EPA
United States
Environmental Protection
Agency

Learning by Doing – Phase I of ToxCast

Robert Kavlock
Director, NCCT

Office of Research and Development
Computational Toxicology Research Program

OpenTox, Potsdam, Germany
May 30 2010I

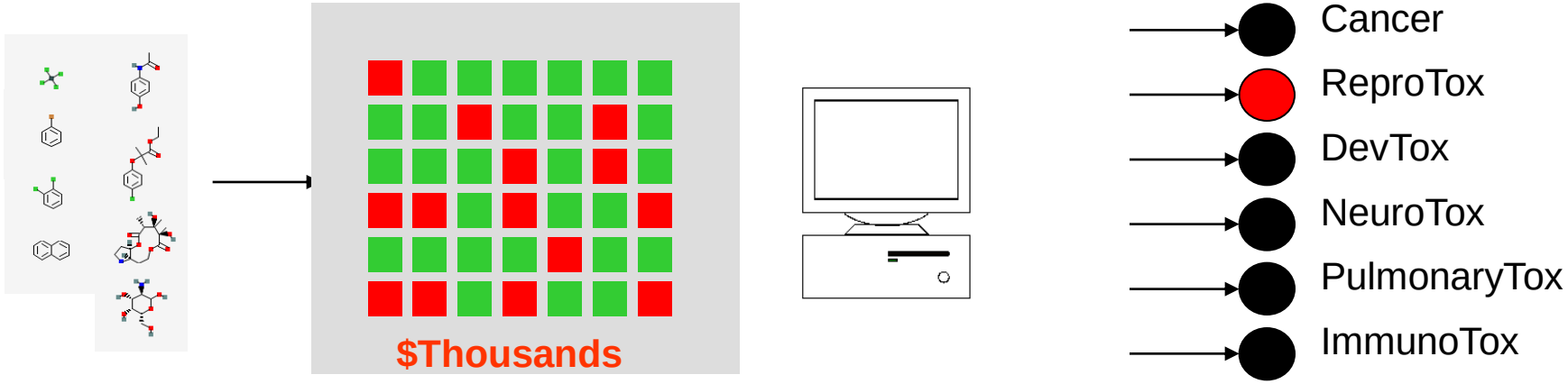
United States Environmental Protection Agency

"...to integrate modern computing and information technology with molecular biology to improve Agency prioritization of data requirements and risk assessment of chemicals"

Providing Decision Support Tools for High-Throughput Screening, Risk Assessment and Risk Management

ToxCast Bioactivity Profiling

in vitro testing *in silico* analysis

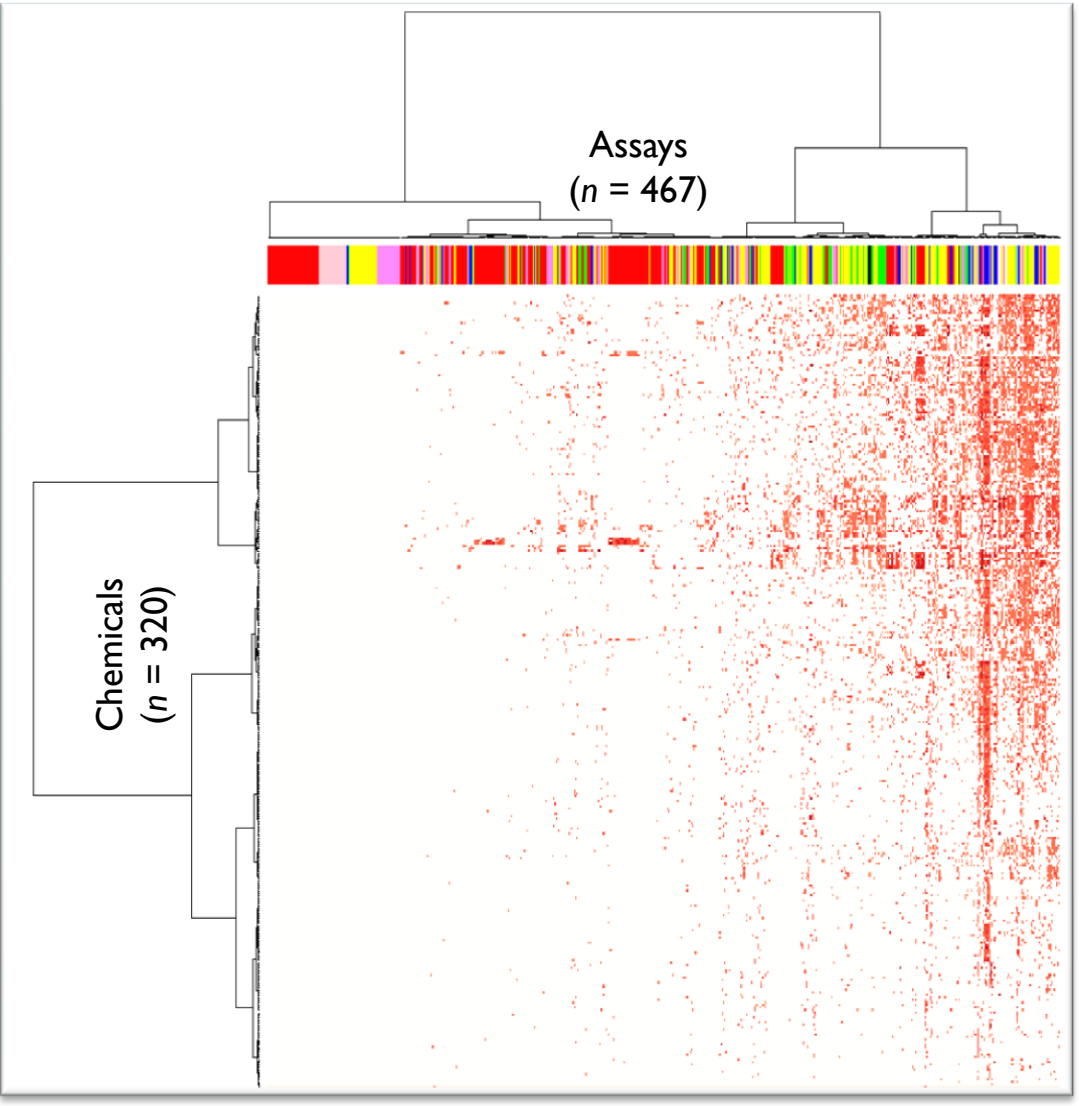


HTS
-omics

Bioinformatics/
Machine Learning

Dix *et al*, Tox Sci 95:5-12 (2007)
www.epa.gov/ncct/toxcast

ToxCast Phase I HTS Results



Cellular Assays

- **Cell lines**
  - HepG2 human hepatoblastoma
  - A549 human lung carcinoma
  - HEK 293 human embryonic kidney
- **Primary cells**
  - Human endothelial cells
  - Human monocytes
  - Human keratinocytes
  - Human fibroblasts
  - Human proximal tubule kidney cells
  - Human small airway epithelial cells
- **Biotransformation competent cells**
  - Primary rat hepatocytes
  - Primary human hepatocytes
- **Assay formats**
  - Cytotoxicity
  - Reporter gene
  - Gene expression
  - Biomarker production
  - High-content imaging for cellular phenotype

