CEBS Database: a structure for data integration across studies

26 June 2013

Chemical Effects in Biological Systems (CEBS)

- Relational public database
  - User interface
  - FTP site with data domain “bundles”
- Private database + file system (Bins)
- Owner is the US National Toxicology Program. Committed to the use of CEBS to store and serve to the public data from NTP studies.
- CEBS: http://cebs.niehs.nih.gov

Example 1

- See all studies in CEBS related to a test article (Toluene)
- How studies are organized in CEBS
- How to access study-level data

National Institute of Environmental Health Sciences

www.cebs.niehs.nih.gov

Search GO

HOME | HEALTH & EDUCATION | RESEARCH | FUNDING OPPORTUNITIES | CAREERS & TRAINING | NEWS | ABOUT NIEHS

- Research
  - Resources for Scientists
    - Databases
      - Alu Pairs Database
      - Biomarkers of Oxidative Stress Study
      - **Chemical Effects in Biological Systems (CEBS)**
        - Citing CEBS
        - Terms of Use
      - Genetic Alterations in Cancer (GAC)
      - Human DNA Polymerase Gamma Mutation Database
      - Microarray Group cDNA Clone Search
      - SNPinfo Web Server
      - Spin Trap Database
- All Scientists
- All Research Groups

Chemical Effects in Biological Systems (CEBS)

A+ A- Share

The CEBS database houses data of interest to environmental health scientists. CEBS is a public resource, and has received depositions of data from academic, industrial and governmental laboratories. CEBS is designed to display data in the context of biology and study design, and to permit data integration across studies for novel meta analysis.

RELATED LINKS

» NIEHS Software and Online Tools Updates

Note: *Users who are using JAWS as accessbility tool are advised to install scripts for using Adobe Flex applications with JAWS before using this application. These scripts have been developed by Adobe for Flex application. You can get the scripts from here (Get JAWS scripts for Flex). Once you have installed these scripts, press the insert key and q together to start the form mode in JAWS. There are number of sound hints available for alerts for opening of dialog boxes and data download. Use Tab key to navigate.*

Open CEBS

Contact CEBS

Download Data

Download Instructions

This FTP site allows you to download study data in CEBS to your computer. You can access it through your browser by clicking the Download link or entering the URL ftp://157.98.192.110/ntp-cebs/datatype/

IF YOU ARE A MAC USER, use Firefox to download data from the FTP site.

This FTP site does not work well with most versions of Safari. Most likely, a Safari user will be able to see the folders or file icons but be unable to open or download the files.

DO NOT try to enter a username or password at the FTP Download login screen.

IT DOES NOT REQUIRE A USERNAME OR PASSWORD.

If your browser shows a login screen on this site asking for a username and password, do not type anything in those boxes.

- If you are using Safari (not recommended) as your browser, click the radio button next to guest and then click connect.
- If you are using other browsers: Close, then Restart your browser, and Clear the Browser Cache before clicking the download link.

If you still see the login screen, please contact fostel@niehs.nih.gov.

Contact

Jennifer Fostel, Ph.D.

CEBS Home Page

A+ A- Share

CEBS Home | Open My Data | Chemical Effects in Biological Systems

Search Study

Search studies using test article, CEBS accession number, NTP Study ID, NTP Study Type, CEBS data domain or institution.

Test Article | NTP Study ID | CEBS accession number

NTP Study Type | CEBS data domain | Institution

Workflows

Tox21 (Phase I) | Explore Conclusion | Site Association

Search Across Studies

Search across the database for subjects meeting your criteria, filtering by:

- **Characteristics** (of studies, participants, stressor, or protocols)
- **Response** (of individual assays or pathology diagnosis)
- **Test Articles** (Chemicals or non-chemical stressors)

Combining search results and then viewing or downloading the results.

Go

Help

Identify and download the list of all studies producing a positive result in the Ames test

Add selected search result data to the Workspace

Download additional depositor data

Accessing Tox21 Phase 2 workflow

Depositing Data to CEBS

Tox21 Phase I: Filtering assay results by activity call

Tox21 Phase I: Find all data for an assay

Data Status (Last updated on May 2012)

Data Source	Total	Searchable Pathology
GeneTox	6306	no pathology
Non-NTP studies	96	25 with pathology
NTP Bioassay - from CBDS	1315	1315
NTP Bioassay - from TDMS	1312	441
NTP ImmunoTox	89	no pathology
NTP Special studies	470	no pathology
Tox21 phase 1	15	no pathology

Select Test Article

CEBS Home | Open My Data

Search Study

Search studies using test article, CEBS accession number, NTP Study ID, NTP Study Type, CEBS data domain or institution.

Test Article

A. Enter part of test article or cas number

toluene | Submit

B. Select from list of matches and click on search button

CAS Number	Test Article Name	Matched Attribute	Attribute Value	# Study(s)
602-01-7	2,3-Dinitrotoluene	NTP_USE	Intermediate in the manufacture of toluene diiscyanate; plasticizer in propellants	16
584-84-9	2,4-Toluene diisocyanate	STRESSOR_NAME	Toluene-2,4-diisocyanate	19
108-88-3	Toluene	STRESSOR_NAME	Toluene	24
6369-59-1	2,5-Toluenediamine sulfate	STRESSOR_NAME	2,5-Toluenediamine sulfate	8
TOLUENES	Toluenes (2,4-diaminotoluene, p-nitrotoluene, 4-chloro-o-phenylenediamine	CAS_NUMBER	TOLUENES	0
25013-15-4	Vinyl toluene	STRESSOR_NAME	Vinyl toluene	23

Search | Cancel

Investigations using that Test Article

2-year rat study

Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Characteristics | Design | Timeline | Data

Accession Number: 002-02916-0005-0000-5

Study Conclusion:

Stressor Name: No stressor

More

Press More button to see other study stressors

Study Variables: DOSE|DOSE_UNIT|TREATMENT_GROUP_TYPE

Characteristic	Characteristic Value
INSTITUTION	NTP
LABORATORY	IRDC
NTP_CHEMTRACK_NUM	C07272B
NTP_STUDY_TYPE	General Toxicology - Long Term
NTP_TDMS_NUMBER	05701-02
NTP_TOXICOLOGY_TYP	Long-Term
START_DATE	09/27/82

Publications:

Title	PubMed

Design tab – explains the groups and the comparators

Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Characteristics | Design | Timeline | Data

TRIAL	TREATMENT_GROUP_TYPE	DOSE_UNIT	DOSE
GROUP5	Test Chemical	PPM	600
GROUP3	Test Chemical	PPM	1200
GROUP6	Test Chemical	PPM	1200
GROUP1	Vehicle Control	PPM	0
GROUP2	Test Chemical	PPM	600
GROUP4	Vehicle Control	PPM	0

Timeline Tab – gives the study events and protocols

Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Characteristics | Design | Timeline | Data

Protocol | day:	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
STRESSOR PROTOCOL				✓																		
PREPARATION																						
OBSERVATION	✓					✓		✓	✓				✓		✓	✓				✓	✓	✓
DISPOSITION																						
CARE				✓																		

Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Characteristics | Design | Timeline | Data

705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733
	✓		✓		✓	✓		✓		✓		✓		✓			✓	✓			✓			✓	✓	✓	✓	✓
	✓		✓	✓	✓	✓	✓	✓		✓		✓		✓			✓	✓		✓	✓			✓	✓	✓	✓	✓
																								✓	✓		✓	✓

Timeline Tab – gives the study events and protocols

Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Characteristics | Design | Timeline | Data

Protocol | day:	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
STRESSOR_PROTOCOL				✓																		
PREPARATION																						
OBSERVATION	✓					✓		✓	✓				✓		✓	✓				✓	✓	✓
DISPOSITION																						
CARE				✓																		

