

EPA Chemoinformatics *Community of Practice (CoP)*

Ann Richard
National Center for Computational Toxicology
US Environmental Protection Agency





Envirofacts Data Warehouse

[Recent Additions](#) | [Contact Us](#) | [Print Version](#) EF Search: **GO**

[EPA Home](#) > Envirofacts

Welcome to Envirofacts, your one-stop source for environmental information.

The Toxic Release Inventory (TRI) 2004 Data has been released. For further details please visit <http://www.epa.gov/tri/tridata/tri04/index.htm>

Select **Quick Start** to retrieve a sampling of information available pertaining to your area, or choose a **Topic** for more in-depth questions and answers about your area. If you are an experienced user, select **Advanced Capabilities**.

- EF Overview
- Queries, Maps, & Reports
- TRI eFDR
- Data Update
- Technical User
- Site Map
- Contact Us

Quick Start!
View environmental information for any ZIP Code, City, or County. Use State Abbreviations.

- ZIP Code
 - City, State Abbr
 - County, State Abbr
- GO**

Customer Satisfaction

Topics

- Waste
- Water
- Toxics
- Air
- Land
- Radiation
- Compliance
- Other
- Maps

Advanced Capabilities

Queries

- [Multisystem](#)
- [AIRS/AFS](#)
- [BR](#)
- [Brownfields](#)
- [CERCLIS](#)
- [ECHO/IDEA](#)
- [EMCI](#)
- [RadNet](#)
- [>Customized](#)
- [FRS](#)
- [>EZ](#)
- [GICS](#)
- [ICR](#)
- [Locational Information](#)
- [>LRT Viewer](#)
- [NCOD](#)
- [PCS](#)
- [>Customized](#)
- [RCRAInfo](#)
- [RADInfo](#)
- [SDWS](#)
- [TRI](#)
- [>Customized](#)
- [>eFDR](#)
- [>EZ](#)
- [>Form R](#)
- [UV Index](#)

[EPA Home](#) | [Privacy and Security Notice](#) | [Contact Us](#)

Last updated on Tuesday, April 11th, 2006
URL: <http://www.epa.gov/enviro/>

Envirofacts Master Chemical Integrator (EMCI)





Envirofacts Data Warehouse

[Recent Additions](#) | [Contact Us](#) | [Print Version](#) EF Search: **GO**

[EPA Home](#) > Envirofacts

Welcome to Envirofacts, your one-stop source for environmental information.

The Toxic Release Inventory (TRI) 2004 Data has been released. For further details please visit <http://www.epa.gov/tri/tridata/tri04/index.htm>

- EF Overview
- Queries, Maps, & Reports
- TRI eFDR
- Data Update
- Technical Us
- Site Map
- Contact Us

Quick Start
View enviro
information
Code, City,
Use State
Abbreviatio

- ZIP Code
- City, State
- County, State
- GO**

Customer S

Envirofacts Master Chemical Integrator (EMCI)

[Recent Additions](#) | [Contact Us](#) | [Print Version](#) EF Search: **GO**

[EPA Home](#) > [Envirofacts](#) > [EMCI](#) > Query Form



Query Form

Search the EMCI Database

The Chemical Query Form allows you to obtain the acronyms, chemical identification numbers, and chemical names reported by the Envirofacts databases (AFS, PCS, RCRAInfo, and TRIS) using the Envirofacts Master Chemical Integrator (EMCI). You may see if the chemical is included in other groups, or is made up of other components.

[User's Guide](#)

Chemical Selection

You may enter one or more name fragments under the Chemical Name search option, separated by a space. If you enter more than one name fragment, the "Containing" Radio Button has to be selected. All chemical names in the EMCI are searched concurrently, including CAS index names, common names, and chemical names and descriptions used by program office systems. RCRA hazardous waste codes can also be searched as name fragments. More information about entering multiple fragments is available in the [user's guide](#).

Chemical Search Option:

Chemical Option Value:

- Beginning With
- Exact Match
- Containing



- Recent Additions
- Newsroom
- Search IRIS
- Multiple Substance Reports
- What is IRIS?
- IRIS Guidance Documents
- Related Links
- Download IRIS
- IRIS Track
- Help

Integrated Risk Information System

[Recent Additions](#) | [Contact Us](#) | [Print Version](#) Search:

[EPA Home](#) > [Browse EPA Topics](#) > [Human Health](#) > [Health Effects](#) > [IRIS Home](#) > IRIS Search

Search IRIS

The Search IRIS page enables users to find IRIS files, identify substances with similar toxicological properties, and conduct other comparative analyses of IRIS data. The [CASRN](#) and Keyword Search options enable users to determine whether a substance not included on the List of IRIS Substances is listed under a synonym or addressed in IRIS as part of a broader substance category (e.g., Lead and compounds, PCBs). The [Keyword](#) Search can also be used to locate information on general topics either within the IRIS summaries and Toxicological Reviews themselves or throughout the entire IRIS Website, including IRIS guidance documents and other background information.