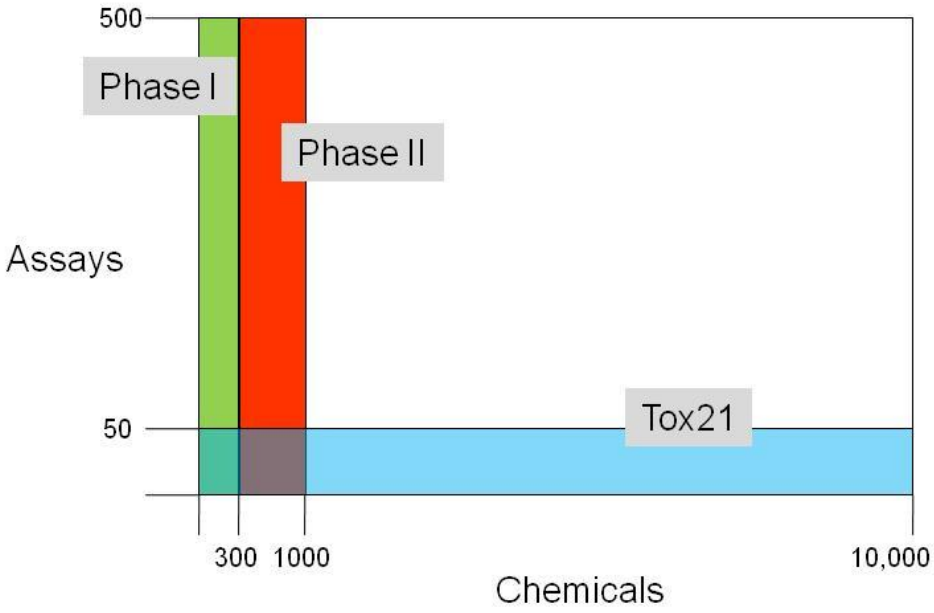
Biochemical Assays

- **Protein families**
  - GPCR
  - NR
  - Kinase
  - Phosphatase
  - Protease
  - Other enzyme
  - Ion channel
  - Transporter
- **Assay formats**
  - Radioligand binding
  - Enzyme activity
  - Co-activator recruitment

Judson *et al* EHP 118:485-492 (2010)
http://www.epa.gov/ncct/toxcast/

From Phase I to Phase II and Tox21

	Phase I	Phase II	Tox21
Actives	272	120	700
Inerts	24	100	1000
Antimicrobials	33	100	500
HPV	35	170	1300
MPV	7	60	1500
Green	4	60	500
PCCL	73	150	500
Pharmaceuticals	0	100	2500
Consumer Products /Food additives	0	0	1500
Total	**309**	**~700**	**~10000**



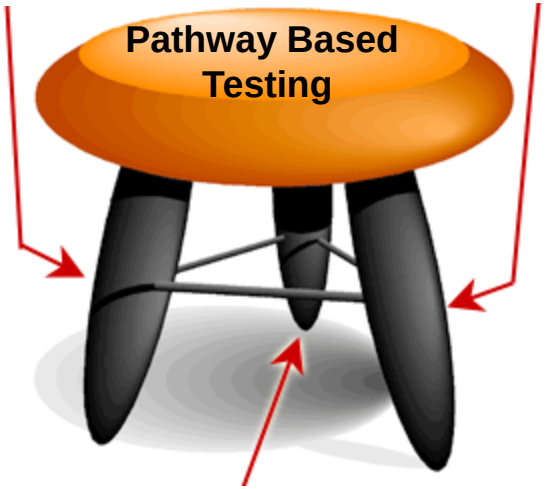
Overview

- Information Management
- Transparency
- Unexpected Results
- Team Effort

Enabling the Transformation in Toxicology

High Throughput Screening

Information Technology and Management

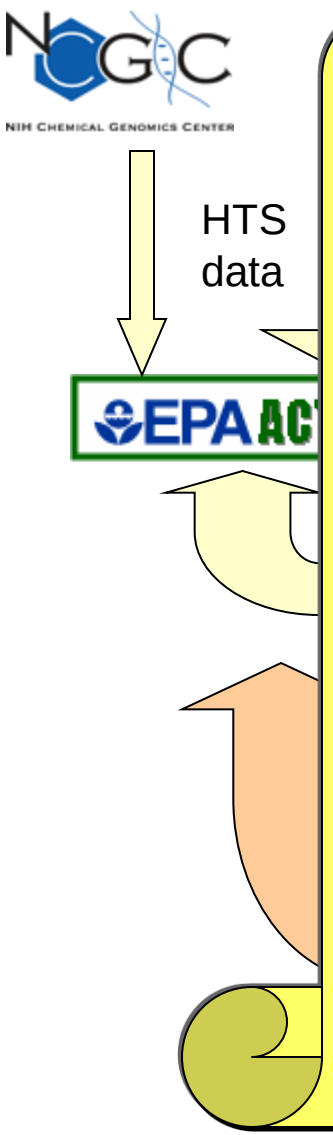


Molecular, Cellular and Systems Biology

Databases & Tools

- DSSTox – Distributed Structure Toxicity Data
- ACToR – Aggregated Computational Toxicology Database
- ToxRefDB – In vivo guideline study data
- ToxMiner – ToxCast data and analyses
- EDB – NCCT Endocrine Database
- VTKB – Virtual Tissue Knowledgebase
- Chemical Registry

ToxCast/Tox21: Chemical Annotation & Data Publication



Cheminformatics Needs:

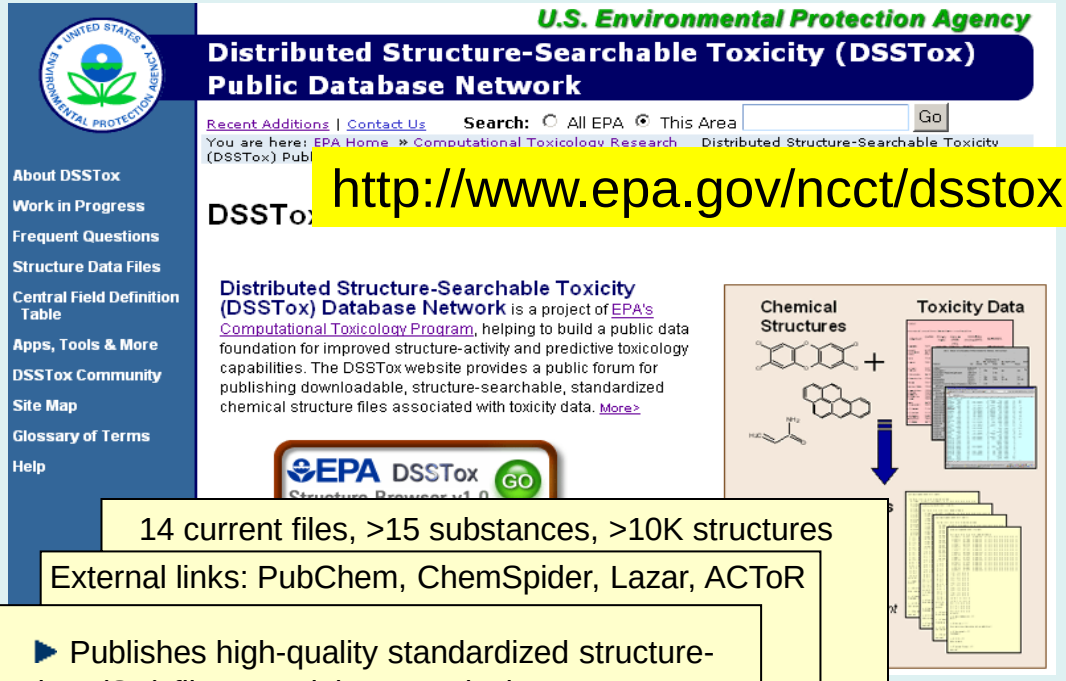
- Select chemicals for testing
  - ✓ high environmental / tox interest
  - ✓ suitable for testing
  - ✓ chemically diverse
  - ✓ analogs, metabolites, etc
- Standardized chemical structure/substance annotation across all testing & reference inventories
- Chemical annotation QC
- Analytical QC
- Public reporting of data & QC

nnotation
C
king

rized
t data for
SAR
g

DSSTox: Distributed Structure-Searchable Database Network Project…& ToxCast™

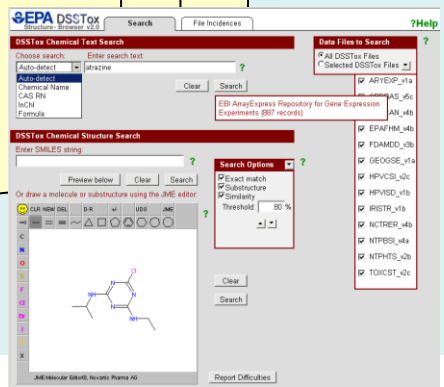
United States Environmental Protection Agency



