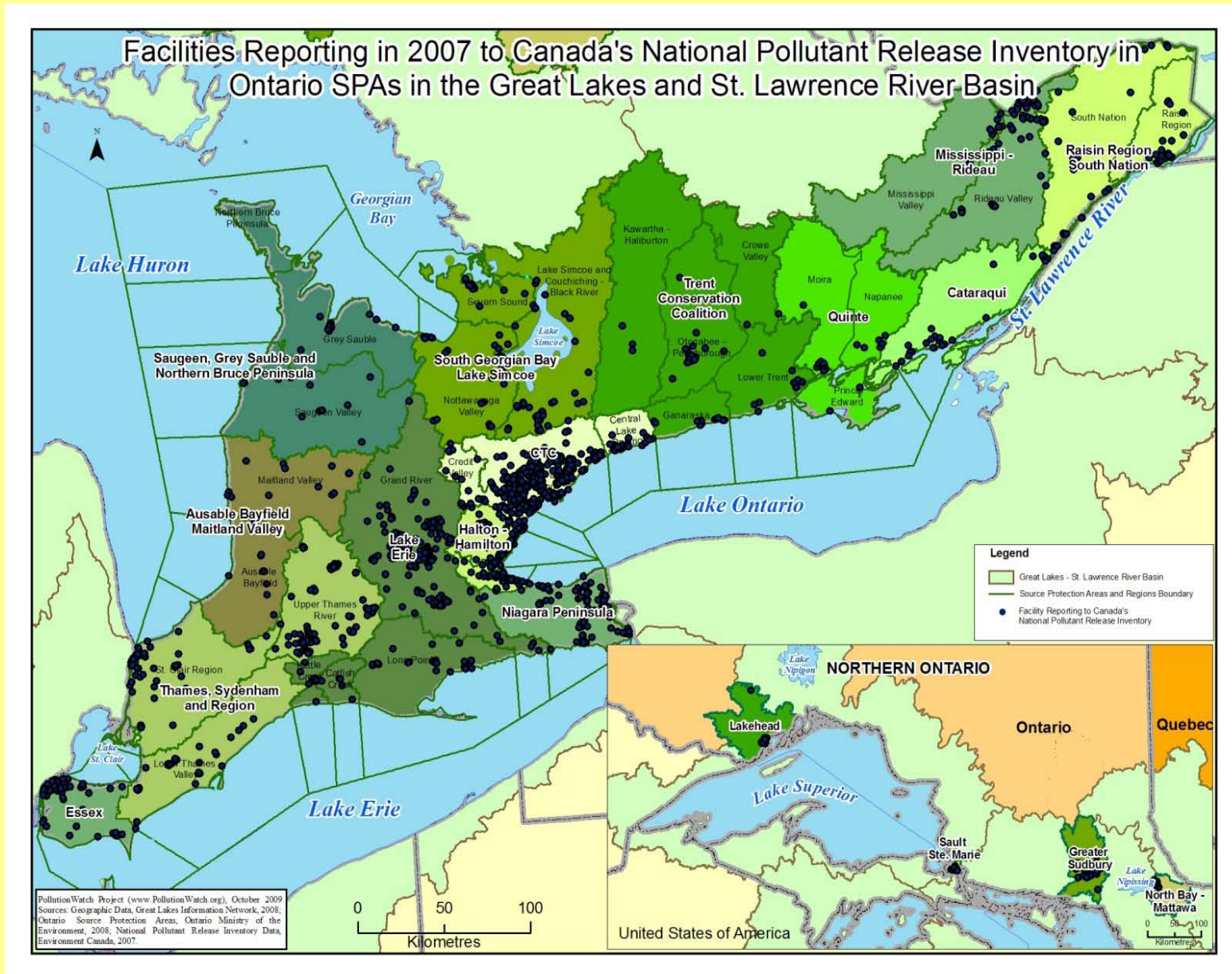
From PollutionWatch report, "Partners in Pollution 2: An Update on the Continuing Canadian and United States Contributions to Great Lakes-St. Lawrence River Ecosystem Pollution" December 2009, www.pollutionwatch.org

However, Canadian and US data aren't completely compatible...

National coverage can vary.