Search IRIS by Keyword

[List of IRIS Substances](#)

Full IRIS Summaries/Toxicological Reviews
 Entire IRIS Website

The Search IRIS page also enables users to find substances with similar toxicological properties. The [Critical Effect/Precursor Effect/Tumor Type](#) Search enables the user to find substances that effect similar organs or tissues in similar ways. The [Evidence for Human Carcinogenicity](#) Search enables the user to find substances by the weight-of-evidence characterizations used in the 1986, 1996, and 1999 Carcinogen Risk Assessment Guidelines. The [Quantitative Toxicity and Related Values](#) Search enables users to identify substances within ranges of cancer and noncancer dose-response values, including RfDs, RfCs, slope factors, and unit risks. The [Uncertainty and Modifying Factors](#) Search enables users to analyze sources of uncertainty and variability and the application of uncertainty factors for derivation of RfDs and RfCs across the IRIS database.

Search IRIS by Substance Name or CASRN [back to the top](#)



The Search by Substance or CASRN allows IRIS users to enter any substance name or CASRN and get a list of IRIS Substances that match or are a synonym for that substance name or CASRN. [Search Tips](#)

Enter a Substance Name or CASRN:



Integrated Risk Information System

[Recent Additions](#) | [Contact Us](#) | [Print Version](#) Search:

[EPA Home](#) > [Browse EPA Topics](#) > [Human Health](#) > [Health Effects](#) > [IRIS Home](#) > IRIS Summaries

IRIS Substance List

The substances are listed in alphabetical order. You can click on the first letter of the substance you want which will bring you to the section that the substance is in, or use your browser's Find command to search for a substance name or Chemical Abstracts Service Registry Number (CASRN). These substance files are typically about 15K to 40K in size, within a range from less than 10K up to about 120K.

(To search the IRIS database, go to the [Search page](#))

Search IRIS by Keyword

[List of IRIS Substances](#)

Full IRIS Summaries/Toxicological Reviews
 Entire IRIS Website

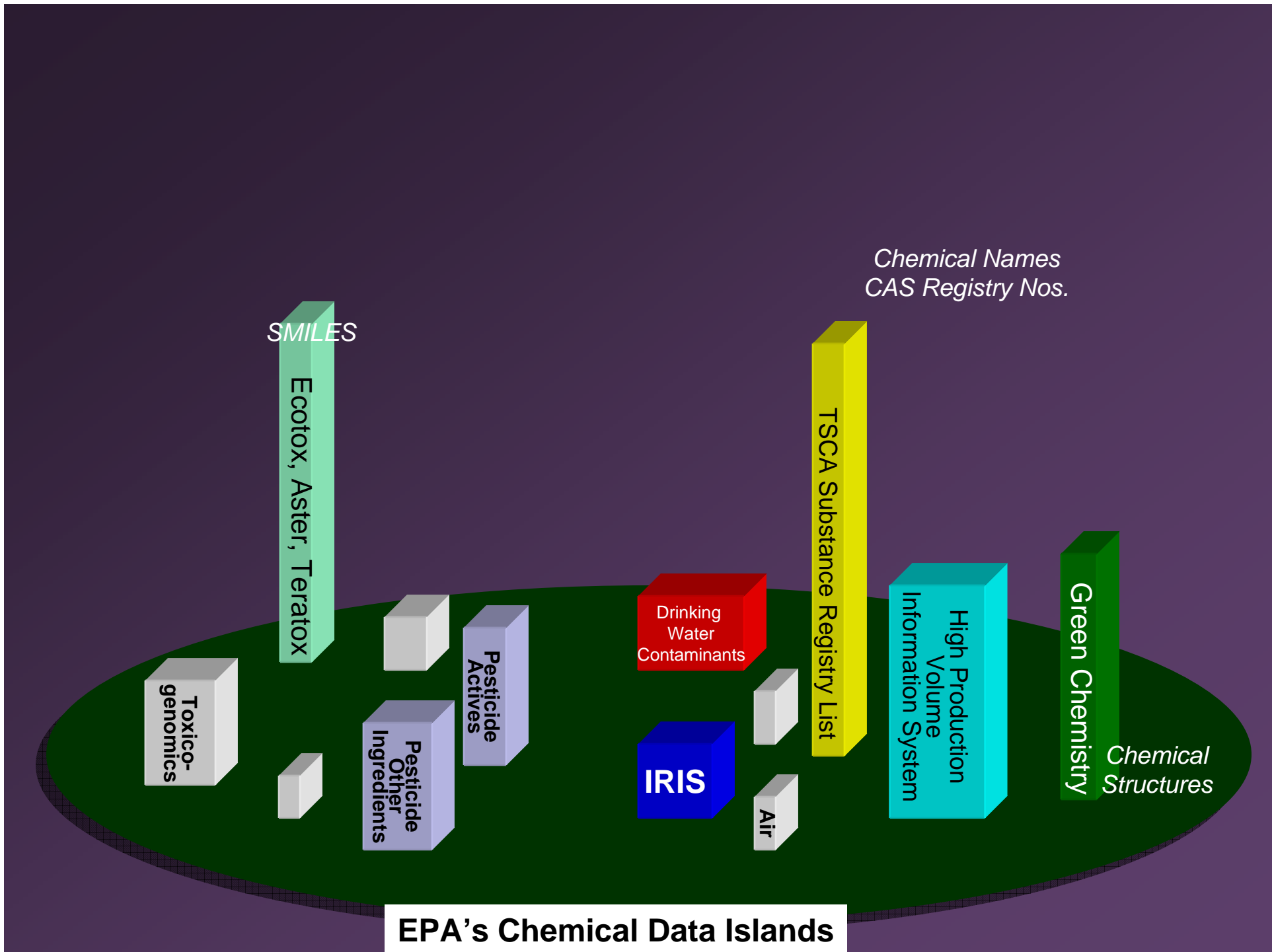
[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