14 current files, >15 substances, >10K structures

External links: PubChem, ChemSpider, Lazar, ACToR

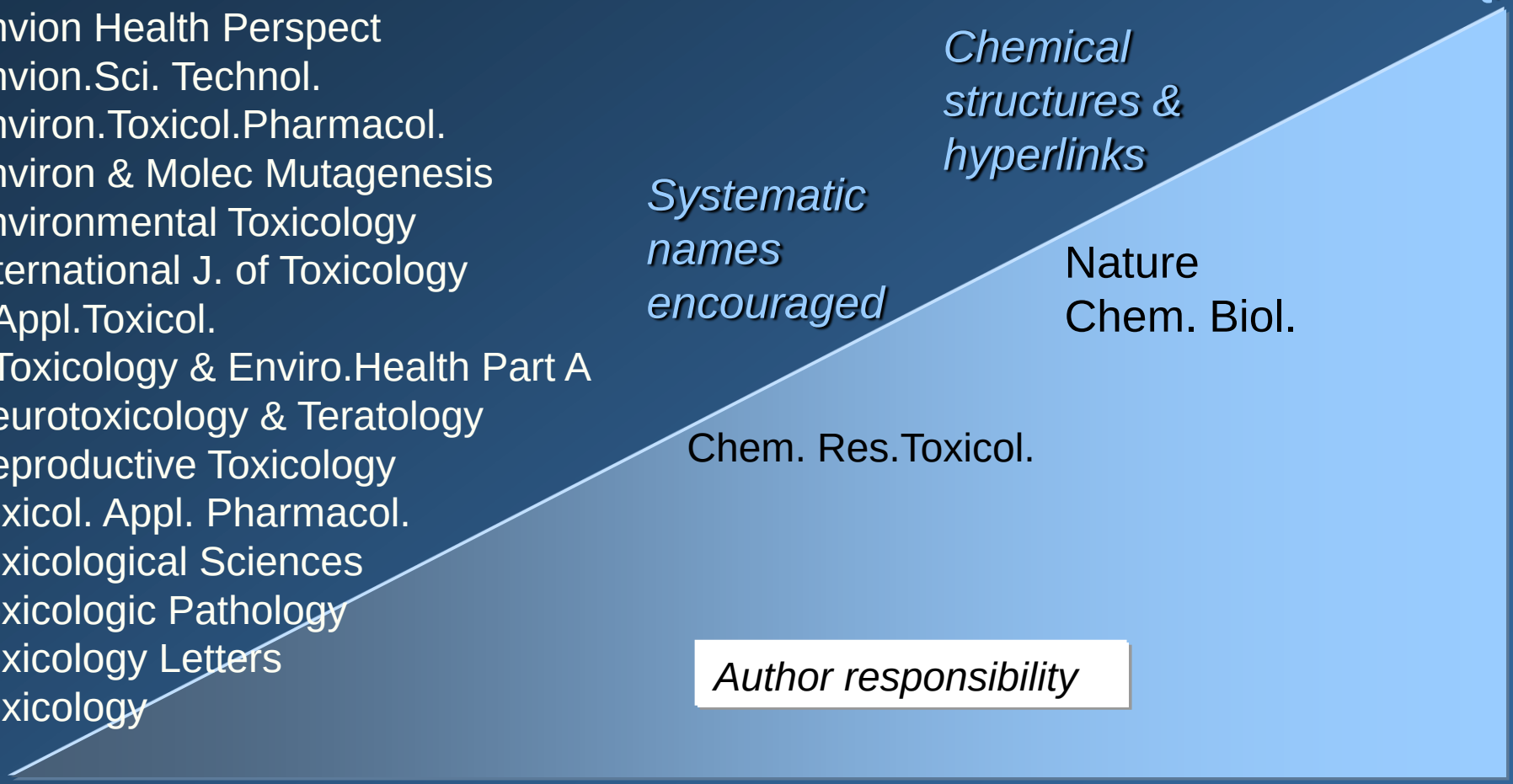
- Publishes high-quality standardized structure-data (SD) files pertaining to toxicology:
  - *EPA, HPV-IS, IRIS, NTP, FDA, NCBI, EBI, …*
- Strict chemical QA procedures
- Public substance/structure ID registry system
- DSSTox Structure-Browser
  - *Search by structure, name, CAS, …*
  - *Similarity & fragment search*
  - *Search within a DSSTox file*
  - *Link-in and link out of browser*



- Chemical information QA and structure annotation
  - *ToxCast Phase I & II*
  - *ToxRefDB*
  - *Tox21*
- Primary source of QC'd structures for ACToR
- Facilitate external linkages and data publication
- Publish summary activities and chemical classifiers for modeling
- PubChem Source depositor for structures & "assays"

Current Guidelines & Chemical Standards in Toxicology & Biological Effects Publishing:

Archives of Toxicology
Brain Research Bulletin
Ecotoxicol. Environ. Saf.
Envion Health Perspect
Envion.Sci. Technol.
Environ.Toxicol.Pharmacol.
Environ & Molec Mutagenesis
Environmental Toxicology
International J. of Toxicology
J.Appl.Toxicol.
J Toxicology & Enviro.Health Part A
Neurotoxicology & Teratology
Reproductive Toxicology
Toxicol. Appl. Pharmacol.
Toxicological Sciences
Toxicologic Pathology
Toxicology Letters
Toxicology



"Authors should provide sufficient detail to allow the work to be reproduced."

nature chemical biology

Editorial

Nature Chemical Biology **3**, 297 (2007)
doi:10.1038/nchembio0607-297

A new look for chemical information

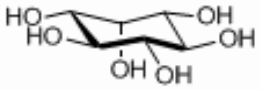
***Nature Chemical Biology* is committed to enhancing interdisciplinary communication and features online content to increase the accessibility of chemical information for our readers.**

"**Chemical compound information**
For all significant compounds included in Nature Chemical Biology original research papers, a compound data page, linked directly from the compound reference in the full text, appears in the online journal.

Compound Data Index

From the following article
Small molecules enhance autophagy and reduce toxicity in Huntington's disease models
Sovan Sarkar, Ethan O Perlstein, Sara Imarisio, Sandra Pineau, Axelle Cordenier, Rebecca L Maglathlin, John A Webster, Timothy A Lewis, Cahir J O'Kane, Stuart L Schreiber & David C Rubinsztein

Nature Chemical Biology **3**, 331-338 (2007) Published online: 7 May 2007
doi:10.1038/nchembio883

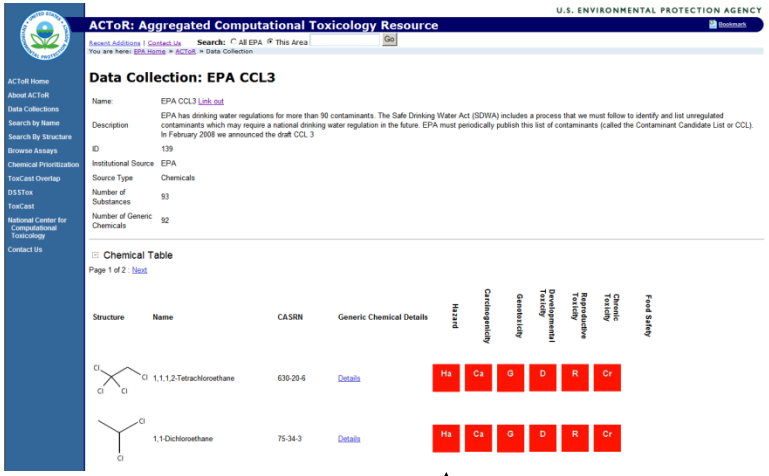


Compound 1
Inositol
View in PubChem
View compound page (2 KB) | View in 3D (2 KB) | Download ChemDraw file of structure (2 KB)