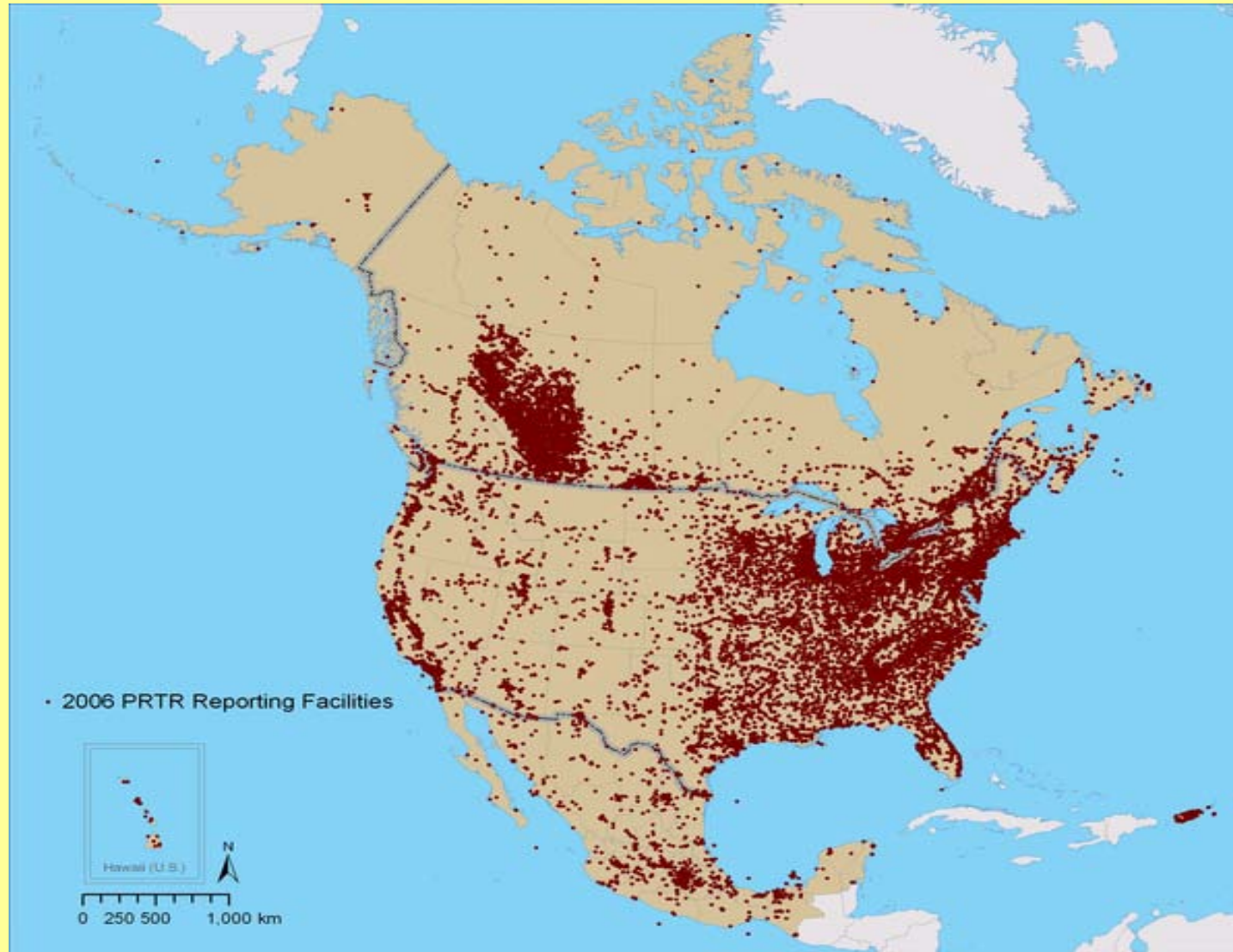
For example, in several respects, the Canadian NPRI has greater coverage, since it requires reporting from municipal sewage plants and landfills, as well as oil/gas exploration facilities and additional chemicals not on the TRI.

