Data Standards Implementation

...how to score in the data standards game!

Jeff Kohn
kohn.jeffrey@epa.gov
Data Standards Branch
Office of Environmental Information

The Lighter Side of Standards

- Why is the US Standard railroad gauge 4 feet, 8.5 inches?

- Why did the British build them like that?

- Why did they use that gauge then?

- Why did the wagons use that odd wheel spacing?

The Lighter Side of Standards

- So who built these old rutted roads?

- And the ruts?

- And thus, the answer

  - "Wide enough to accommodate the back-ends of two war horses"

Data Standards: What Are They?

- Standards are documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines or definitions of characteristics to ensure that materials, products, processes, and services are fit for their purposes -- *International Organization for Standardization (ISO)*

- Data Standards are documented agreements on representations, formats, and definitions of common data -- *EPA*
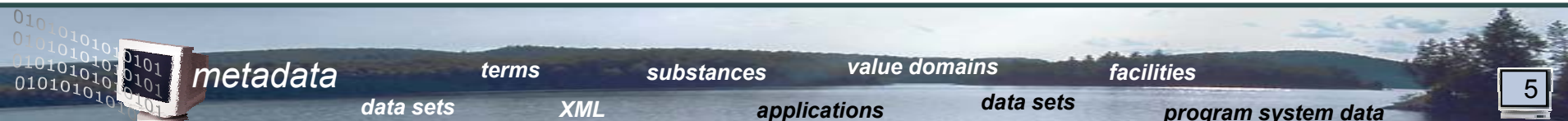
What Standards Do and Don't Do?
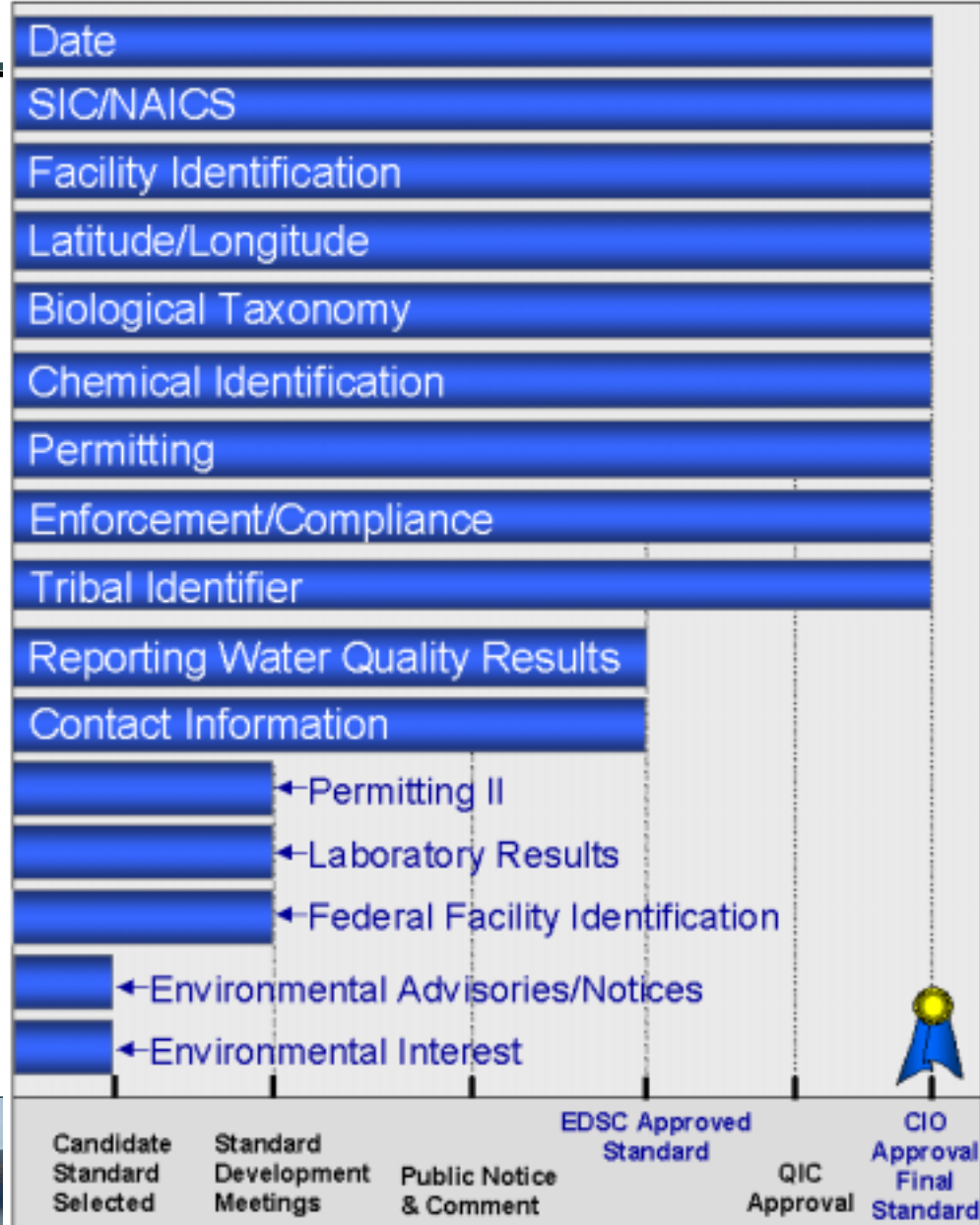
- **Standards Do:**