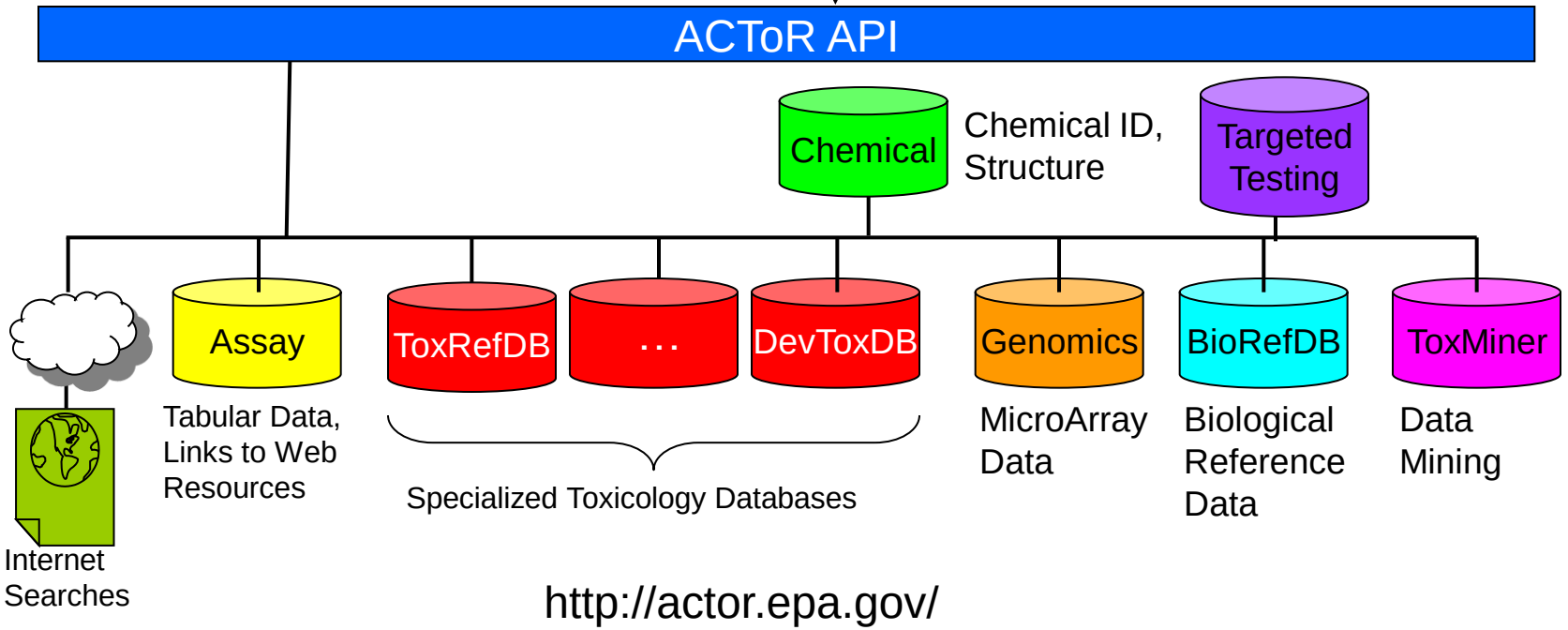
A PubChem link allows users of Nature Chemical Biology to go, in a single click, from the mention of a molecule in a paper to a rich and growing collection of information about chemical structures and their biological assay results, hosted by the NCBI.

"Authors ensure that the chemical compound information within their papers is complete, scientifically accurate and appropriately formatted."

ACToR
Aggregated Computational Toxicology Resource



ACToR Web Browser



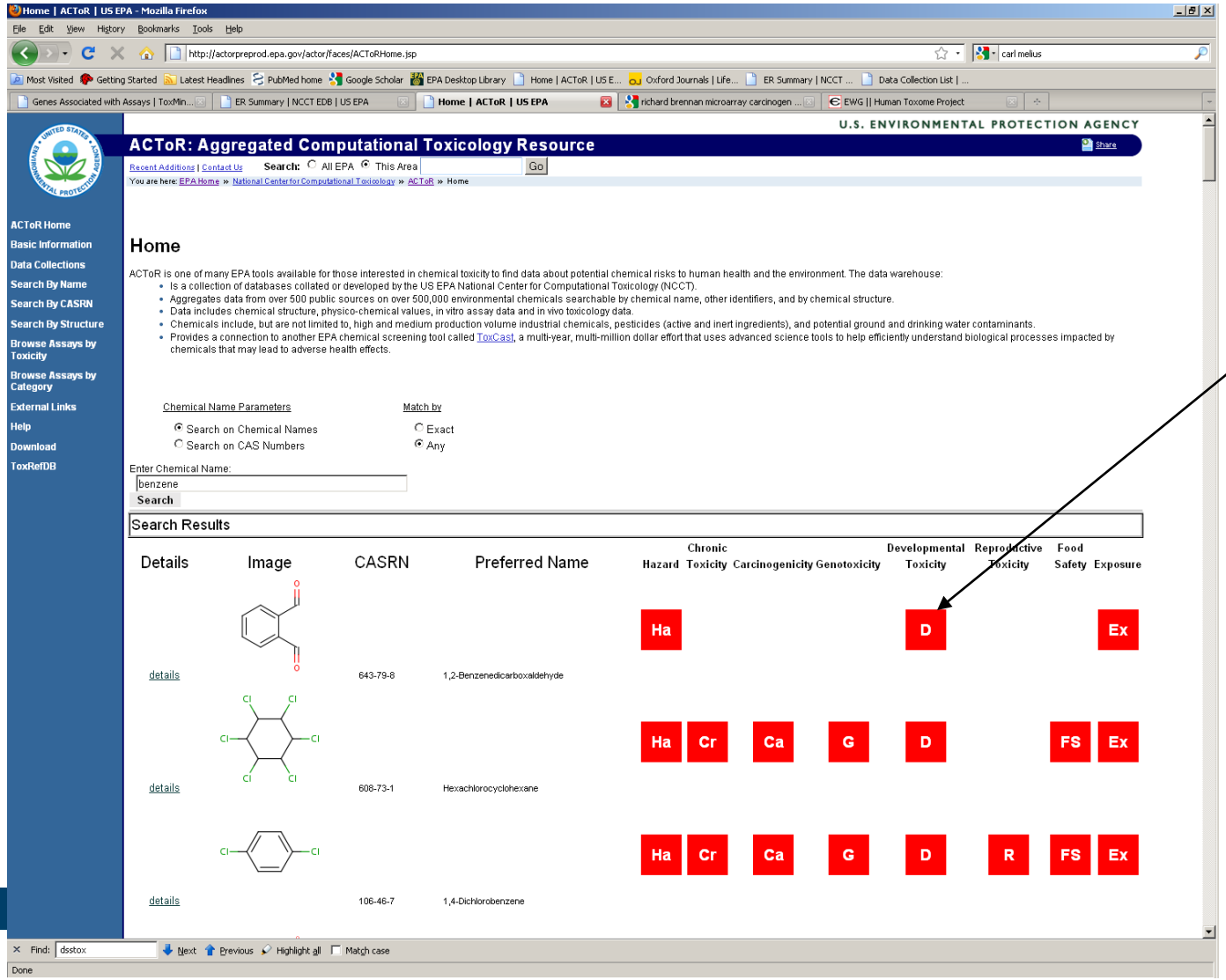
http://actor.epa.gov/

The Reach of ACToR

Category	Count
Data Collections	580
Substances	2,139,297
Compounds	1,044,834
Generic Chemicals	546,956
Generic Chemicals with Structure	438,756
Assays	3,213
Assay Components	7,221
Assay Results	6,662,296