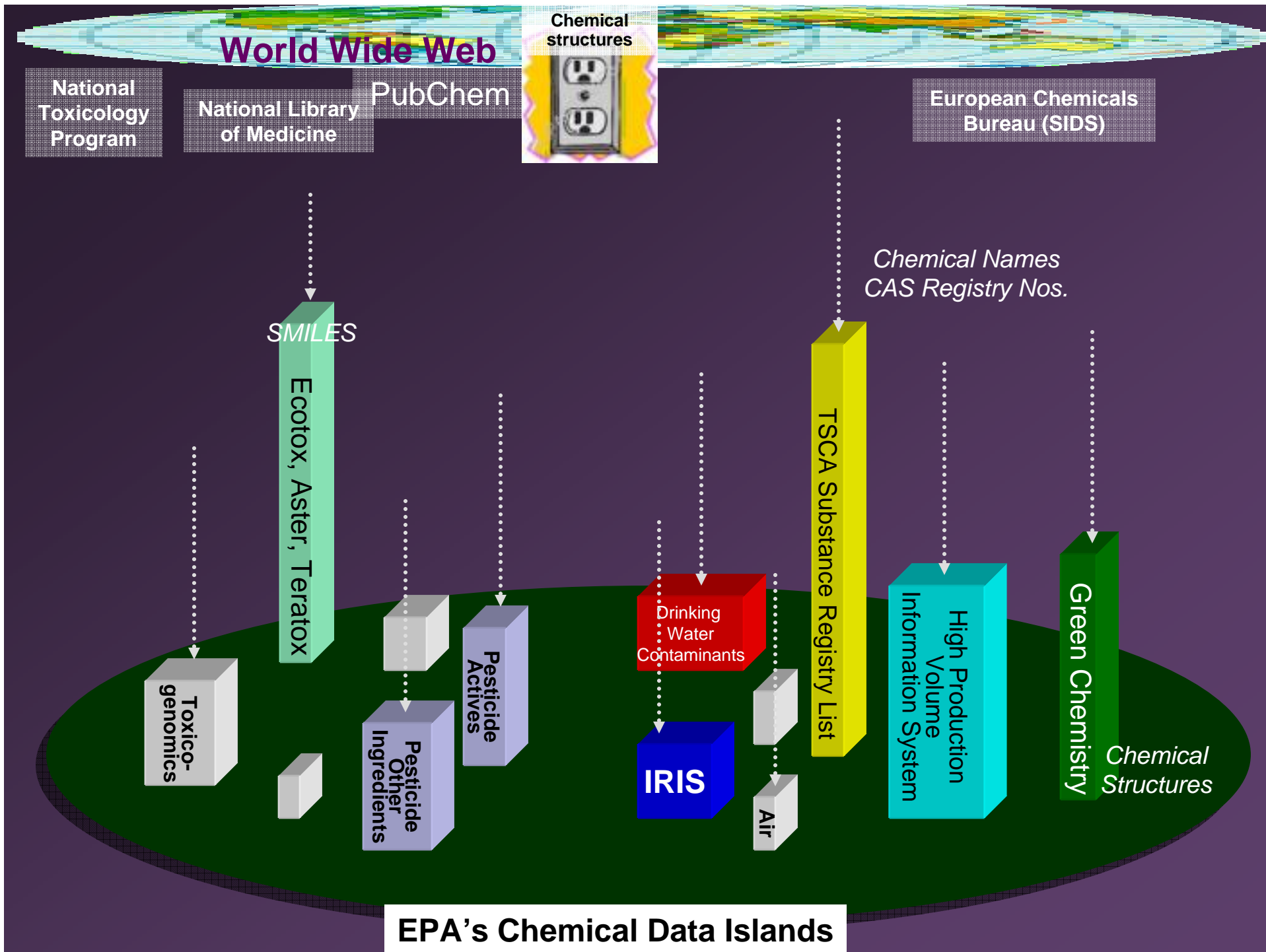
Substance Name	Chemical Abstracts Service Registry Number (CASRN)	Last Significant Revision**
Acenaphthene	CASRN 83-32-9	11/01/1990
<ul style="list-style-type: none"> • QuickView 		
Acenaphthylene	CASRN 208-96-8	01/01/1991
<ul style="list-style-type: none"> • QuickView 		
Acephate	CASRN 30560-19-1	05/01/1989
<ul style="list-style-type: none"> • QuickView 		
Acetaldehyde	CASRN 75-07-0	10/01/1991
<ul style="list-style-type: none"> • QuickView 		