  - ◆ Identify data of common interest
  - ◆ Establish standard definitions and acceptable values
  - ◆ Provide a standardized set of information about the data

- **Standards Do Not:**

  - ◆ Tell people how to define their unique data needs
  - ◆ Solve data quality problems
  - ◆ Establish reporting requirements

metadata

terms   substances   value domains   facilities

data sets   XML   applications   data sets   program system data

Data Standards Status

Overview of EDSC Data Standards

Basic Information Standards

- Date

- Facility Identification

- Contact Information

- Standard Industrial Codes/North American Industrial Classification Code System (SIC/NAICS)

- Latitude/Longitude

- Tribal Identifiers

Substance Identification

- Chemical Identification

- Biological Taxonomy

Program-Specific Information

- Enforcement/Compliance

- Permitting

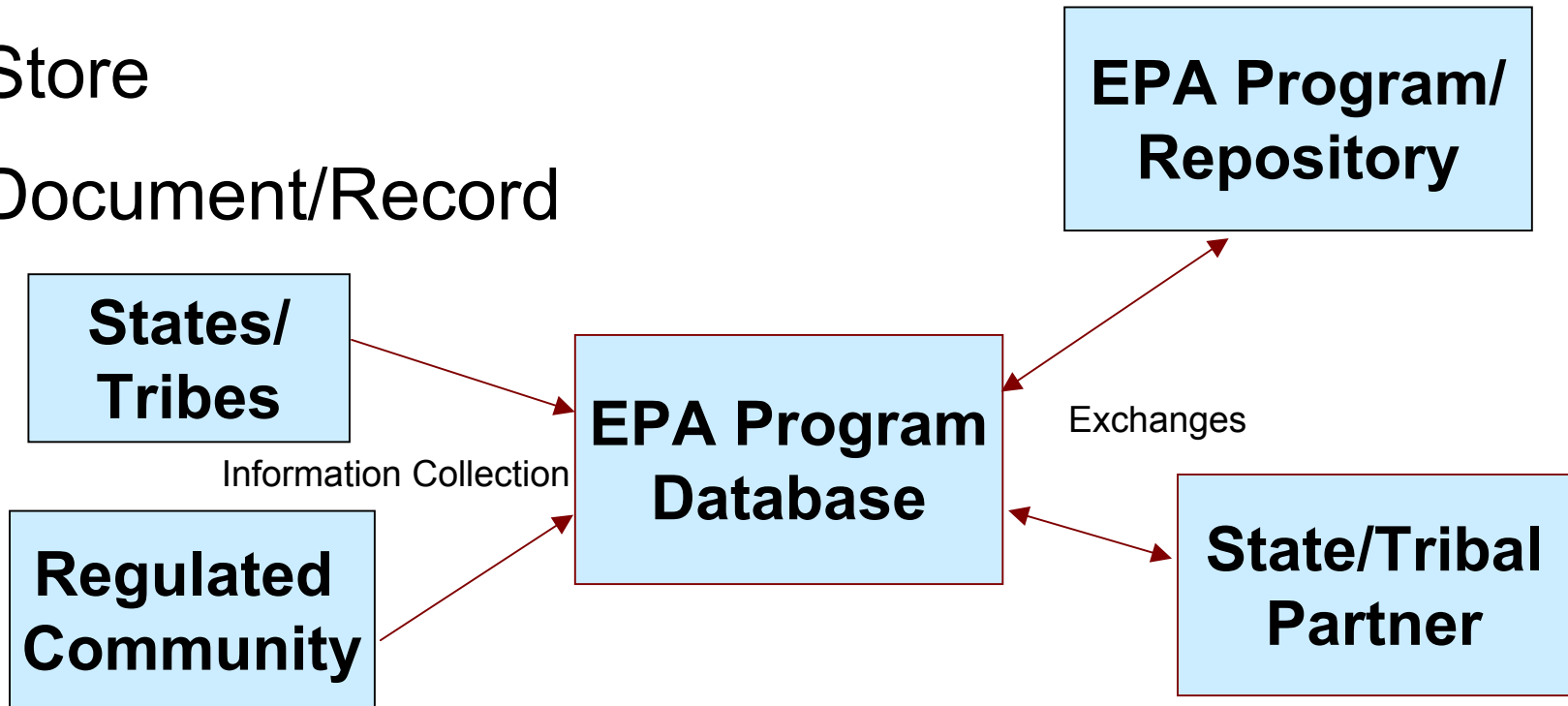- Reporting Water Quality Results for Chemical and Microbiological Analytes (RWQRCMA)

Data Standards Under Review

- Permitting II

- Federal Facility Identification

- Laboratory Results

Key Terms

- Exchange/Data Transfer

- Information Collection

- Store

- Document/Record


```
States/           Information Collection      EPA Program
Tribes        ──────────────▶  EPA Program  ──────────▶  Repository
                                Database
Regulated     ──────────────▶              ◀────────▶  State/Tribal
Community                                   Exchanges    Partner
```

Implementation Challenges

- Federal vs. State Issues

- Misunderstanding of requirements (mapping vs. database changes)

- Seeing beyond an individual program's business needs (my data vs. our data)