Protocol Type/Name/Attributes	Value
▼ CHEM_STRESSOR_PROTOCOL	
▶ Toluene-0	
▼ Toluene-1200	
VEHICLE NAME	AIR
IS VEHICLE CONTROL	F
ROUTE ADMIN	RESPIRATORY EXPOSURE WHOLE BODY
DOSE PER ADMIN UNIT	PPM
DOSE PER ADMIN	1200
STRESSOR NAME	Toluene
▼ Toluene-600	
DOSE PER ADMIN UNIT	PPM

Study data by “domains”

Chemical Effects in Biological Systems (CEBS)

Study Information	
CEBS Accession Number:	002-02916-0005-0000-5
Chemical Name:	Toluene
CASRN:	108-88-3
NTP Study Type:	General Toxicology - Long Term
NTP Study ID:	C07272B

View Study Details

Animal Number	Dose (PPM)	Species	Strain	Sex	Removal Reason
GROUP1_CAGE2_M_16561	0	Rat	F 344/N	Male	Terminal Sacrifice

Organ and Accountable Site Status
MISSING
Organ
Adrenal gland
NORMAL
Organ
Blood vessel
Bone
Bone marrow
Brain
Epididymis
Esophagus
Intestine, large
Intestine, small
Islets
Mammary gland
Parathyroid gland
Salivary gland
Skin
Stomach
Thymus
Thyroid
Trachea
Urinary bladder

Observations			
Organ	Organ Site	Observation	Qualifiers/Severity
Adrenal gland	Cortex	Hyperplasia	Mild
Heart		Cardiomyopathy	Mild Chronic
Kidney	Renal Tubule	Pigmentation	Mild
Liver	Bile Duct	Hyperplasia	Mild
Lung	Arteriole	Mineralization	Mild
Lymph node	Mandibular	Hyperplasia	Mild Lymphoid
Nose	Olfactory Epi	Degeneration	Mild
Pancreas	Acinus	Atrophy	Minimal
Pituitary gland	Pars Distalis	Hyperplasia	Mild
Preputial gland		Inflammation	Mild Chronic
Prostate		Inflammation	Mild Chronic Active
Spleen		Pigmentation	Mild
Stomach	Forestomach	Hyperplasia	Mild Squamous
Testes		Atrophy	Mild

Study data by “domains”

Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Toxicity Evaluation of Toluene (108-88-3) on F 344/N Rat

Characteristics | Design | Timeline | Data

HISTOPATHOLOGY

Group	Subject	Collection Time (day)	COMMENTS	ORGAN	PRIMARY ORGAN	SUB TOPOGRAPHY	ORIGINAL ...	SEVERITY
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Adrenal gla		Cortex	Hyperplasia	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Testes		Interstit Ce	Hyperplasia	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730	Chronic	Kidney			Nephropath	Moderate
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Kidney		Renal Tubu	Pigmentatic	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Liver			Focal Cellul	
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Liver		Bile Duct	Hyperplasia	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730	Chronic	Liver		Portal	Inflammatic	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Lung			Congestion	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730	Lymphocyte	Lung			Infiltration	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730	Chronic Act	Lung		Interstitium	Inflammatic	Minimal
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Lung		Arteriole	Mineralizati	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730	Lymphoid	Lymph nod		Mandibular	Hyperplasia	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Nose		Olfactory Ep	Degenerati	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Nose		Respirat Ep	Degenerati	Mild
GROUP1_CAGE2_M	GROUP1_CAGE2_M_16561	730		Nose		Lumen	Hemorrhag	Moderate

Example 2

- Find data for male rats with liver injury
- Example of searching across studies using CEBS
  - Search using pathology results, also get all data from these animals
- Example of limitations of data in CEBS
  - Legacy NTP terminology

Searching across studies

A+ A- Share

CEBS Home | Open My Data | Chemical Effects in Biological Systems

Search Study

Search studies using test article, CEBS accession number, NTP Study ID, NTP Study Type, CEBS data domain or institution.

Test Article | NTP Study ID | CEBS accession number
NTP Study Type | CEBS data domain | Institution

Workflows

Tox21 (Phase I) | Explore Conclusion | Site Association

Search Across Studies

Search across the database for subjects meeting your criteria, filtering by:

- **Characteristics** (of studies, participants, stressor, or protocols)
- **Response** (of individual assays or pathology diagnosis)
- **Test Articles** (Chemicals or non-chemical stressors)

Combining search results and then viewing or downloading the results.

Go

Help

Identify and download the list of all studies producing a positive result in the Ames test

Add selected search result data to the Workspace

Download additional depositor data

Accessing Tox21 Phase 2 workflow

Depositing Data to CEBS

Tox21 Phase I: Filtering assay results by activity call

Tox21 Phase I: Find all data for an assay

Data Status (Last updated on May 2012)

Data Source	Total	Searchable Pathology
GeneTox	6306	no pathology
Non-NTP studies	96	25 with pathology
NTP Bioassay - from CBDS	1315	1315
NTP Bioassay - from TDMS	1312	441
NTP ImmunoTox	89	no pathology
NTP Special studies	470	no pathology
Tox21 phase 1	15	no pathology

CEBS Home | My Workspace | Open My Data | Chem

Characteristics | Response | Test Articles

A. Select the assay domain

HISTOPATHOLOGY

B. Select the diagnosis

-- All --

C. Select the organ

Liver

D. Select the organ part

-- All --

E. Select the sub topography

-- All --

F. Select the cells

-- All --

G. Select the distribution

-- All --

H. Select the severity

-- All --

Search

Searches

Search SEX=Male

Search GENUS=Rattus

Search RESPONSE where HISTOPATHOLOGY obser

Show Studies | Clear Search(s)

Add to Workspace

Advanced Options

Advanced Options

Searches	# Studies
Search SEX=Male	3120
Search GENUS=Rattus	54
Search RESPONSE where HISTOPATHOLOGY observations have	2787
male rats with liver toxicity	15

Combine Searches | Add To Workspace | Back

A. Select search

male rats with liver toxicity

B. Retrieve

- (•) Participants in search
- ( ) Search participants plus controls
- ( ) All participants from study

Add | Cancel

The CEBS Workspace

CEBS Home | My Workspace | Open My Data

My Workspace

Help

Source	Study title or Search name		# Studies	# Groups	# Subje...
From Search	male rats with liver toxicity		14	198	911

Show Details | Build Data Matrix | Microarray Analysis | Visual Data Mining | Test Articles List | Study Information

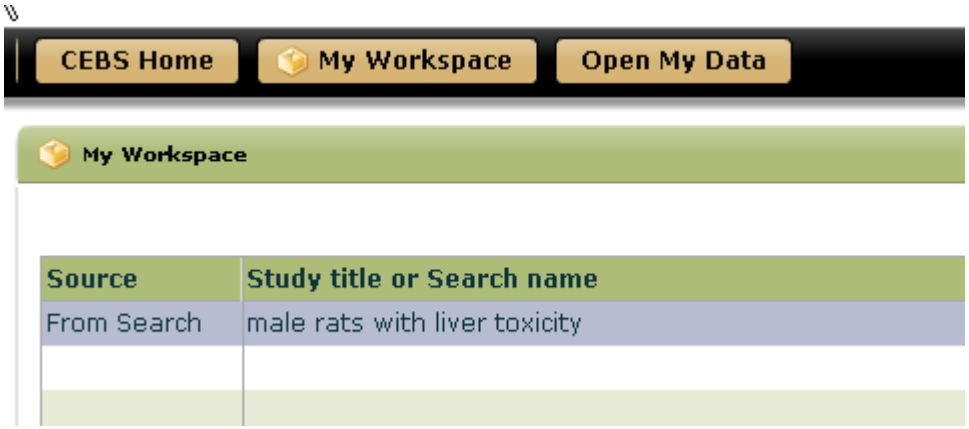
Details for studies in selected list

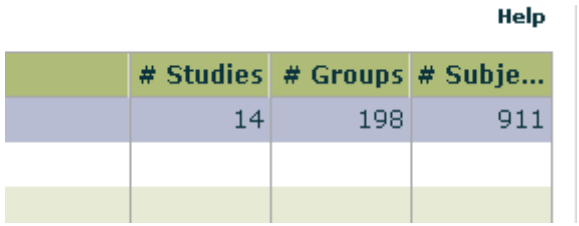
Study Information for male rats with liver toxicity