As of May 2010

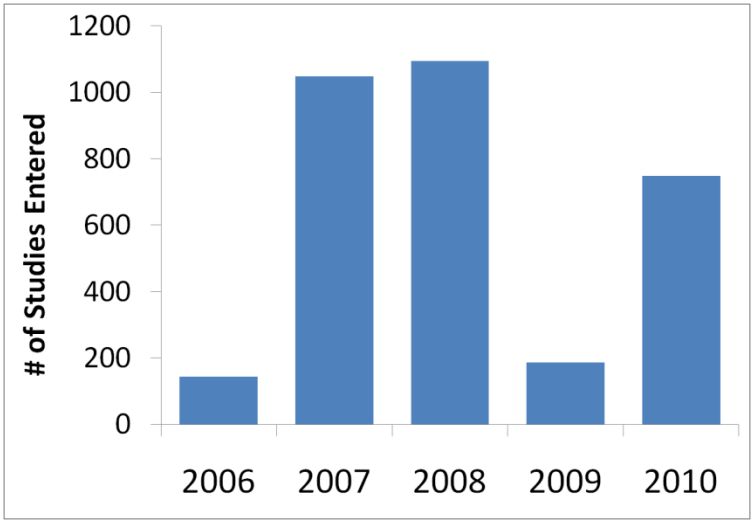
Chemical List View

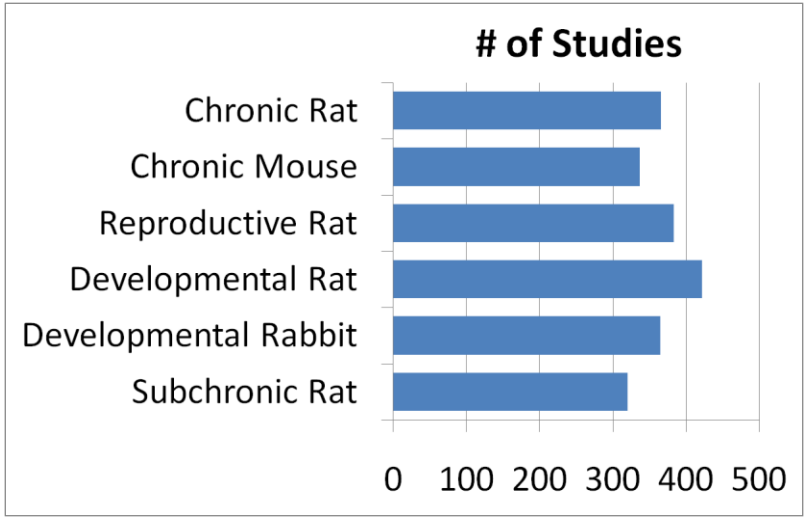


Red box indicates that data is available for that phenotype, not that chemical causes that phenotype

ToxRefDB

- Developed in 2006
  - Database Structure
  - Database Vocabulary
- Manual entry of >3200 studies for > 500 chemicals

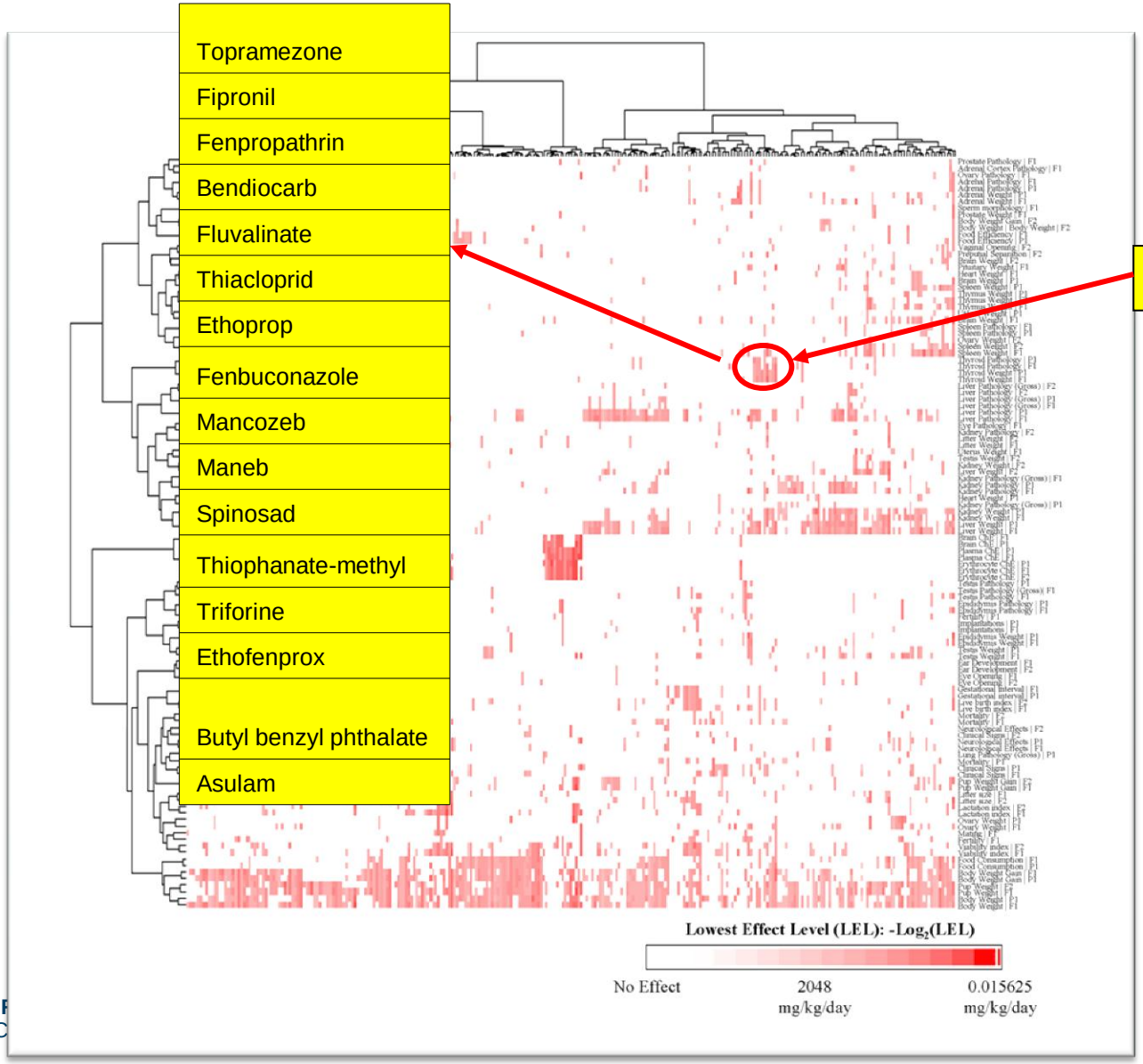




- Datasets released publically in 2009 (3 publications)
- Full database & Web Interface released publically in 2010