Facilities reporting to NPRI in Ontario Special Protection Areas For Drinking Water Sources including Municipal Sewage Treatment Plants



Facilities reporting to Canadian NPRI in Ontario Special Protection Areas from "Protecting the Great Lakes - St. Lawrence River Basin and Drinking Water Sources" a report by PollutionWatch, December 2009, www.pollutionwatch.org

This map shows the lack of information on oil and gas exploration facilities in the western US south of the border, whereas there is information on these facilities north of the border.



2. How might we better correlate international environmental

- (1) Compatible Data Collection Standards
- (1) Compatible Data Collection Regulations
- (1) Joint Special Research Efforts
- (1) Earth Environmental Reference Model
- (5) Earth Environmental Data Service

(1) Compatible Data Collection Standards

Standards efforts are underway in many venues – chemicals, oceans, health effects, etc.

Standards efforts are typically slow and limited in scope.

The original international PRTR effort was intended to develop a common shared approach to track chemical pollutant releases. However, for over half a decade, the US has not strongly supported this effort.

(2) Compatible Data Collection Regulations

Nearly impossible because of legal traditions,
economic priorities, etc.

(3) Joint Special Research Efforts

Joint research is underway in many noteworthy areas, e.g., climate change, and scientists can cooperate on research methods.

Research can be disconnected from political impleme

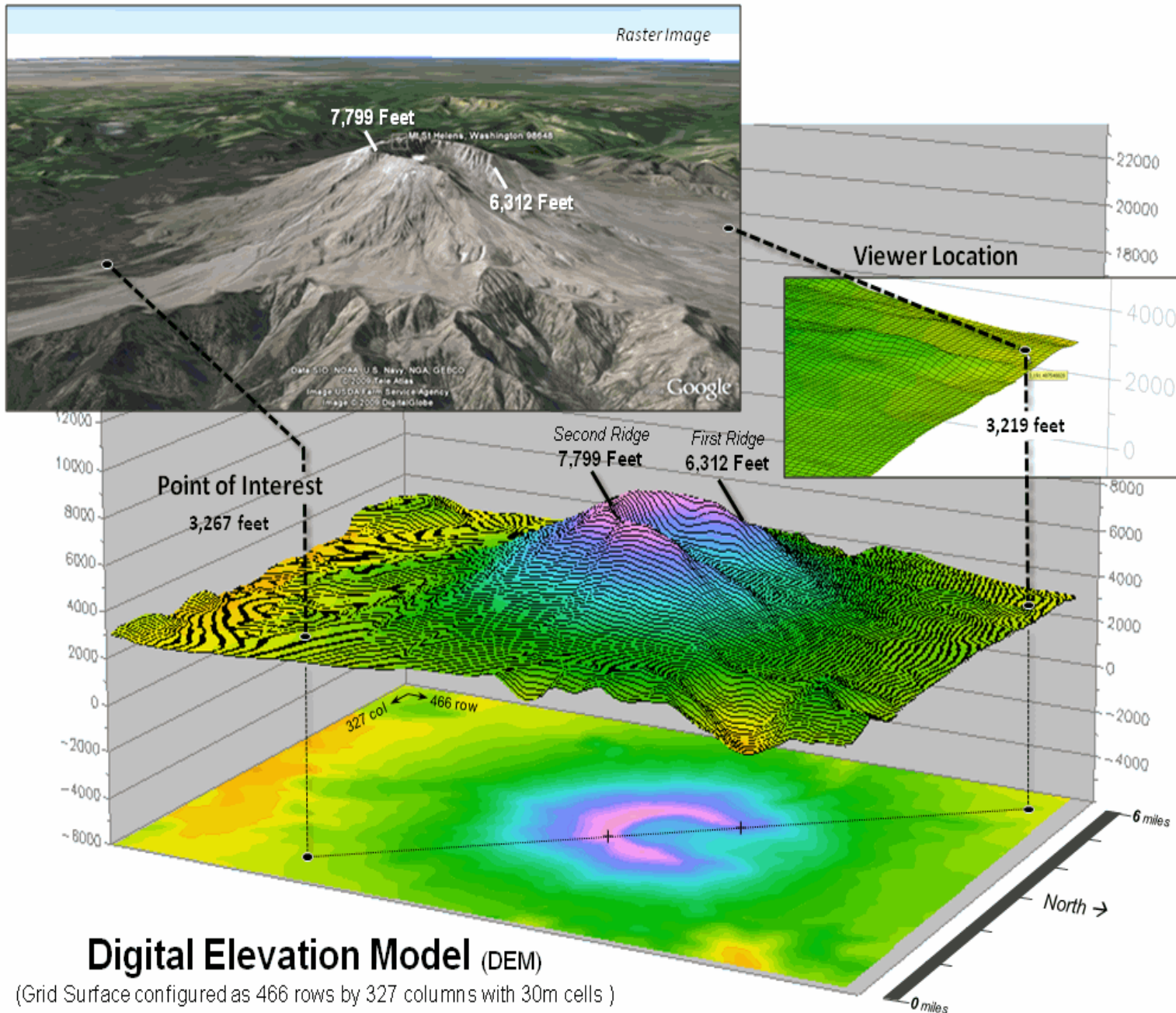
(4) Earth Environmental Reference Model (EERM)