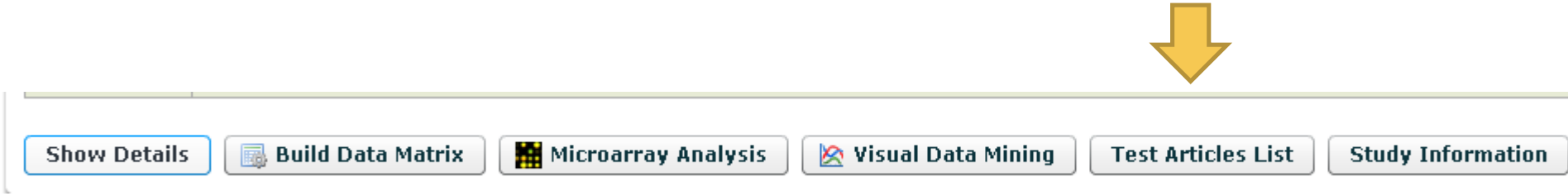
Accession N	Study Title	Conclusion	Start Date	Species Co	Sex	Strain	Study Facto	CEBS URL
008-00001-	A 1-, 3- & 7-		1/1/00		Male	Sprague-Da	Dose|Time	http://tools
008-00001-	Clofibrate:		1-23-01		Male	Sprague Da	Dose|Time	http://tools
008-00001-	Methapyrile		1-23-01		Male	Sprague-Da	Dose|Time	http://tools
008-00001-	Clofibrate: (		1/1/00		Male	Sprague Da	Dose|Time	http://tools
001-00001-	Application			Rat	Male	F/344	DOSE|TIME	http://tools
001-00001-	Application			Rat	Male	F/344	DOSE|TIME	http://tools
001-00001-	Application			Rat	Male	F/344	DOSE|TIME	http://tools
001-00001-	Application			Rat	Male	F/344	DOSE|TIME	http://tools
001-00001-	Application			Rat	Male	F/344	DOSE|TIME	http://tools
001-00001-	Application			Rat	Male	F/344	DOSE|TIME	http://tools
001-00001-	Application			Rat	Male	F/344	DOSE|TIME	http://tools
001-00001-	Application			Rat	Male	F/344	DOSE|TIME	http://tools

Download Close

The CEBS Workspace







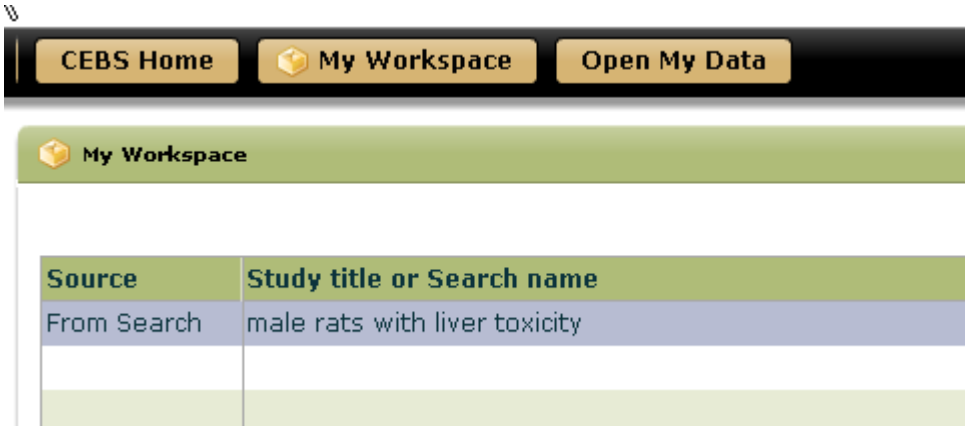
Chemicals in selected studies

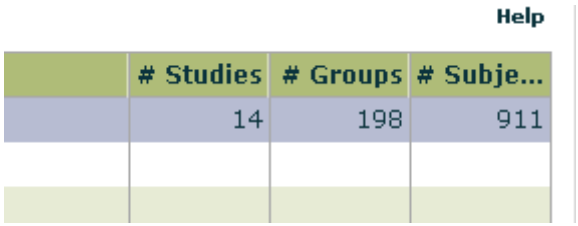
Test Articles List for male rats with liver toxicity

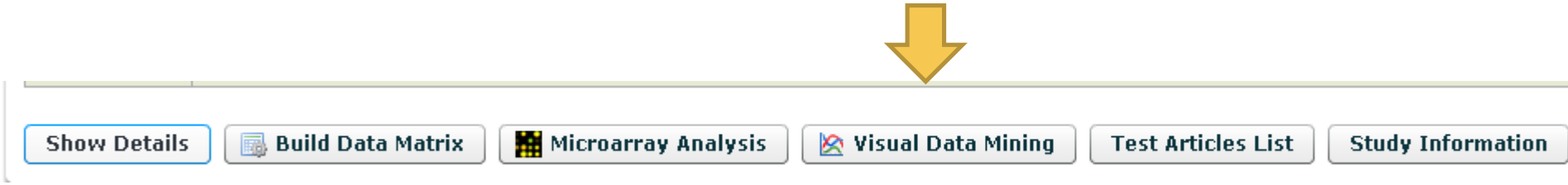
Test Article	Cas Number	Is Vehicle Control?
1,2-Dichlorobenzene	95-50-1	False
1,4-Dichlorobenzene	106-46-7	False
Acetaminophen	103-90-2	False
Bromobenzene	108-86-1	False
Clofibrate	637-07-0	False
Diquat dibromide	85-00-7	False
Methapyrilene	91-80-5	False
Monocrotaline	315-22-0	False
N-Nitrosomorpholine	59-89-2	False
Thioacetamide	62-55-5	False