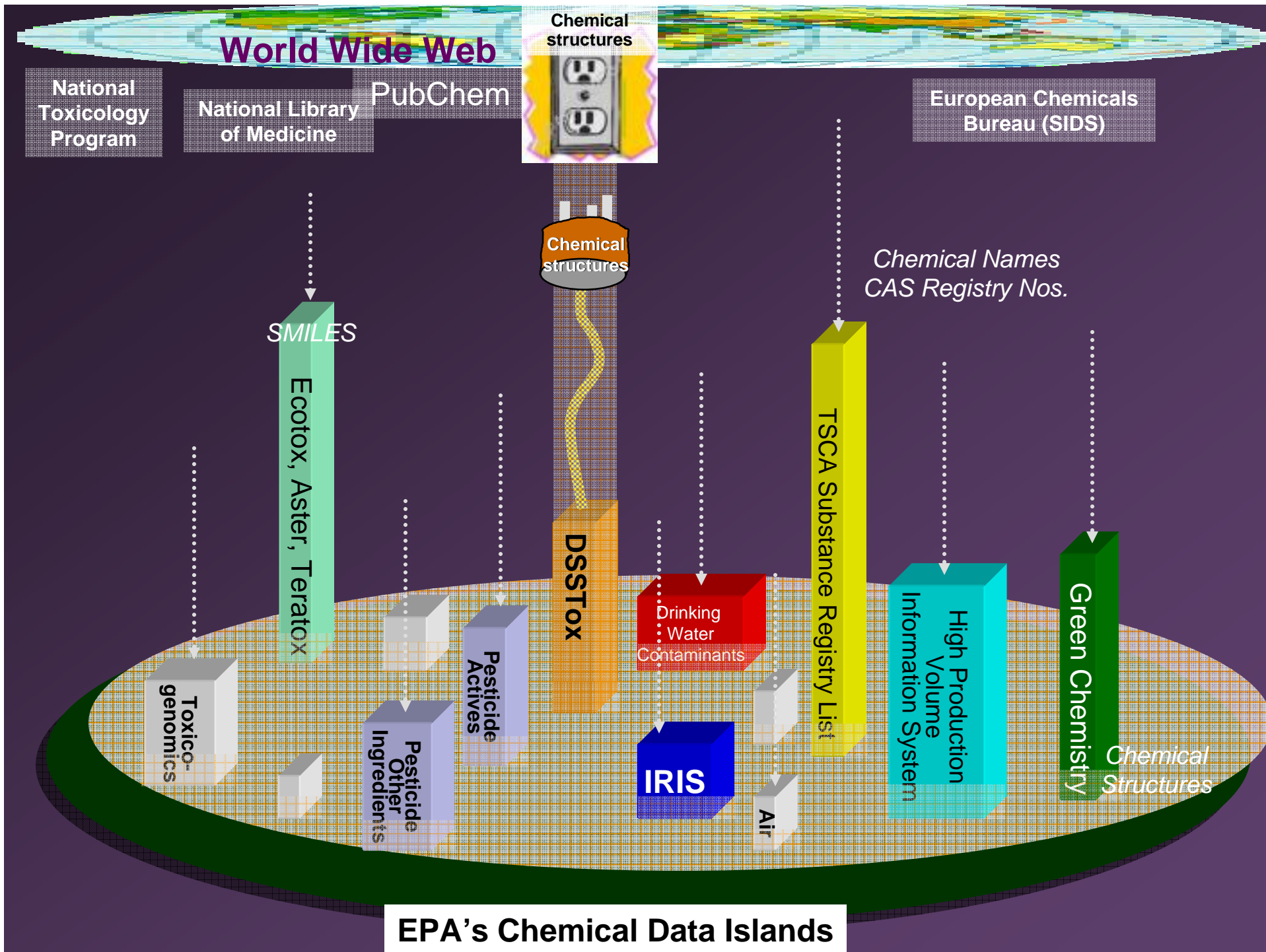
► *Chemical names are non-unique and error-prone*