- An EERM is a means to connect any data element to a shared logical framework to support joint research or monitoring. It can connect what happens anywhere on the earth in a specific location on, under, or above the earth at a specific time.
- An EERM can provide “Rosetta Stone” to crosswalk any data to neutral framework.
- An EERM would consist of a virtual geospatial/temporal/elevation matrix with an object (e.g., facility or chemicals) taxonom.
- The EERM matrix would consist of “cubic” cells permitting latitude/longitude/ elevation/time tagging of event, data string, or analytical objects.
- Analytical models could be based on EERM coordinates and operate upon EERM structured data.

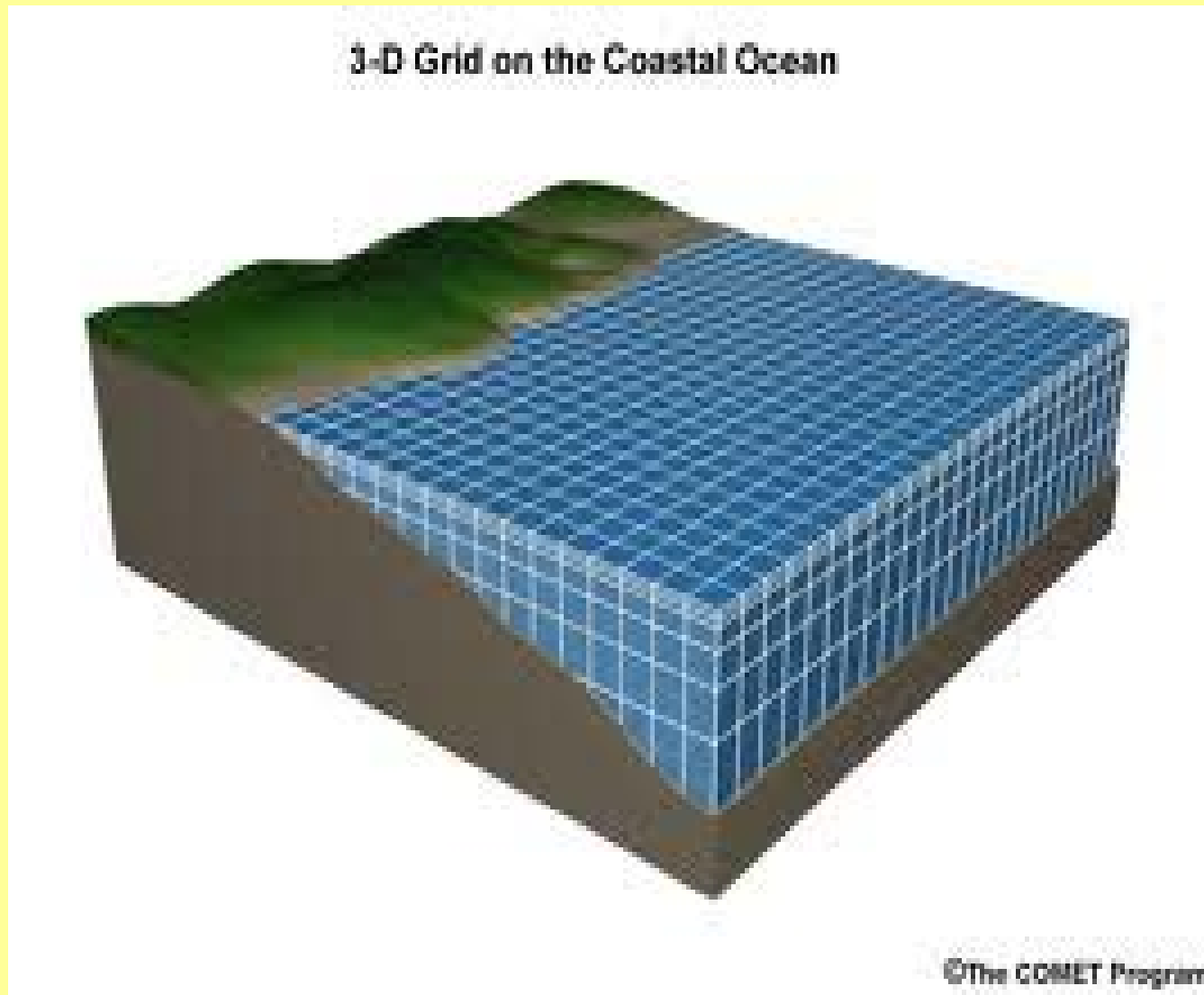
The surface of the Earth is divided into cells:



