

**CUPA Forum Board Issue Paper  
for  
Unified Program Data Management**

December 17, 2003

**Recommendation:**

Cal EPA develop and implement a Unified Program Data Management Action Plan in partnership with the CUPA Forum Board (CFB). Use the Preferred Data Exchange Model developed by the California Conference of Directors of Environmental Health as a model for Unified Program related data management issues

**Key Issues**

- Coordination and communication of Data Management issues and projects by Cal EPA.
- Identification of various projects, local systems, agencies involved, issues, concerns, efforts at all levels.
- Adoption of a Cal CFB / UPAAG Data Management philosophy/direction.
- New or different data requests. Justification, process, interpretation, frequency, etc.
- Understanding local agency concerns and securing assistance to local agencies (eg. grants).
- Electronic vs paper data management and reporting.

---

1. CUPA Forum Board Information & Data Management Philosophy

The CFB is committed to work with all stakeholders to develop necessary reporting and data exchange procedures to meet state and federal requirements with minimum impacts to local agencies. To meet this goal, the CFB requests that state and federal agencies coordinate with the CFB through various workgroups and UPAAG to provide adequate notice and explanation for their information needs and intended use of this information. CFB understands the potential impacts to local agencies and will work with state and federal agencies to ensure that changes in data needs are necessary, that impacts to local agencies are minimized to the extent possible, that adequate lead time for needed changes is provided, and that whenever possible, financial support will be offered to assist local agencies in modifying local data management systems to meet new or different needs.

Further, the CFB recognizes that many local agencies may continue to use non-electronic methods for portions of their local programs and that this data may not be readily available in an electronic format. CFB recommends that whenever possible, state and federal agencies seek funding to assist local agencies in the conversion and upgrading to electronic systems.

To minimize long-term impacts to local programs, CFB recommends that a reasonable schedule be established to review and modify the Title 27 data dictionary.

## 2. Information Needs & Stakeholders

### a) State and federal agencies and their information needs, including:

#### 1 US EPA

- RCRA Large Quantity Generator Data (immediate need)
- RCRA Data for all HWG inspection and enforcement activity (future need-when?)
- EPCRA US EPA requested data
- UST Data
- Facility Registry
- NPDES
- Future data developed to measure the environmental outcomes associated with CUPA program performance rather than just its outputs

#### 2 Cal EPA

- SWRCB
- DTSC
- OES
- SFM

#### 3 Other State Agencies

- DHS
- CIWMG
- 

### b) Local Unified Program Agencies

- 1 Large and small agencies have different available resources and funding
- 2 Urban compared to rural area will have different local agency priorities
- 3 Fire and County Environmental Health agencies may have different perspectives
- 4 Identify and compare local data needs with state and federal agencies. How is local data gathered and managed?

### c) Other stakeholders

- 1 Local Emergency Response Agencies

- 2 Industry-Large and small industry representatives
  - 3 The public
  - 4 Non-Governmental Organizations
  - 5 Elected federal, State and Local Government Officials
- d) Correlate agency needs and program elements
- 1 Identify common information needs and incorporate into data management plan
  - 2 Identify inconsistencies and resolve
- e) Identify existing state & federal information databases and describe their purpose.
- 1 RCRA Info
  - 2 Enforcement and Compliance History Online (ECHO)
  - 3 Online Tracking Information System (OTIS)
  - 4 Windows to My Environment (WME) – future database
  - 5 TRI – business to EPA
- f) Identify industry data reporting requirements to federal, state or local agencies
- 1 HMBP
  - 2 Tier II
  - 3 UST Forms
  - 4
- g) Continue current review of existing data dictionary – add or delete where possible
- 1 Establish a periodic review schedule
  - 2
- h) Software providers: What needs can they address; what opportunities are available for future?
- 1 Envision
  - 2 CUPA DMS
  - 3 Unidocs
  - 4 Customized local systems
  - 5 Others

### 3. Data Projects & Management

- a. Identify key databases used by local unified program agencies
- 1 Envision
  - 2 CUPA DMS
  - 3 Geotracker
  - 4 Unidocs
  - 5 FileMaker Pro
  - 6 Locally developed systems
  - 7 Create and maintain a list of local agencies using non-electronic methods
  - 8

- b. Identify and track pilot projects and key objectives/interests
  - 1 Sacramento County/Cal EPA Challenge Grant re: Enforcement Data Exchange-in process
  - 2 Livermore/USEPA RCRA LQG database comparison-in process
  - 3 San Diego/USEPA RCRA LQG database comparison-pending(?)
  - 4 Unidocs/Cal EPA/USEPA Readiness Grant re: XML Data exchange-pending
  - 5 Unidocs/SWRB re: UST Data integration-pending
  - 6 Unidocs/USEPA re: Tier II Reporting and CAMEO integration-applied
  - 7 Unidocs/USEPA re: RCRA LQG data integration-proposed
  - 8 EPIC Indicator Project
  - 9 Mobile/electronic field inspection efforts-various
- c. Identify and track various committees or working groups working on data dictionary, Title 27, data reporting and their key objectives/interests
  - 1 CFB Forms & Data Committee
  - 2 EPA-ECOS Information Management Workgroup (Cal/EPA is on this group... Gary Arstein-Kerslake)
  - 3
- d. Identify by program element, objectives and requests from State & local agencies for specific data
  - 1 US-EPA Echo information
  - 2 DTSC/US-EPA RCRA LQG information requests
  - 3 SWRCB new Quarterly Report Data requests

#### 4. Data Justification

- a. All specific information requests should be reviewed in the context of the CFB Information & Data Management Philosophy and the CCDEH preferred data exchange model with the goal of minimizing overlapping or duplicative data requests from multiple agencies for similar or different purposes.
- b. Review and discuss state and federal agency requests, utility of data and how it will be interpreted.
- c. Review and discuss current data submittal to Cal-EPA as part of annual reports, quarterly reports, annual audits and performance reviews to examine consistency and overlap issues.
- d. Consider a decision process to evaluate new data needs benefits versus impacts to gather.