► *CAS are proprietary and not necessarily unique*

► *No chemical similarity or analog searching*







Problems across EPA:

- ✦ Little or no chemical structure annotation
- ✦ No structure searching capabilities (Internet or desktop)
- ✦ No coordination of chemical structure information across EPA databases
- ✦ No standard quality review procedures for chemical information

EPA Chemoinformatics CoP Charter

- ✦ Facilitate communications, exchange information and experiences
- ✦ Evaluate and recommend software, tools
- ✦ Propose standards for chemical structure representation to enable cross compatibility of databases
- ✦ Guide creation of consolidated EPA database of chemical structure information

EPA Chemoinformatics CoP Roster (18Feb2006)

Co-Chairs

NCCT Ann Richard /RTP/USEPA/US
NHEERL, MED..... Chris Russom /DUL/USEPA/US

Members

ORD

NHEERL

Jack Jones/ATH/USEPA/US

NERL

Paul Winget /ATH/USEPA/US
Mitchell Kostich /CI/USEPA/US
Gary Walter /RTP/USEPA/US

NHSRC

Chandrika Moudgal /CI/USEPA/US

NCCT

Matt Martin /RTP/USEPA/US
Keith Houck /RTP/USEPA/US
David Dix /RTP/USEPA/US

NRMRL

Todd Martin /CI/USEPA/US
Paul Harten /CI/USEPA/US
William Barrett /CI/USEPA/US

NCEA

George Woodall /RTP/USEPA/US

OEI

John Harman /DC/USEPA/US
Jonathan Barney /R5/USEPA/US
Susan Lundquist /RTP/USEPA/US@EPA
Scott Burks /RTP/USEPA/US@EPA

OPPT

Richard Engler /DC/USEPA/US
Kent Anapolle /DC/USEPA/US
Kelly Mayo /DC/USEPA/US
Yintak Woo /DC/USEPA/US
Bob Boethling /DC/USEPA/US

OPP

Kerry Leifer /DC/USEPA/US
Pauline Wagner /DC/USEPA/US
Brian Montague /DC/USEPA/US
Alberto Protzel /DC/USEPA/US

Chemoinformatics CoP "TeamRoom":

Address https://epaqp.rtp.epa.gov/QuickPlace/copchemoinformatics/Main.nsf/h_Toc/4DF38292D748069D0525670800167212/?OpenDocument Go Links »

CoP Chemoinformatics Home

Ann Richard | Sign Out New... Edit Check Out... Copy Move Delete

EPA Warning Banner
Welcome
Discussion
Library
Calendar
Tasks
Index
Customize
Members

☒ Notify ☐ Print

Welcome to the Chemoinformatics - Community of Practice (COP)

The goals of the Chemoinformatics workgroup are to facilitate, coordinate and integrate efforts to address the challenges of chemical structure annotation (or indexing), retrieval, and mining of chemically-related data and documents, including newer toxicogenomics and metabonomics data, across EPA Program Offices, Labs and Centers.