Download Close

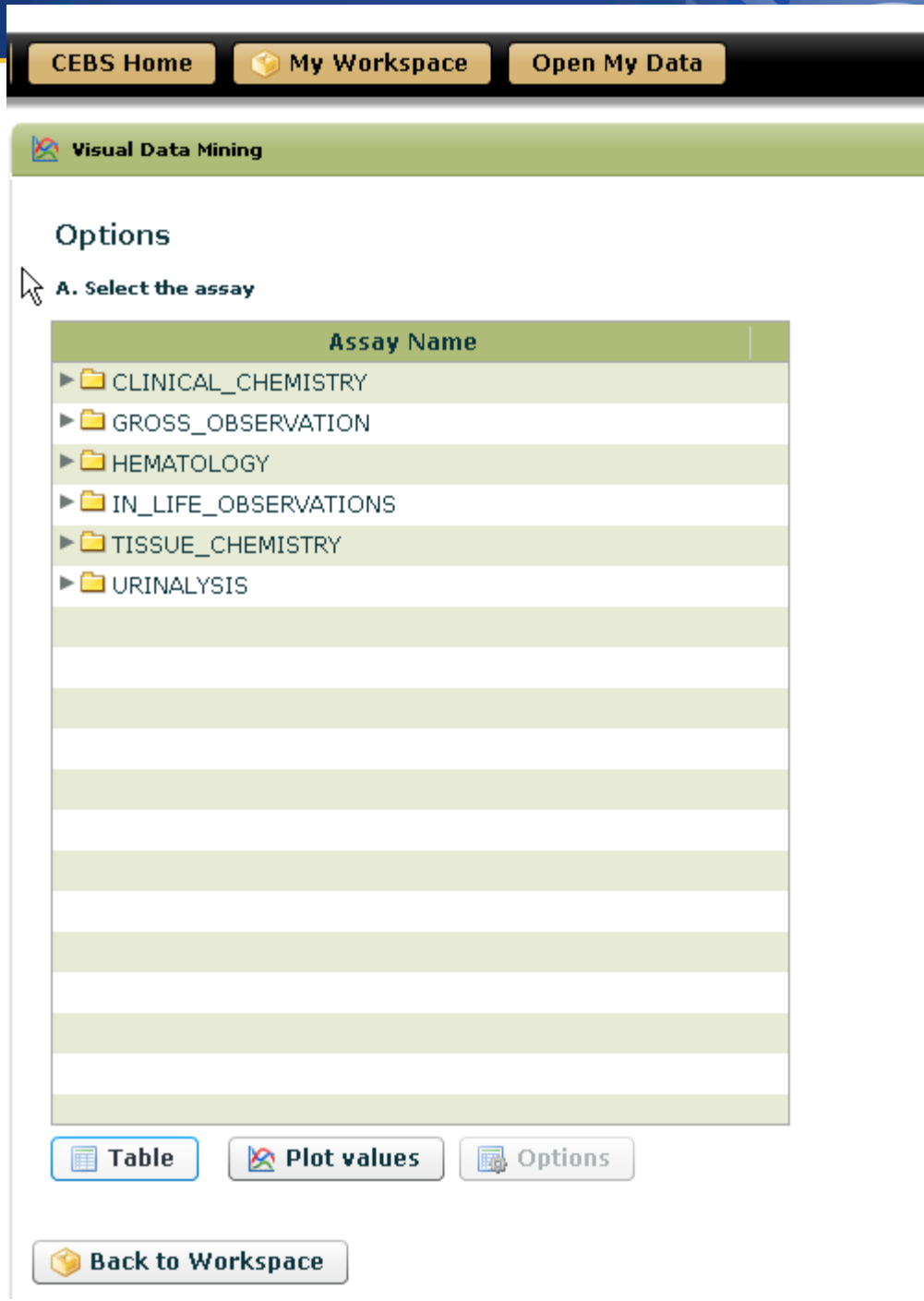
The CEBS Workspace



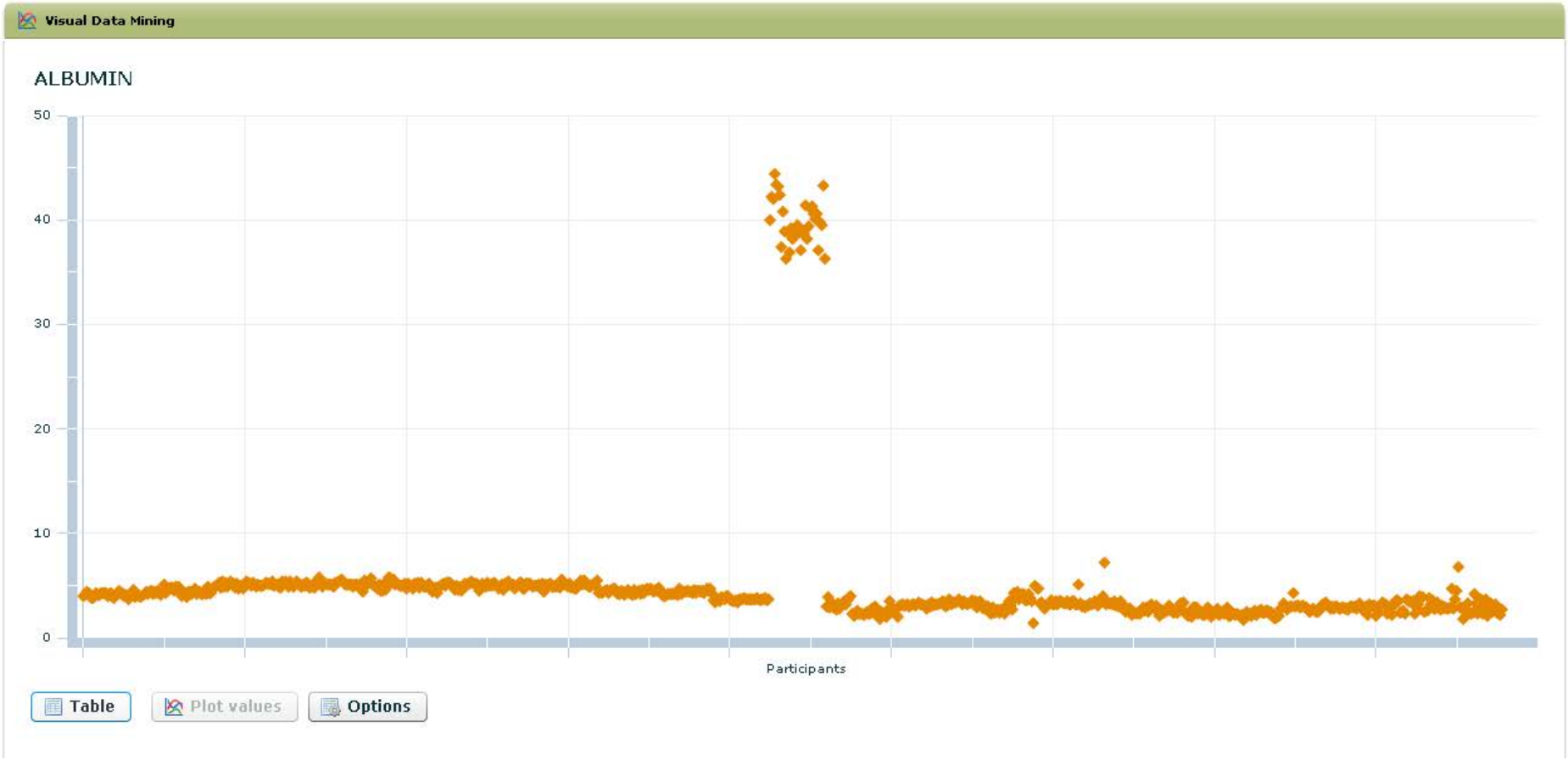




All data are collected



Quick review of clinical chemistry…..



Example 3

- Visual datamining over all data in a domain
- This workflow is implemented for Tox21 phase 1 data
  - 1408 compounds
  - 117 assays
- The Tox21 phase I data was just made public, and this change has not yet been implemented in CEBS