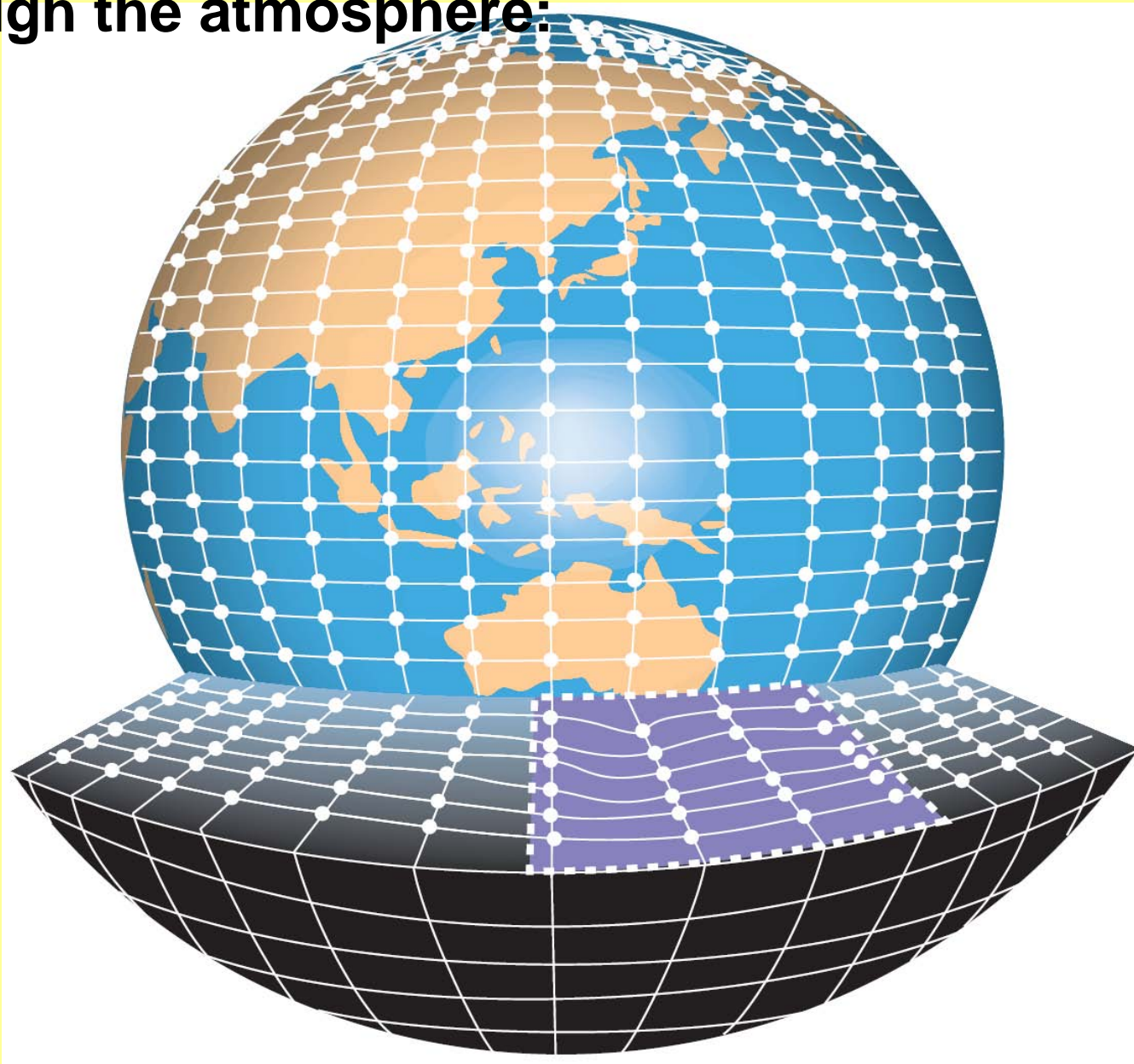
We also need to take into account elevation and topolog