***** Draft Charter & Membership Roster attached below**

Library

Title	Created By	Created On	Updated By	Modified On
Chemistry Freeware List of freeware tools available for use in managing and viewing chemical-structure information, including SD file viewers, chemical drawing programs, InChI generation and viewing capabilities. PowerMV A software environment for statistical analysis,...	Ann Richard	02/26/2006	Ann Richard	02/26/2006
EPA Licensed Software LeadScope Enterprise Data Mining: NCCT has have the client software installed. For more info...	Ann Richard	02/24/2006	Ann Richard	02/24/2006

On-line Chemistry Structure Browsers

PubChem Project - National Center for Biotechnology Information
<http://chem.sis.nlm.nih.gov/chemidplus/chemid>

On-line Chemistry Structure Browsers

Accord by Accelrys <http://www.accelrys.com/>
Information <http://pubchem.ncbi.nlm.nih.gov/>

Publications

CODDD_Richard_15Feb2006_rev_print.pdf
Chemical structure annotation and searching on the internet, with li...

Items 1 - 5 out of 5 (including hidden items)

Published Version | [Draft In Progress](#)

EPA Licensed Software

Created By: [Ann Richard](#) 02/24/2006 - 03:16 PM

Updated By: [Ann Richard](#) 02/24/2006 - 03:32 PM

LeadScope Enterprise Data Mining:

NCCT has purchased a 1 year site license for multiple simultaneous users. You must submit a request to the Computer Support Help Desk to have the client software installed. For more information on this software or to view introductory webcast presentations, visit the LeadScope website:

Main page: <http://www.leadscope.com/>

Demo webcasts on wide range of topics: http://www.leadscope.com/about_us_demo.php

LeadScope FDA CDER 2006 Chronic/Subchronic Database (84 drugs w/352 tests on rat, mice, oral exposures)

NCCT has purchased a 1 year site license for this database, which will be installed on the EPA LeadScope Enterprise server application. Hence, anyone with access to LeadScope will automatically be able to access this application. The chemical content in this initial version is somewhat limited (only 84 compounds, all drugs) but the value of the database is its depth, i.e. the amount of data per chemical and the ability to relationally explore and mine these data.

Scitegic Pipeline Pilot:

NCCT has purchased a 1 year site license for multiple simultaneous users. You must submit a request to the Computer Support Help Desk to have the client software installed. For more information on this software or to view introductory presentations, see attached files below (Exercise files are meant to be viewed from within Pipeline or with an xml viewer) or visit the Scitegic website.