Select assay
-- or --
Select chemical

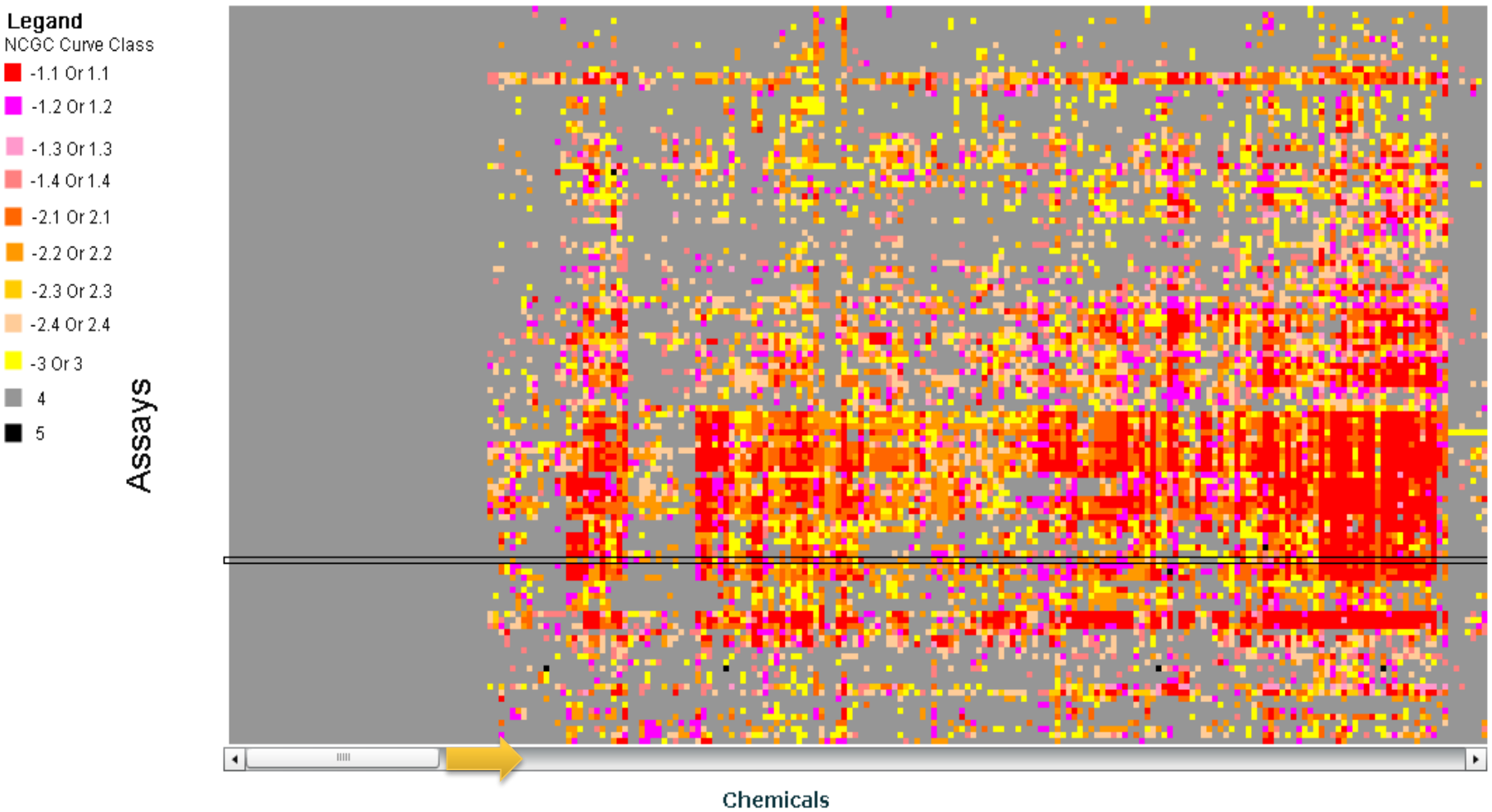
View response
in context of
all assays and
all chemicals

Select the
response level
of interest

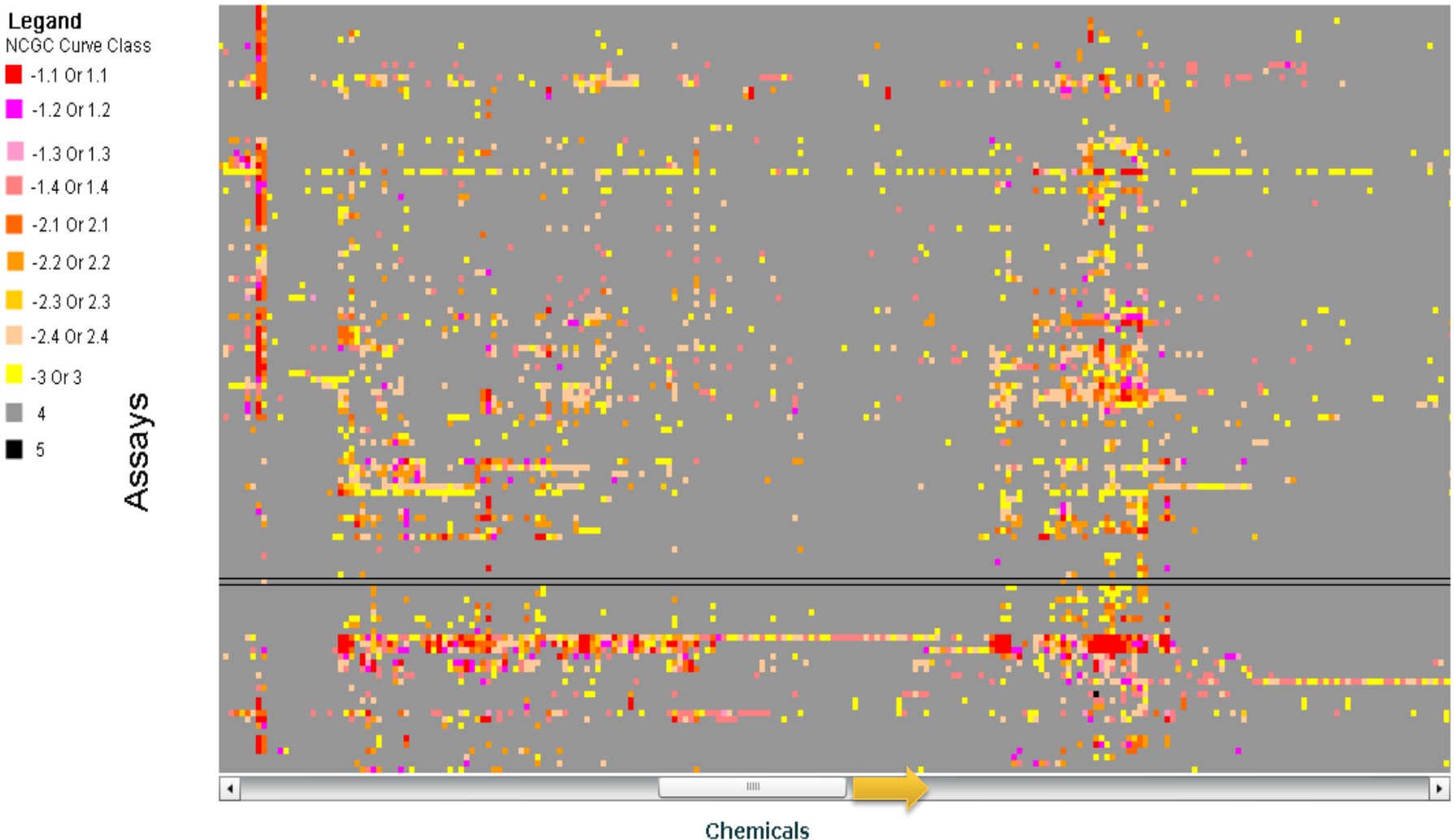
Combine searches

Download

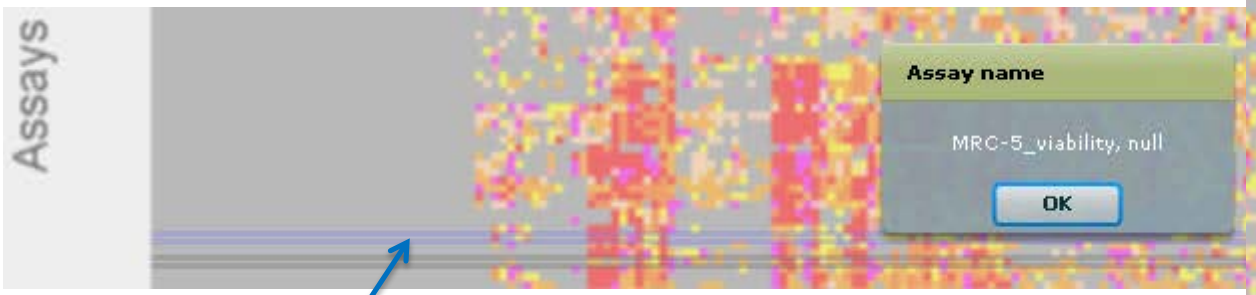
Responses of all assays to all chemicals; target assay is highlighted



Clustered bi-directionally to group assays and chemicals on the basis of response