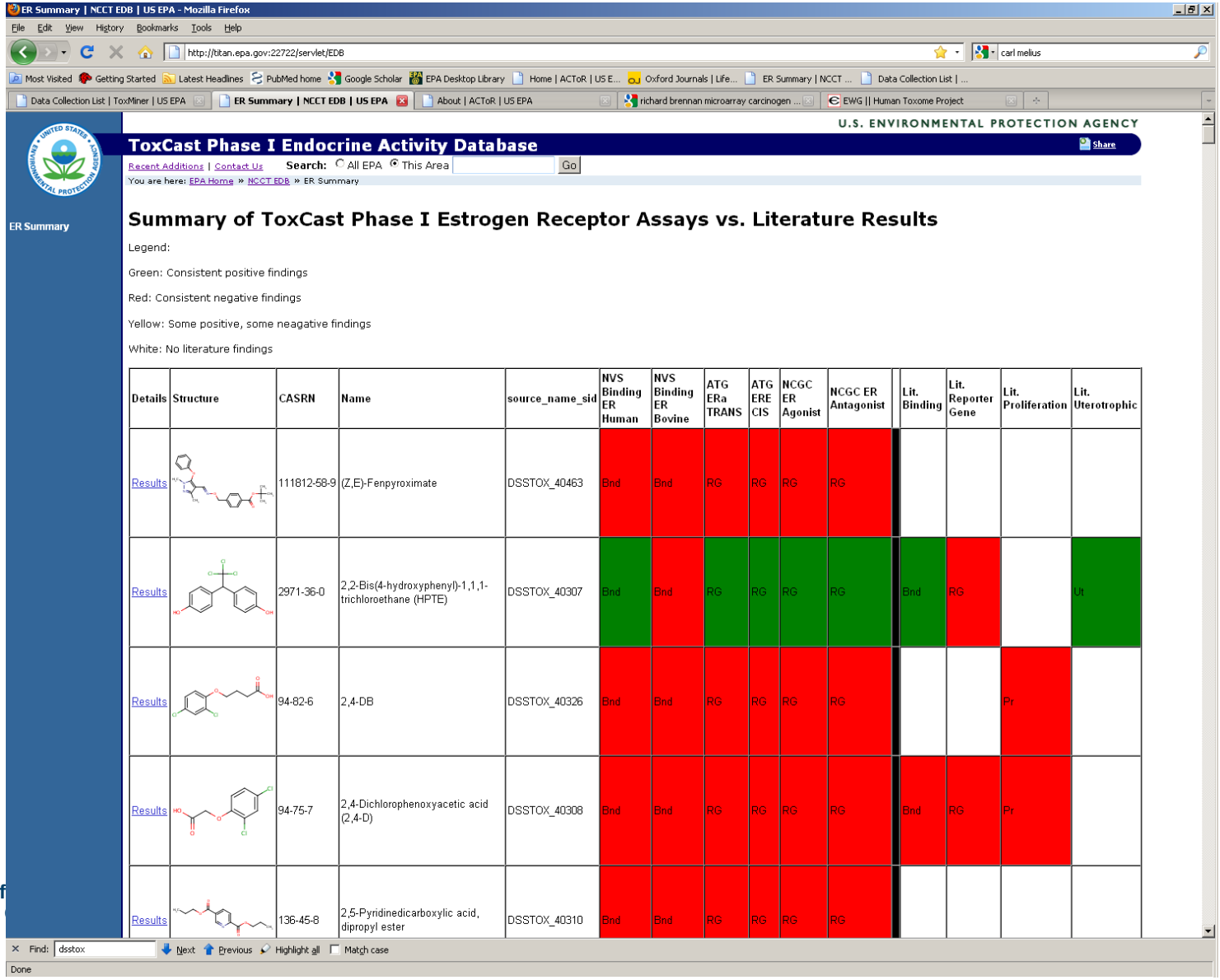
ToxRefDB: A Computable Database

United States Environmental Protection Agency



Martin *et al*, *ToxSci*, 110:181-190 (2009)

NCCT Endocrine Database



ER Summary | NCCT EDB | US EPA - Mozilla Firefox

http://titan.epa.gov:22722/servlet/EDB

U.S. ENVIRONMENTAL PROTECTION AGENCY

ToxCast Phase I Endocrine Activity Database

Recent Additions | Contact Us Search: All EPA This Area Go

You are here: EPA Home » NCCT EDB » ER Summary

ER Summary

Summary of ToxCast Phase I Estrogen Receptor Assays vs. Literature Results

Legend:

Green: Consistent positive findings

Red: Consistent negative findings

Yellow: Some positive, some neagative findings

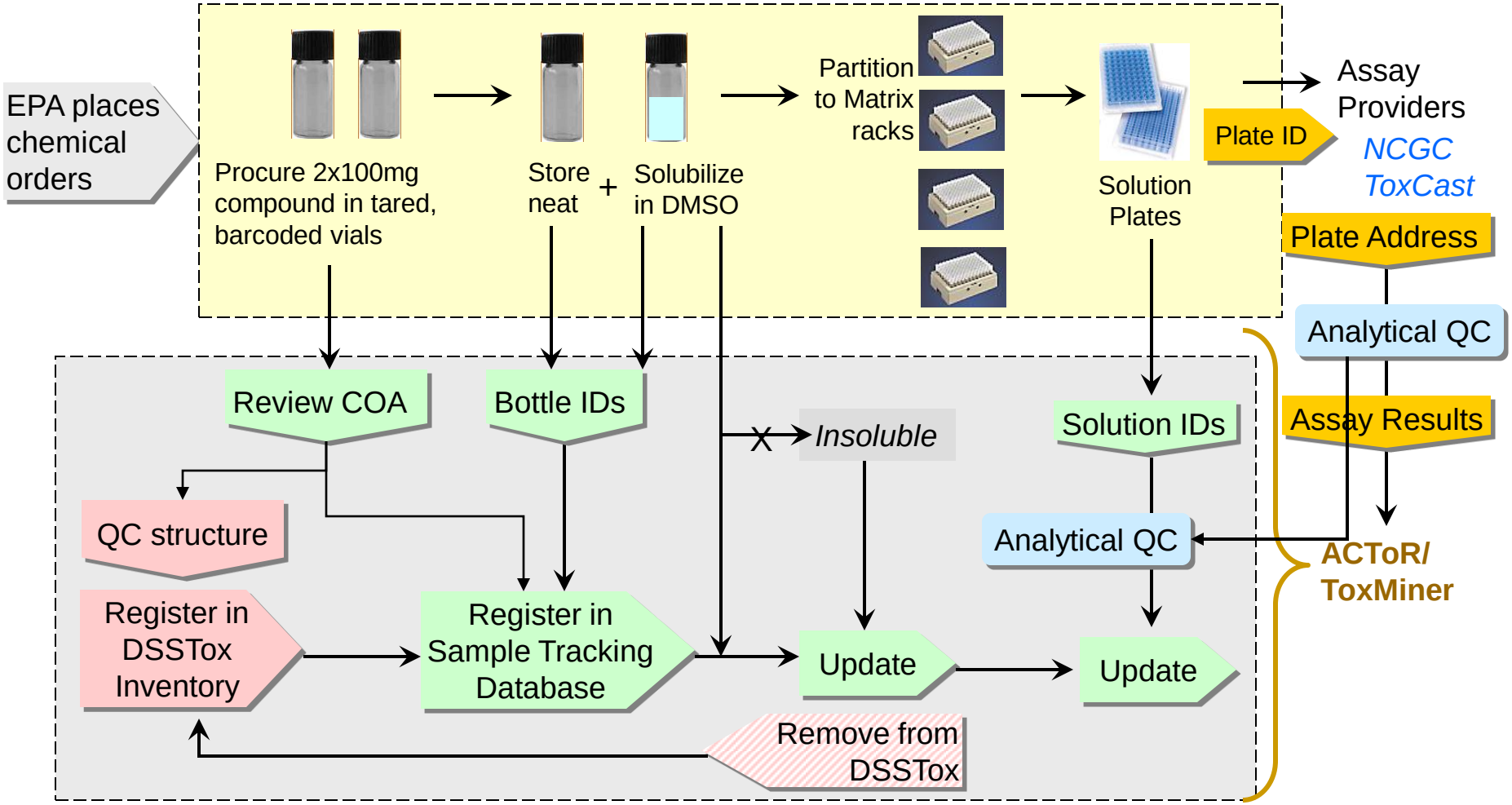
White: No literature findings

Details	Structure	CASRN	Name	source_name_sid	NVS Binding ER Human	NVS Binding ER Bovine	ATG ERa TRANS	ATG ERE CIS	NCGC ER Agonist	NCGC ER Antagonist	Lit. Binding	Lit. Reporter Gene	Lit. Proliferation	Lit. Uterotrophic
Results		111812-58-9	(Z,E)-Fenpyroximate	DSSTOX_40463	Bnd	Bnd	RG	RG	RG	RG				
Results		2971-36-0	2,2-Bis(4-hydroxyphenyl)-1,1,1-trichloroethane (HPTE)	DSSTOX_40307	Bnd	Bnd	RG	RG	RG	RG	Bnd	RG		Ut
Results		94-82-6	2,4-DB	DSSTOX_40326	Bnd	Bnd	RG	RG	RG	RG			Pr	
Results		94-75-7	2,4-Dichlorophenoxyacetic acid (2,4-D)	DSSTOX_40308	Bnd	Bnd	RG	RG	RG	RG	Bnd	RG	Pr	
Results		136-45-8	2,5-Pyridinedicarboxylic acid, dipropyl ester	DSSTOX_40310	Bnd	Bnd	RG	RG	RG	RG				