An EERM should be based on a 3-D analytical view:



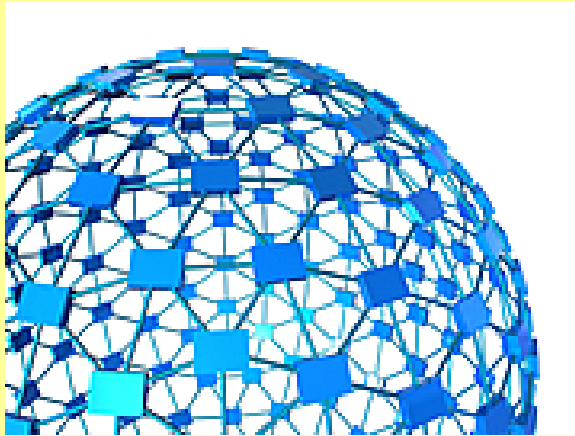
An EERM can be thought of as a 3-D stack of boxes with depth, reaching from sub-surface zones up through the atmosphere:



Major Feature Requirements of an EERM

- a. A standard cell size is agreed upon – e.g., 1 cubic meter.
- b. No maximum elevation limits atmospheric data.
- c. No maximum depth limits geological data.
- d. No maximum time frame or temporal resolution limits historical values – individual cells are repeated for time variation.
- e. Algorithms are developed by participating data contributors to migrate traditional databases to EERM format.
- f. Custom conversion tables are developed for specialized taxonomic issues – e.g., chemical names.

(5) Earth Environmental Data Service - EEDS



- a. Data-providing participants would make their data retrievable in real time through direct data calls (e.g., QL) to their stored data systems.
- b. The Internet would serve as the transport level to assure that separate data services could produce a virtual Earth data service with data pre-linked to EERM cells.
- c. Analytical applications, either workstation-based or server-based, would correlate and present retrieved data.
- d. Applications could present data in EERM format or any layout.
- e. Models could be published in manner similar to data, and permit their execution independent of original model source and rely upon data retrieved by EEDS.

3. What are the implications for US data collection?

(1) A preferred information architecture is required, based on a synthesis of conceptual frameworks from multiple requirements analyses.