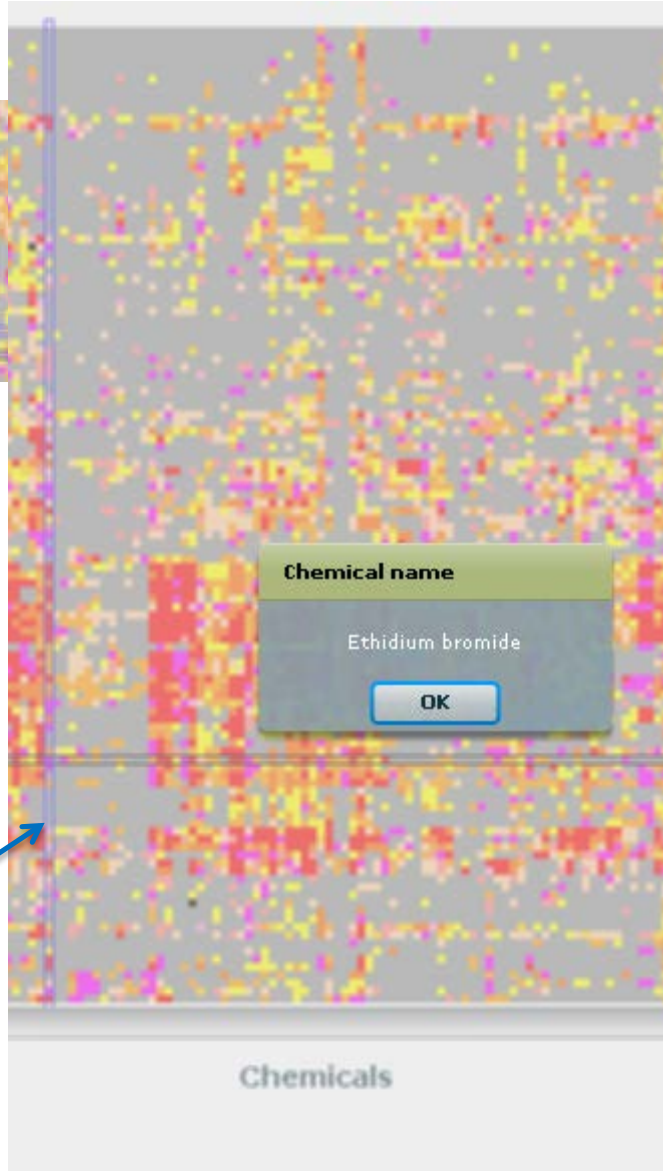


Additional functionality: look up chemical or assay.



Assay Lookup

Click to get “cursor”
To highlight assay of interest



Chemical Lookup

After reviewing responses select response level(s) of interest:

Select NCGC curve class

- [x] -1.1 Or 1.1
- [x] -1.2 Or 1.2
- [x] -1.3 Or 1.3
- [x] -1.4 Or 1.4
- [ ] -2.1 Or 2.1
- [ ] -2.2 Or 2.2
- [ ] -2.3 Or 2.3
- [ ] -2.4 Or 2.4
- [ ] -3 Or 3
- [ ] 4
- [ ] 5

Select | Cancel

list of chemicals producing that response:

Assay	Chemical	Single Channel
Assay=Mesenchymal cell_viability, Curve class=		
Mesenchymal cell_viability	Dimenhydrinate	-1.2
Mesenchymal cell_viability	13-cis-Retinal	-1.1
Mesenchymal cell_viability	2,2',4'-Trichloroacetophenone	-1.1
Mesenchymal cell_viability	2-Octyl-3-isothiazolone	-1.1
Mesenchymal cell_viability	3,3'-Dichlorobenzidine dihydrochloride	-1.2
Mesenchymal cell_viability	4'-(Chloroacetyl)acetanilide	-1.1
Mesenchymal cell_viability	4,4'-Thiobis(6-tert-butyl-m-cresol)	-1.1
Mesenchymal cell_viability	4-(Hydroxyphenyl)retinamide	-1.1
Mesenchymal cell_viability	4-Hydroxy-tamoxifen	-1.1
Mesenchymal cell_viability	5-Azacytidine	-1.2
Mesenchymal cell_viability	6-Mercaptopurine monohydrate	-1.2
Mesenchymal cell_viability	6-Thioguanine	-1.2
Mesenchymal cell_viability	C.I. Acid red 114	-1.2
Mesenchymal cell_viability	CI Basic red 9 monohydrochloride	-1.1
Mesenchymal cell_viability	Captan	-1.1
Mesenchymal cell_viability	Captan	-1.1

Look up chemical to find the assays with response of selected level(s):

Select NCGC curve class

- [x] -1.1 Or 1.1
- [x] -1.2 Or 1.2
- [x] -1.3 Or 1.3
- [x] -1.4 Or 1.4
- [ ] -2.1 Or 2.1
- [ ] -2.2 Or 2.2
- [ ] -2.3 Or 2.3
- [ ] -2.4 Or 2.4
- [ ] -3 Or 3
- [ ] 4
- [ ] 5

Select | Cancel

Assay	Chemical	Single Channel
Chemical=Malachite green oxalate, Curve clas		
HUV-EC-C_viability	Malachite green oxalate	-1.1
DNA_repair_915_p2	Malachite green oxalate	-1.1
BJ_viability	Malachite green oxalate	-1.1
N2a_viability	Malachite green oxalate	-1.1
DNA_repair_1782_p2	Malachite green oxalate	-1.1
Estrogen_receptor_alpha_antagonist	Malachite green oxalate	-1.1
HepG2_viability	Malachite green oxalate	-1.1
MRC-5_viability	Malachite green oxalate	-1.1
Proximal tubule cell_viability	Malachite green oxalate	-1.1
endotoxin-viability-ntp-o3	Malachite green oxalate	-1.1
endotoxin-antagonist-ntp-o2	Malachite green oxalate	-1.1
Hek293_viability	Malachite green oxalate	-1.1

Examples

- Example 1
  - Data for an individual study
  - CEBS standard and user-defined report formats
- Example 2
  - Cross-study query
  - Use of My Workspace to review aspects of query results
- Example 3
  - Cross-domain review of results
  - Selection of responsive assays or responsive chemicals

How is CEBS organized?

- Meta data
  - Biological context, design, protocol and participant information
- Data
  - Based on assays (on specimen)
  - Or observations made during in-life phase
- Database organization
  - Study description and Assay domains
- Data format for loading
  - SIFT: Simple Investigation Formatted Text
- CEBS Data Dictionary

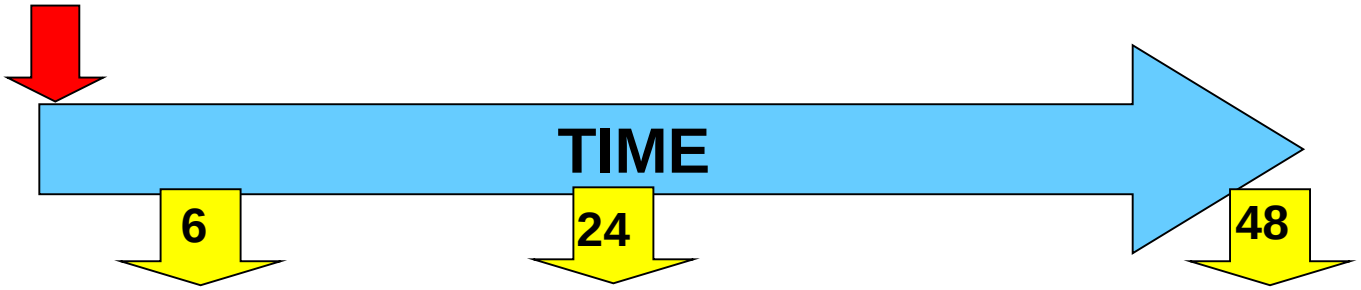
Biological context:

Care: feed, housing
time of day of dosing

Treatment:

0, 50, 150, 1500 mg/kg Acetaminophen by gavage
5 male Sprague-Dawley rats per group

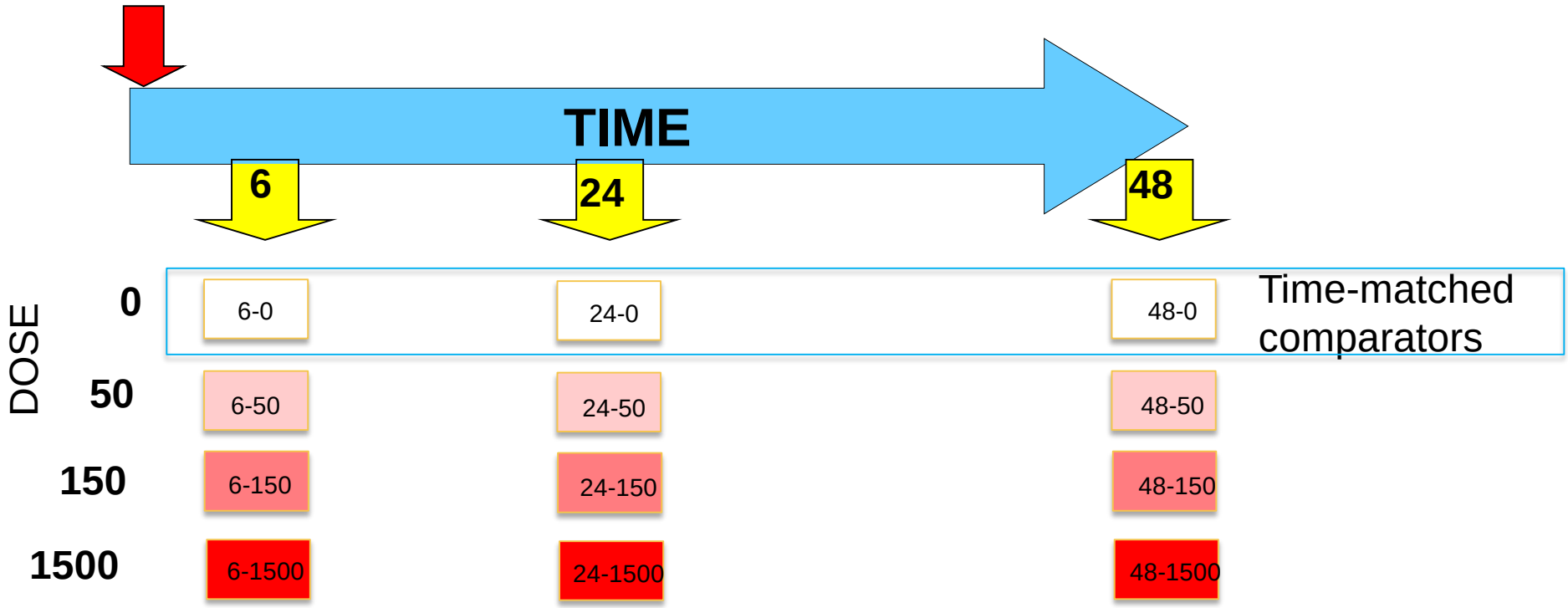
In life observations: in-cage morbidity, behavior



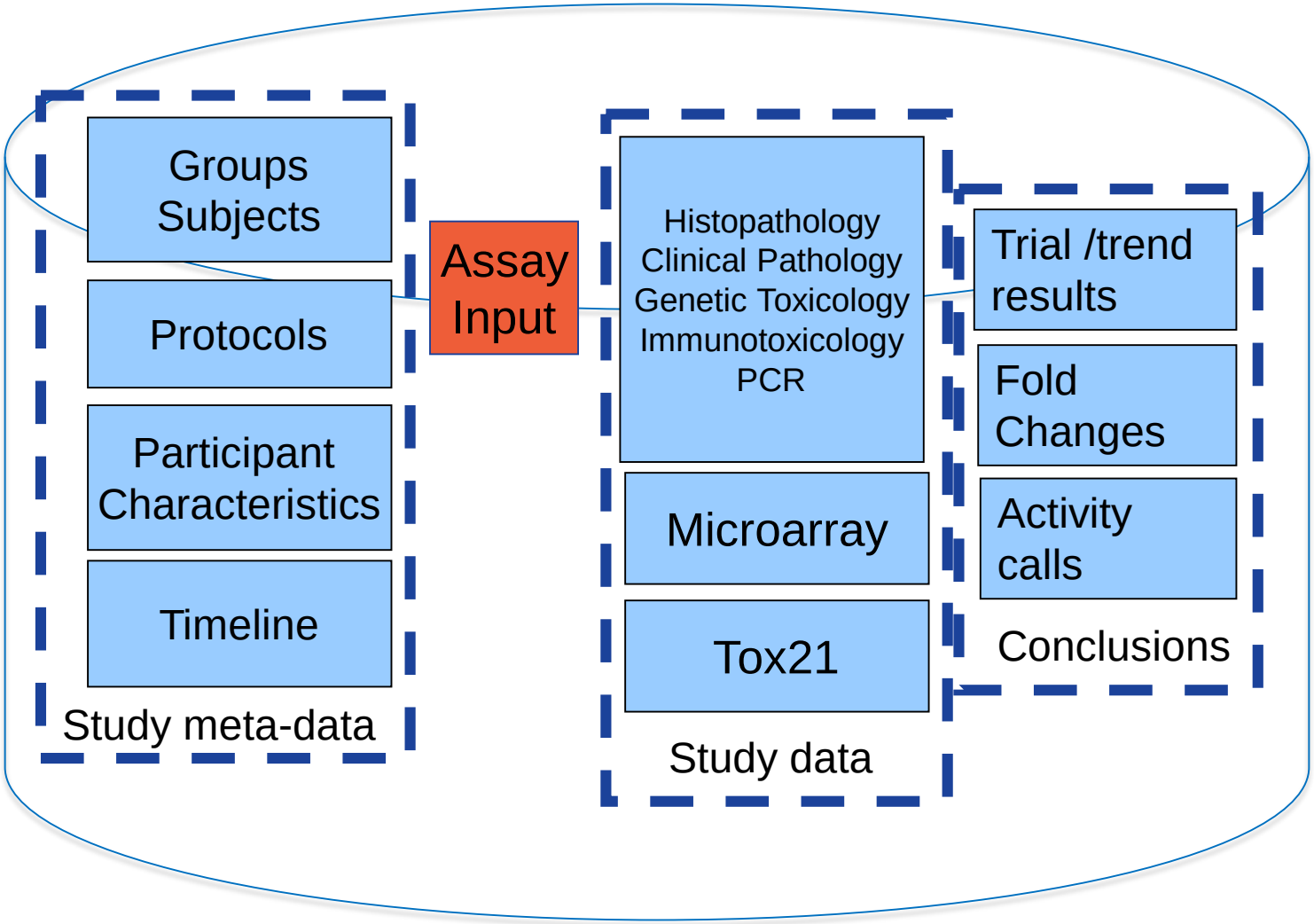
Sacrifice: using anesthesia, time of day

Take specimens: liver, blood, kidney,
for histopathology and microarray

Study design: groups, comparators, and study factors



CEBS database design



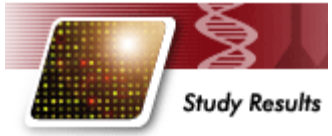
Original data
In file system

SIFT: Standard format for loading CEBS

- Simple Investigation Formatted Text
- Syntax based on SOFT (GEO; *www.ncbi.nlm.nih.gov/**geo**/*)
- Used to capture
  - Study meta-data
  - Study data
  - Data transformation
- Tools (java)
  - SIFT Loader (loads CEBS)
  - SIFT validator
  - CEBS SIFTBuilder
- Uses CEBS data dictionary to validate