<http://www.scitegic.com/>

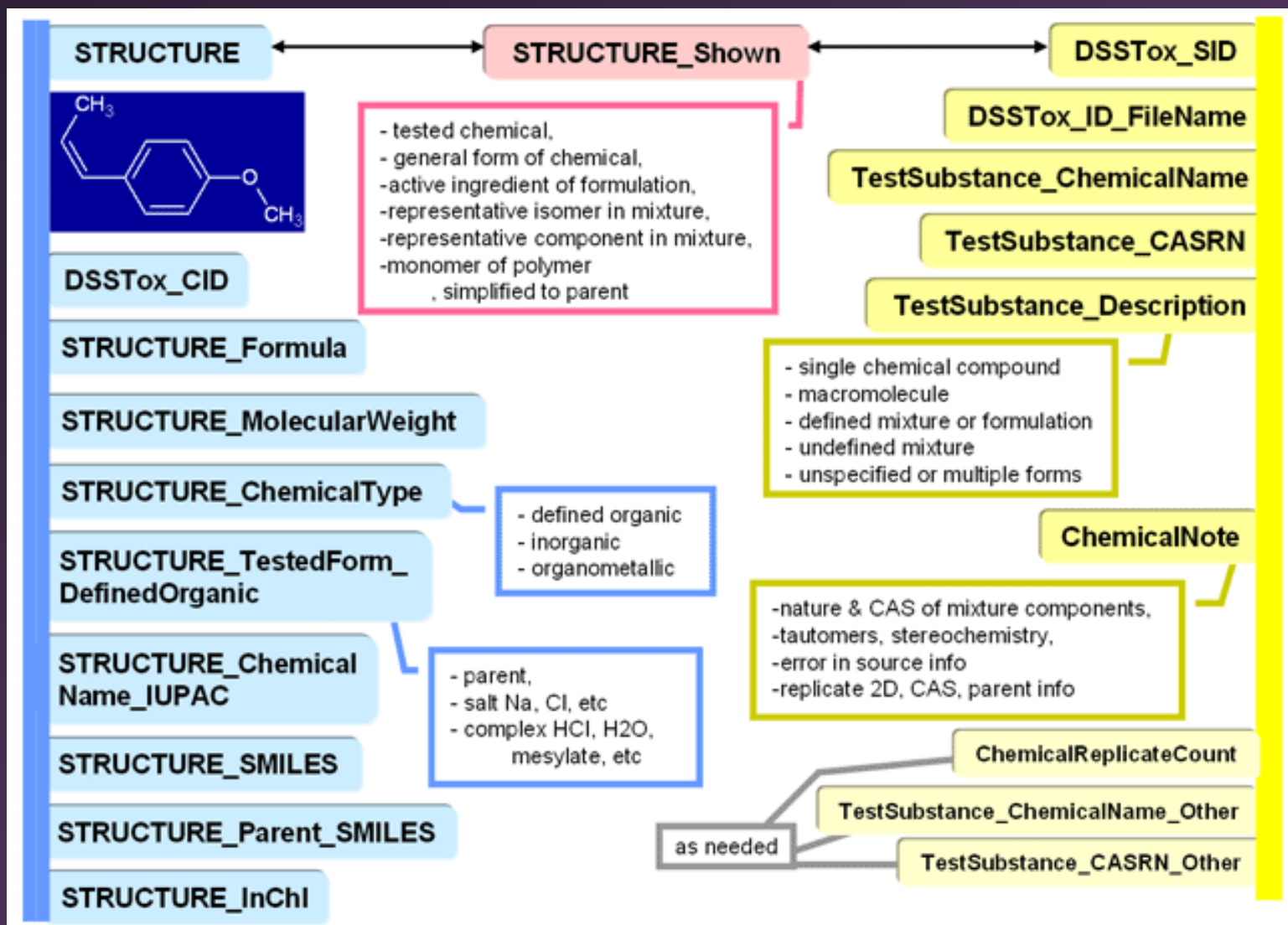
<http://www.scitegic.com/product>

File Name	Size	Modified
Fundamentals of Pipeline Pilot - Chem 5.ppt	6325KB	1/30/2006 2:14:36 PM
Exercise1.xml	113KB	1/26/2006 4:37:50 PM
Exercise10.xml	40KB	1/26/2006 4:37:50 PM
Exercise2.xml	106KB	1/26/2006 4:37:50 PM

Propose DSSTox Project be used to advance CoP objectives:

- ✦ Chemical structure & information standards
- ✦ Standard quality review procedures for chemical information
- ✦ Extensive collaboration network
- ✦ Central DSSTox Master Structure-Index File
- ✦ Public Internet website – information portal

DSSTox Chemical Standards



8804 total records

DSSTox Master Structure-Index File

Carcinogenic Potency Database
 EPA Disinfection By-products Cancer Predictions ✨
 EPA Fathead Minnow Acute toxicity ✨
 NCTR Estrogenic Activity
 FDA Drug Maximum Daily Dose
 EPA High Production Volume SI ✨
 NTP Bioassay SI
 EPA IRIS SI ✨
 NTP High Throughput Screening SI
NTP Immunotoxicity Testing Battery
DEMETRA Pesticides Ecotoxicity ✨
NTP Genetic Toxicity - Zeiger
NCTR Androgenic Activity
CEBS SI
ICCVAM Endocrine Disruption Set SI
EPA Pesticide ToxCast Candidates SI ✨
EPA Developmental Neurotox SI ✨
EPA Pesticide Active Ingredients SI ✨
EPA Drinking Water Contaminants SI ✨
EPA Pesticide Other Ingredients SI ✨

Published
Databases

Published
Structure-Index Files

Databases
in development

Unpublished
Structure-Index Files

#	NAMEID	DSSTox Master_v1a
1	CPDBAS	v3b_1481
2	DBPCAH	v3b_209
3	EPAFHM	v3b_617
4	IICTRER	v3b_232
5	FDAMDD	v2b_1217
6	HPVCSI	v1a_3548
7	IITPBSI	v1a_2415
8	IRISSI	v1a_544
9	IITPHTS	v1a_1408
10	IMMTOX	v1a_87
11	ECODEM	v1a_399
12	IITPGTZ	v1a_1931
13	IICTRAR	v1a_202
14	CEBSSI	v1a_20
15	ICCVAM	v1a_87
16	EPAPTC	v1a_826
17	EPADIIT	v1a_82
18	EPAPAI	v1a_873
19	EPADWC	v1a_66
20	EPAPOI	v1a_441