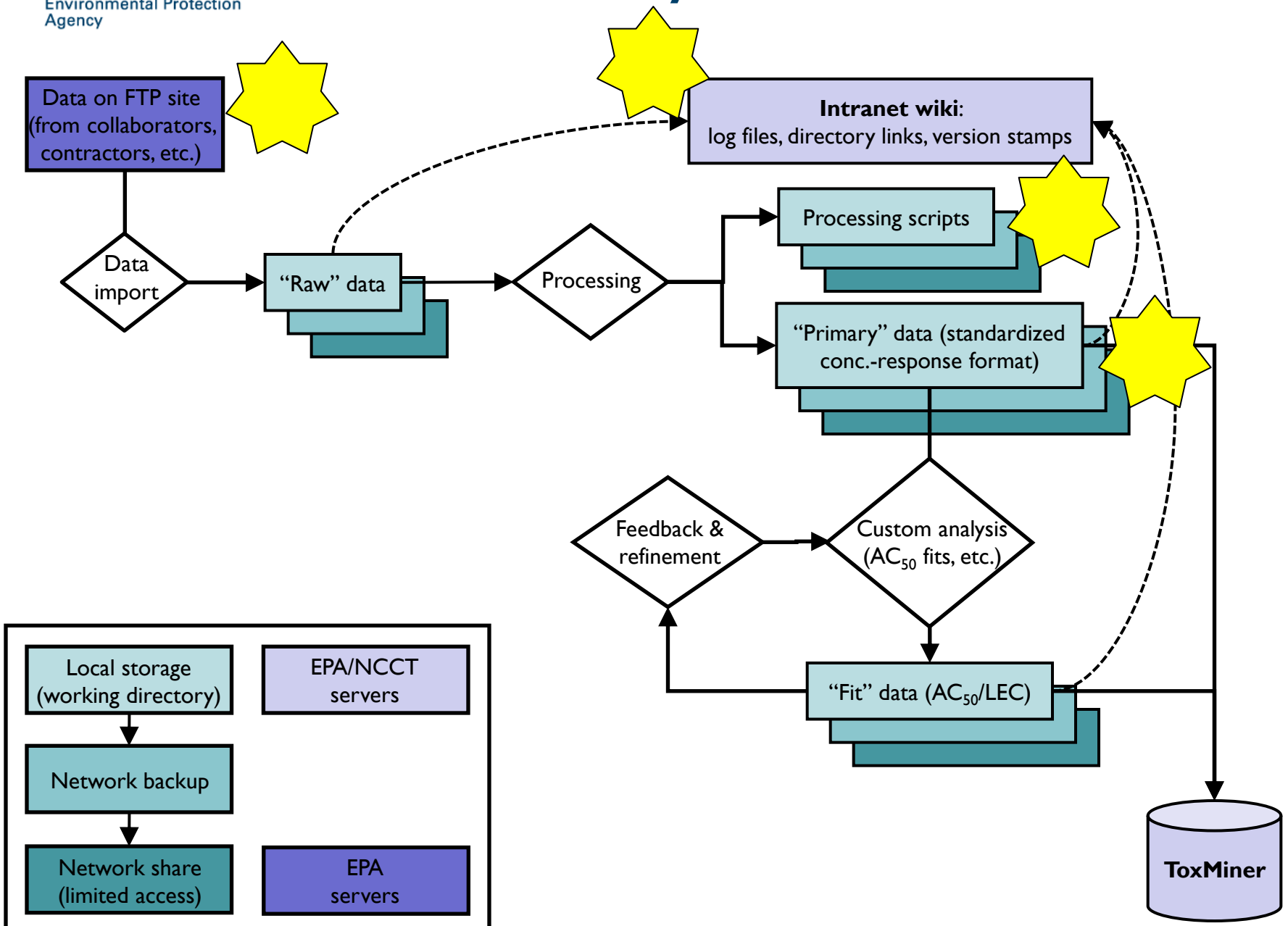
Find: dsstox Next Previous Highlight all Match case

Done

Chemical Sample Registration Workflow



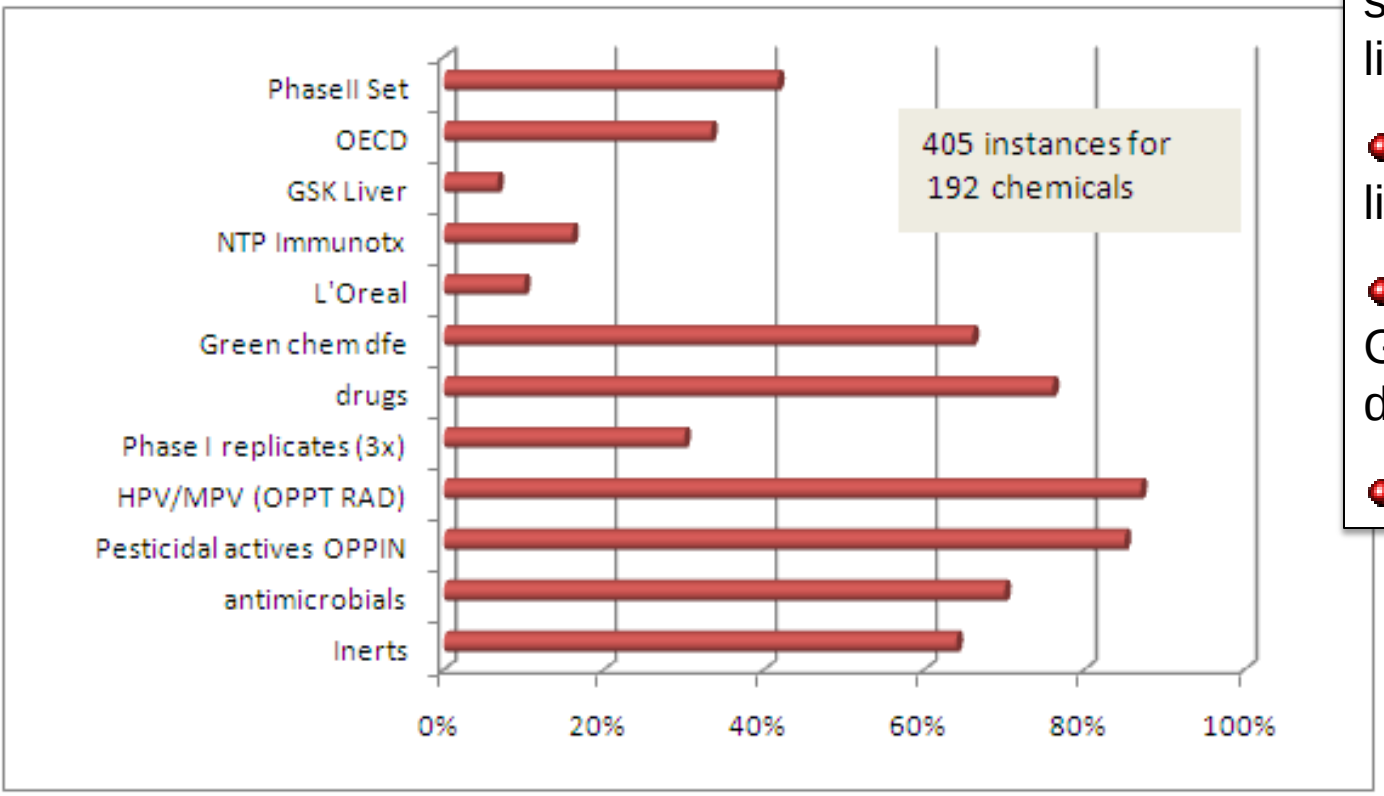
ToxCast Assay Data Workflow: Phase-II



Tox21 & ToxCast Phase II Chemicals

- 3800 chemicals “On hand”, registered or in process
- Solubilize_1 set (1616 compounds): 135 DMSO insolubles
- Solubilize_2 set (1300 compounds): submitted 5/17/2010
- 114 pharma compounds (failed drugs) received and solubilized from Merck, GSK, Pfizer, Sanofi-Aventis
- 192 chemicals from Solubilize_1 set + 114 pharma compounds plated for Phase II
- 10 Phase II triplicates “On hand”, 3 included in 192

ToxCast Phase II Chemical Selection

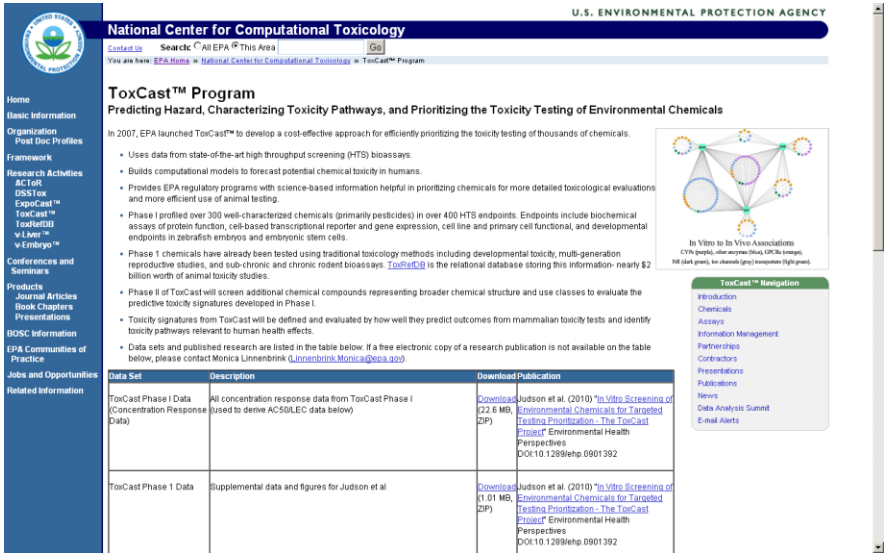


- Phase II chemicals being selected from Tox21 library
- High-priority EPA lists + In vivo Data
- Pfizer, Merck, GSK chem w/clinical data (approx 110)
- Defined structures