Repositories for Toxicity Data

- In Vivo Data Warehouse
  Lilly Greenfield Research Labs



- TDMS and ClinChem DB
  US National Toxicity Program

CEBS
CHEMICAL EFFECTS *in* BIOLOGICAL SYSTEMS

Data Dictionary

Formats for Exchange of Toxicity Data with Regulators

- SEND Standard for Exchange of Nonclinical Data
  CDISC, Pharmquest

- 21CFR part11-compliant repositories,
  e.g. Tox/Path from Xybion Medical Systems

Other Toxicogenomics Database

- MIAME/Tox & Tox/ArrayExpress
  HESI Committee: Application of Genomics to Mechanism based Risk Assessment
  US National Center for Toxicogenomics
  The EMBL European Bioinformatics Institute

Toxicity Data Indexed by Chemical Structure

- DSSTox, US EPA
- ToxML, LIST Consortium

LIMS systems

- MAPS (NIEHS)
- TSP (US EPA)

CEBS DD: toxicological terms and synonyms from various sources

One entry in the CEBS-DD

Content of Row in CEBS-DD	*Name of Column in CEBS-DD*
Required	CEBS-minimal flag
Study	CEBS entity
Study title	CEBS term
The descriptive experiment title for the Stu	CEBS term definition
text	CEBS expected content
Study	SysTox table name
StudyTitle	SysTox Field names
	SysTox table name
	SysTox Field names
tProject	TSP Table Name
	TSP Link ID
	TSP LookUp Table
ProjectName	TSP Field
TEST_ARTICLES	NTP_DATA Table
TEST_ARTICLE_NO	NTP_DATA link
CT_CHEMTRACK_DATA	TDMSE Table
TDMS_STUDY_NO	TDMSE link
	TDMSE Table 2
	TDMSE Table 2 link
	TDMSE TABLE 2 target field
	TDMSE Table 3 / comments
	TDMSE TABLE 3 or comments
All	LEADSCOPE WORKBOOK
StudyTitle	LEADSCOPE FIELD
The StudyTitle of the report	LEADSCOPE Description
Protocol	Xybion-module
Protocol Information (general information)	Xybion-table
ZT.GEN	Xybion-code
Study Title	Xybion-ID
SS	SEND DOMAIN
SSTESTCD	SEND LABEL
STTITL	TERM
SSORRES	RESULTS
Study TitleCharQualifierTitle of study. Example: "91-Day Feeding Study with Compound XYZ in Fischer 344 Rats".	SEND variable label or usage notes
STUDY	Lilly-entity
STUDY NAME	Lilly-entity/attribute
The precise and unambiguous label or specification used to identify a particular (i.e. instance of) STUDY; the title of a STUDY.	Lilly-definition

CEBS-DD: Common Nomenclature Informed by Different Formats, Different Aims and Interests

Term = "Study Title"
Belongs to "Study"
Definition: "The descriptive title for the Study"

Synonyms are found in:

- TDMS: STUDY_NO in CHEMTRACK_DATA
- Lilly: Study Name in STUDY
  (Toxicity Data Repositories)
- Xybion Path/Tox: Study Title in ZT.GEN
- SEND: STTITL, SSORRES in SS
  (Exchange Formats for Regulators)
- Tox-ML: Study Title in all workbooks
  (Chemical Structure Index)
- TSP: Project name, in tProject
  (example LIMS system)

CEBS DD evolution

- Started in 2003
- Originally text based, now XML
- Based on definition tables in CEBS
- Correlation with other syntaxes used to build parsers for various data formats

- Currently aligned with
  - SEND (Standards for Exchange of Nonclinical Data)
  - OBI (Ontology for Biomedical Investigations
- SIFT uses original MAGE-Tab for microarray data

Curation

- Pipelines
  - The CEBS team has four human curators who verify that studies have biologically sensible information in the CEBS presentation screen.
- Individual studies
  - If an individual provides a study for depositon into CEBS, that person is requested to visually confirm that the study presents correctly
- Loads from NTP databases
  - Automated scripts compare the number of studies, observations, subjects, etc.
- Term consolidation
  - Woroking with pathologists and ontologist

Towards Linked Data

Start with SIFT Tabular Data (implicit relationships)

Participant characteristics (meta data)

$GROUP_NAME	$COMPARATOR_NAME	$IS_CONTROL_GROUP	$COMPOUND	$DOSE	$DOSE_UNIT	$ROUTE	$SPECIES_COMMON_NAME	$SEX	$STRAIN
GROUP8	GROUP2		beta-Picoline	1250	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP7	GROUP1		beta-Picoline	1250	mg/l	DOSED WATER	Mouse	MALE	B6C3F1
GROUP6	GROUP2		beta-Picoline	625	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP5	GROUP1		beta-Picoline	625	mg/l	DOSED WATER	Mouse	MALE	B6C3F1
GROUP4	GROUP2		beta-Picoline	312.5	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP3	GROUP1		beta-Picoline	312.5	mg/l	DOSED WATER	Mouse	MALE	B6C3F1
GROUP2		T	beta-Picoline	0	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP1		T	beta-Picoline	0	mg/l	DOSED WATER	Mouse	MALE	B6C3F1

PARTICIPANT_NAME	ORGAN	HISTOPATHOLOGY
GROUP1_CAGE1_1	Lung	Hyperplasia
GROUP1_CAGE1_1	Nose	Metaplasia
GROUP1_CAGE3_3	Adrenal corte:	Adenoma
GROUP1_CAGE3_3	Nose	Atrophy
GROUP1_CAGE7_7	Nose	Hyperplasia

Histopathology observations (data)

…convert to RDF triples with explicit relationships in the data

```
:group1-cage1-1
  rdfs:label "GROUP1_CAGE1_1" ;     # This subject
  rdf:type obo:NCBITaxon_10090 ;    # is a mouse
  sift:member-of :group1-cage1 ;    # member of sub-group :group1-cage1
  bfo:0000159 obo:PATO_0000384 ;    # has quality at all times: male
  sift:dose :dose-0 ;               # has dose :dose-0
  sift:death-status :terminal-sacrifice ; # was sacrificed
  sift:death-date :study-day-731    # died on day 731

:group1-cage1-1-lung
  rdf:type obo:MA_0000415 ;         # This is a mouse lung
  sift:part-of :group1-cage1-1 .    # part of :group1-cage1-1

:group1-cage1-1-lung-hyperplasia
  rdf:type obo:MPATH_134 ;          # This is a mouse hyperplasia
  obo:BFO_0000052 :group1-cage1-1-lung ; # in :group1-cage1-lung
  sift:severity sift:mild .         # has severity: mild
```

Start with SIFT tabular Data (implicit relationships)

Create CEBS Application Ontology, Rules to capture SIFT relationships and tools to convert SIFT to ttl …..James A Overton

Final aims

- RDF triple store
- This is Linked Data*, structured and computer-readable

*Tim Berners-Lee (2006-07-27). "Linked Data—Design Issues". W3C. Retrieved 2010-12-18 at http://en.wikipedia.org/wiki/Linked_data

- Over which we can write SPARQL and other computational queries to produce summary reports
- Which will be available for use on the Semantic Web
- Which will be available for download
- Which users can query in new user interface

Recapping

- Examples of use
  - Data for an individual study
  - Cross-study query
  - Cross-domain review of results
- CEBS practices
  - Capture of meta data and data
  - CEBS data dictionary
  - SIFT files
- CEBS project towards Linked Data

Thanks

- Laura Hall (POB)
- Vistronix team – Asif Rashid
  - Julie Berke-Law, Phyllis Brown, Kevin Church, James Faison, Cari Favaro, Hui Gong, Mary Hicks, Isabel Lee, Justin Johnson, Jeremy McNabb, Robert Mullen, Anand Paleja, Veronica Pham, Mike Shaw
  - James A Overton (CEBS Application Ontology)