Chemoinformatics CoP Progress To-Date:

DSSTox

- ▶ DSSTox Master File and several major EPA Data Files published
- ▶ Automated QA procedures for chemical information review
- ▶ Major update to DSSTox website content (13Jun2006)
- ▶ Initiated project to develop a public EPA Chemical Database & Structure-Browser

CoP

- ▶ Draft charter approved
- ▶ Membership roster spanning EPA
- ▶ EPA Team Room created
- ▶ Introductory conference call & emails
- ▶ Coordinating chemical data management software purchase with OPPT

Next Steps:

- ◆ Expand CoP membership beyond EPA
- ◆ Training, improved communication
- ◆ Bring in software
- ◆ Agency-wide structure browser
- ◆ Expand collaborations

Chemoinformatics CoP:

EPA

IRIS
EcoTox
Pesticides
TSCA Inventory
Green Chemistry
Drinking Water Contaminants
Pollution Prevention & Toxics
ORD Labs & Centers

NIH/NIEHS:
National Toxicology Program

NTP High-throughput
Screening Program

NIH/NCGC Molecular
Libraries Roadmap
Initiative

PubChem

ATSDR

NCI

FDA

European
Chemicals Bureau

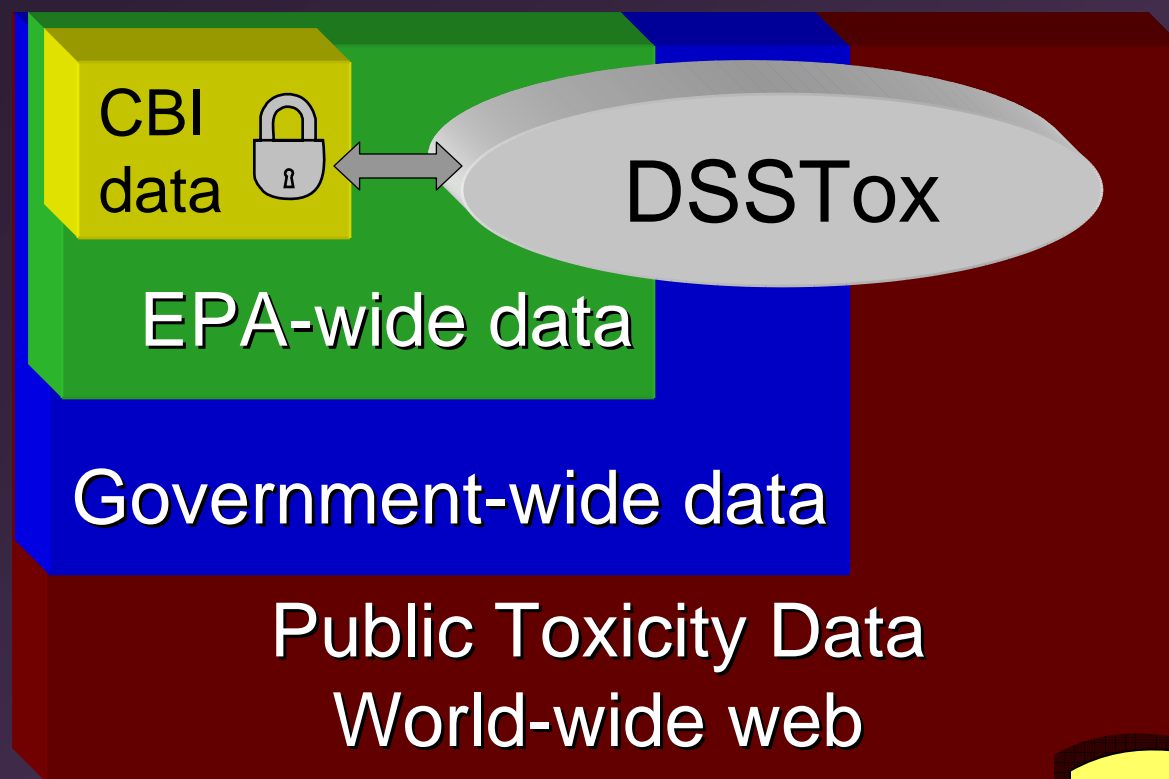
Joint Research
Council

Proctor &
Gamble

ACD
Labs

Leadscope

EPA's "Structure-Indexed" Future:



Standardized
chemical structure
indexing of EPA data

Structure-searching &
viewing capabilities

Full integration with
Internet chemical
resources

**Improved Toxicity Data
Mining & Prediction
Capabilities**