Updated May 15 2010

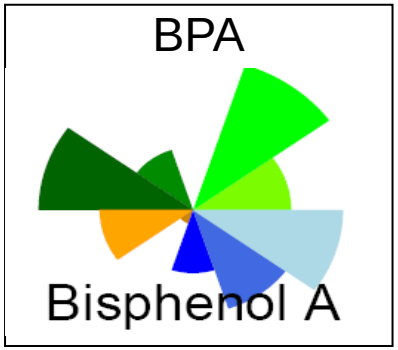
Transparency

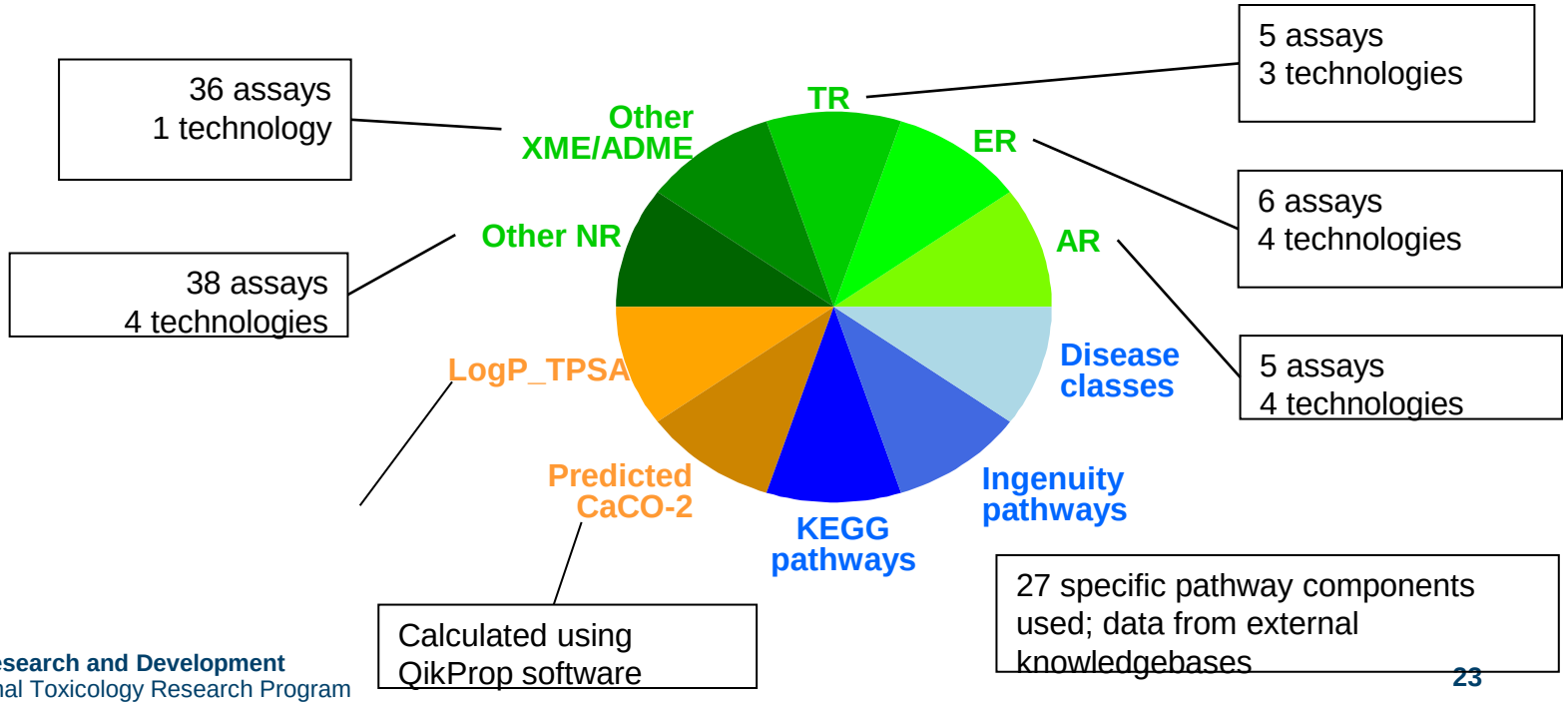
- Website
- Full data release
- Use of open source software
  - R, MySQL, PERL
- Chemical Prioritization Community of Practice
- ToxCast Data Analysis Summit
- Communications
  - Presentations, presentations, presentations
  - Media releases



Unexpected Results

- Breakdown of materials in DMSO
  - Need for constant quality assurance
- Need to orthogonal methods
  - False positives and false negatives
  - Advantage of multiple platforms





United States Environmental Protection Agency

Data and Tool Interoperability

Data and Knowledgebases

ToxRefDB

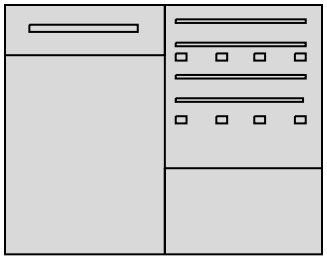
ACToR

ExpoDB

Virtual Tissue KB

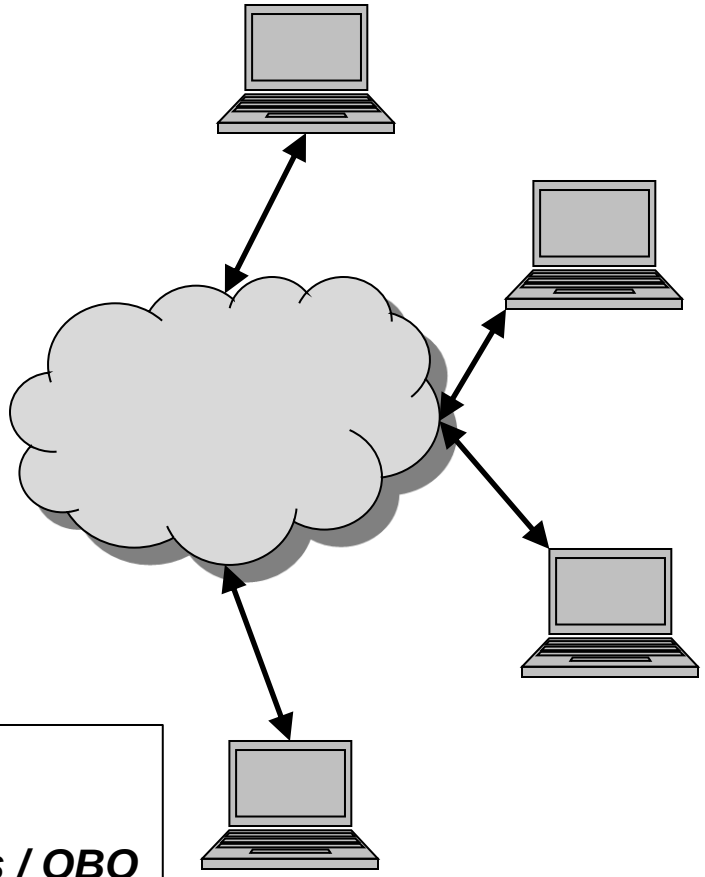
Common Ontologies

Disease, ADR
Assay
Physiology



Analysis Tools

Web Pages
Web Services



OpenTox API

Open Ontologies / OBO

Open Source Tools