#### 5. Data Exchange Issues

- a. XML reporting expectation.
- b. US-EPA Facility Registry Numbers and other data dictionary consistency issues.
- c. Electronic vs. hard copy.
- d. Real time data transfer versus periodic reports
- e. Summary data versus raw data, which should be reported and why
- f. Platform issues or concerns – who is working to upgrade their systems to be able

to handle new requirements.

g. Costs / Local Mandates / Grants / Assistance

6. Prioritization Scheme

7. Timeline – Identify short (less than 1 year) medium (1-3 years) and long (greater than 3 years) term issue resolution/project implementation.

8. Define workgroups –Consider establishing several to bring in all key stakeholders.

## **Attachment**

### **CALIFORNIA CONFERENCE OF DIRECTORS OF ENVIRONMENTAL HEALTH**

#### **Preferred Data Exchange Model (July 15, 2003 Draft)**

##### **BACKGROUND**

Local Environmental health agencies and their state and federal counterparts spend millions of dollars each year collecting, storing and reporting data. Local agencies are required by statute, regulation or program needs to collect and report ever-increasing volumes of data to these oversight agencies – increasing demands on the regulated community and local staff. Currently local environmental health agencies administer as many as twenty different programs and report to as many as eight separate state data management systems in the fields of solid waste, hazardous materials, drinking water, and food safety.

Efforts to create statewide “mega-databases” have failed due to three key factors: prohibitively high costs; huge data volumes and complexities; and incompatibility of individual systems with other existing or proposed data systems – especially local data management systems which are often designed to be multi-functional agency budgeting, staff management and regulatory information tools.

Recent technological advances in data sharing now allow for the focus in data exchange to move away from uni-directional “reporting up the chain” to true data sharing between “data partners” without the heavy dependence on data system structure. It is this group's vision that current and future data interchanges become unobtrusive, opening each agency's data systems to the well-authenticated trading partners who require it - such that data can be pulled on real time on-demand basis instead of pushed in periodic batches.

Collection and management of data in electronic format is becoming commonplace in both industry and government – facilitating rapid data transfer and accuracy. This approach provides a golden opportunity for collaboration between federal, state and local agencies and the businesses they regulate.

##### **DATA EXCHANGE MODEL**

In order to make all environmental health data management more effective and more efficient, CCDEH is advocating the adoption of the following Data Exchange Model:

- A. Specification of XML statewide data file format;
- B. Establishment and codification of Data Dictionaries for each program;
- C. Promotion of electronic data collection, storage and exchange through technical and financial assistance to local agencies.

XML Format: XML stands for Extensible Markup Language and is a self-describing data file format that, when combined with associated modern technologies such as Web Services, promote machine-to-machine data interchange in real time. The benefits of XML and web services are well documented:

- 1 Free and open industry standard - XML is the standard for business integration <http://www.w3.org/XML/> consistent with Cal/EPA's Integrated Data Environmental Assessment (IDEA) project, US EPA's Environmental Data Standards Council [http://oaspub.epa.gov/emg/EDSCDATASTDSS\\$.startup](http://oaspub.epa.gov/emg/EDSCDATASTDSS$.startup), US EPA's Central Data Exchange <http://www.epa.gov/cdx/> and <http://www.exchangenetwork.net/common/default.asp>, and many other initiatives. All are potential trading partners.
- 2 Accommodates all types of data
- 3 Built-in presentation capabilities - An XML document can be viewed in a web browser in a nicely organized human-readable format
- 4 Many freely-available readers/editors/parsers
- 5 Strict validation - Any entity that works with data can be certain that the data adheres to the data dictionary
- 6 Positions the Agency to implement Web Services as a preferred means for machine-to-machine data exchange - This would clearly make Cal/EPA a data management leader in the state of California
- 7 Machine to machine exchanges are preferred over human intervention - XML and Web Services were created for this purpose

To ensure universal participation in this format, the use and/or conversion to XML will need to be mandatory for all data partners.

Data Dictionaries: Local programs are currently using data dictionaries in several programs, most notably the Unified Program and standardized restaurant inspection reporting. These “dictionaries” provide a finite number and configuration of data fields for each data set. Formalizing data dictionaries through statute or regulation gives these standards the weight of law and permanent standing which will result in:

- 1 Consistent, justified and attainable data expectations;
- 2 Information technology systems developed by public and private industry to organize and present this data in increasingly more efficient ways; and
- 1 An approach that may be used as a model for Electronic Data Interchange in other program areas.

In addition to codifying data dictionaries, the State is encouraged to set the bar high in defining data transport and to insist on machine-to-machine exchanges that are more efficient and cost effective than the store and forward batched processes most often seen.

Electronic Data Management: Local agencies are by far the largest source of data in environmental health in the State. While most of the counties and cities have some electronic data management capability, may still lag behind in their ability to collect data

in the field and to store and manipulate such data electronically. Assisting all local agencies through grants for hardware, technical assistance or other direct funding aid will ultimately improve the quantity, reliability and consistency of the data, reduce data management costs and significantly enhance the speed and convenience of data sharing.

On behalf of the 62 local environmental health agencies that it represents, CCDEH embraces this model as the preferred methodology for data interchange and will promote its adoption at the local, state and where applicable, federal level.