- Costs vary significantly across systems

State vs. EPA Implementation

- State implementation overseen by ECOS

  - States implementation focuses on mapping

  - States access the standards via the EDSC web site,

  - http:// www.epa.gov/edsc

  - Implementation NOT governed by EPA business rules

- EPA implementation overseen by DSB

  - Implementation focuses on mapping, but also storage and mandatory data sets

  - Access the standards via the EDR, http://www.epa.gov/edr

  - Governed by business rules

metadata terms substances value domains facilities
data sets XML applications data sets program system data

Conformance Approach

- Goal is to conform during exchanges with other EPA or State/Tribal Partners

- Data collection is typically affected

- How are data standards implemented?

  - ◆ Exchange Mechanism

  - ◆ Mapping

  - ◆ Acceptance of Data

XML Schema Conformance

- Data Standards Branch currently developing strategy for review

- Showing full conformance with standards using XML schema is tricky

  - Tag names are critical!

  - Format and field length are under consideration

  - DSB is researching methods of demonstrating:

    - Definitions
    - Code Sets

Tools Available

- EDSC Web Site

  - State access portal to Data Standards

- Environmental Data Registry (EDR)

  - Access to all Data Standards (proposed and final) and EPA business rules

  - Data standards fact sheets, FAQs, other documentation

  - Compare Tool (Harmonization tool)

- Substance Registry System (SRS)

  - Register chemical list

  - Compare to other chemical lists

- Terminology Registry System (TRS)

  - Compare terminology with other EPA, state, international entities