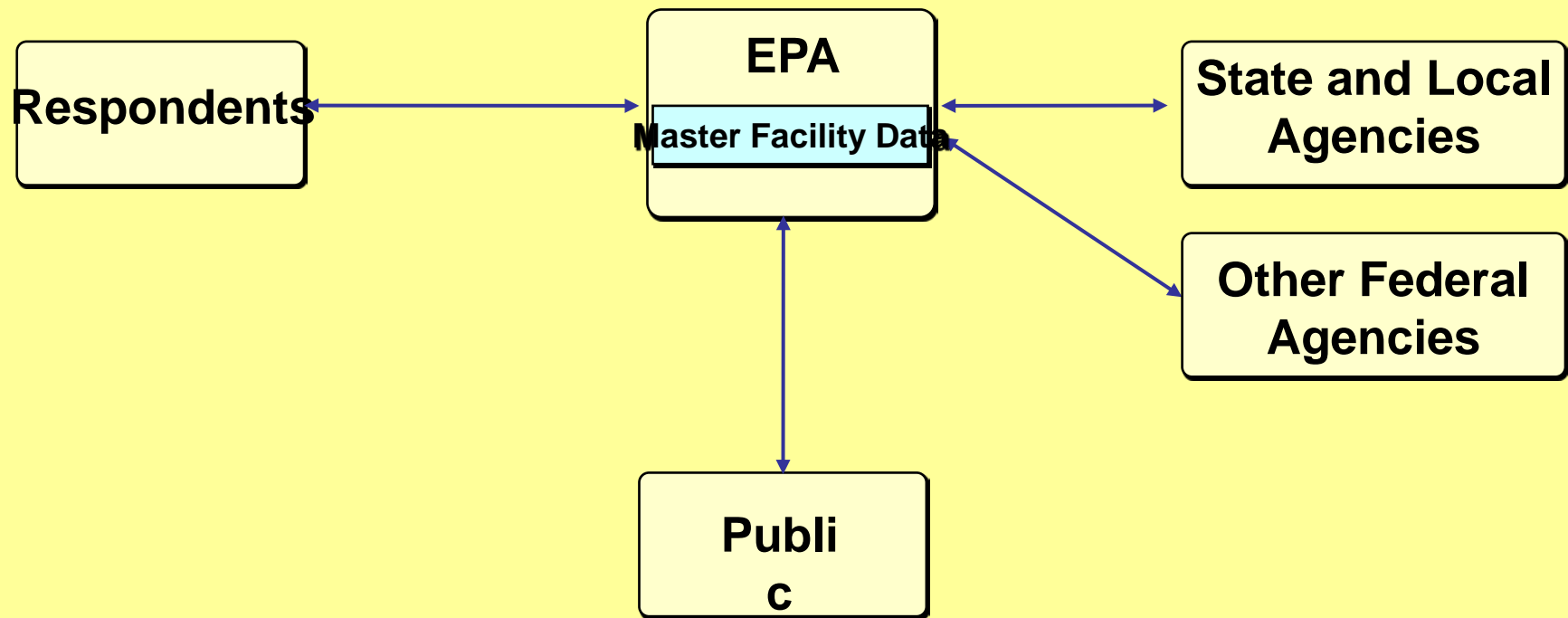
(2) Shared priority registries and common keys must be established.

(3) Shared search syntax and menu systems must be created that reflect business models and the overarching information architecture.

A Re-Invented Facility Registration System is Needed

- a. EPA should implement a Master File system at the national level to house core facility identification (ID) data.
- b. Facility managers will directly submit their changes to the single Master File system under formally promulgated regulations.
- c. All media programs and regional offices will use and share the core facility ID data in the single Master File system and eliminate ID redundancies in their own systems.
- d. All media programs and regional offices will adhere to uniform procedures necessary to maintain the accuracy of the system.
- e. EPA should begin a joint federal/state/respondent/public process to develop facility ID definitions and rules that meet all federal and state legislative and regulatory requirements.
- f. State environmental agencies, at their discretion, can rely on the Master File system as they develop the capacity to maintain and update their own facility ID records.

Proposed Reporting Framework for Master Facility ID System



4. What are the implications for US Data Presentation?

- (1) Fully Integrated Databases, preferably through a consolidated case management system
- (2) Central Registries and Master Keys
 - (1) Facet Reports (chemical, facility, regulatory program, regulated entity, ecosystem, etc.)
- (4) Integrated Search Syntax (EPA and Federal)
- (5) Report Writer and “Canned Reports”

5. Can We Find Consensus on Possible Next Steps?

- (1) We should seek ways to develop new capabilities for data.gov demonstrate EPA leadership, in particular with integration and query capabilities.
- (2) EPA should begin internal development of a conceptual EERM and EEDS.
- (3) EPA should initiate international forum efforts to develop an EERM and EEDS.
- (4) EPA should proceed immediately to modify EPA data systems for compatibility with future data.gov, EERM, and EEDS operations.