Downloading Data from the CEBS FTP Site

Page Navigation (Quick Links):

Function

The FTP site in CEBS allows users to download study data for many studies at once and raw data for individual studies. Data on the FTP site is divided by the data type, individual studies and raw data in file lists. Within the data types, or CEBS Data Domains (e.g. In Life Observations, Histopathology, and Genetic Toxicity), users can download the individual animal data for similar assays. For example, the In-Life Observation data from the NTP contains information about the treatment group, body weight, clinical observations, and food consumption for all subjects in all complete and public NTP studies in CEBS for which in life data was collected. The downloadable data on the FTP site can be used for independent analyses, or identification of results of interest.

For most CEBS Data Domain folders on the FTP Sites, the following file types can be found:
- PDF document describing that CEBS Data Domain
- Files for each institution that collected the data
  - Readme File: explanation of file types and data columns
  - Zip file: experimental data in tab delimited text file(s)
  - Sample file: small sample dataset of the tab delimited text file, provided as an example format for the full dataset
  - Keys file: column header definitions for the zip and sample files.

From the CEBS homepage (http://tools.niehs.nih.gov/cebs3/ui/), and select the Download Data via FTP Server button.

Note: *This FTP site does not work well with most versions of Safari. We recommend Mac users use Firefox or Chrome to download data from the FTP site.*

Note: *Do not try to use a username or password at the FTP Download login screen. It does not require a username or password. If you see a login screen, close then restart your browser. Then clear the browser's cache before accessing the FTP site.*

1) Navigating the FTP Site

Use these instructions to navigate the FTP site and locate desired data.

The links above lead to the folder containing all of the NTP Data Domain folders, as well as some other files and information. Click to open or download these files (depending on your browser settings), which will have an icon other than a folder.

a) Select a Data Domain

To navigate to downloadable data, first select a CEBS data domain by clicking on that domain. Data domains can change. Use the table below for descriptions of some of the common data domains.

CEBS Data Domains:

CEBS Data Domain	Description
Clinical Chemistry	Clinical chemistry refers to the processes used to measure levels of biochemical components in bodily fluids such as serum and plasma. This data domain contains parameters related to lipid metabolism in the liver and kidneys, and to hormone (e.g., thyroid) and electrolyte (e.g., potassium) levels.
Developmental	Developmental toxicity refers to the adverse effects on the developing organism. This data domain contains information relating to prenatal and developmental assessments. Examples of endpoints: fetal examination, markers of puberty (e.g. vaginal opening), and neuromuscular (e.g., grip strength) and neurobehavioral (e.g., motor activity) endpoints.
Genetox-Bacterial Mutagenicity	The bacterial mutagenicity test, or Ames Assay, is used to assess whether a test article causes genetic damage. These data are bacterial colony counts.
Genetox-Micronucleus	The peripheral blood erythrocyte micronucleus test provides an indication of whether a test article is capable of inducing structural and/or numerical chromosomal damage, specifically micronuclei damage. This data domain contains the frequency of micronucleated erythrocytes (red blood cells) in either bone marrow or peripheral blood in response to exposure to a test article. Examples of endpoints include cell counts, percentages of micronucleated cells, and percentages of immature erythrocytes (polychromatic erythrocytes or PCEs) and mature erythrocytes (normochromatic erythrocytes or NCE) within the total

2

CEBS Data Domain	Description
	erythrocyte population.
Gross Observation	Gross observations identify the morphologic alterations in tissues that are visible without the aid of a microscope. This data domain includes changes in organs observed, terminal body weight, and absolute and relative organ weights at necropsy.
Hematology	Hematology is the analysis of cells that circulate in the bloodstream – namely red blood cells, white blood cells, and platelets. Hematology analysis is done on whole blood and includes a complete blood count (CBC) and blood smear evaluation. Examples of endpoints: cell counts, red blood cell size, and hemoglobin concentration.
Histopathology	Histopathology is the study of microscopic changes or abnormalities in tissues. This dataset domain contains descriptions of tissue abnormalities (diagnosis), site of abnormality (location), and severity of lesions. Examples of endpoints: diagnosis, distribution, and severity of lesions, and indication of whether the observed lesion is neoplastic, metastatic, systemic, or malignant.
Immunology	Immunotoxicology identifies the potential for a test article to modulate immune function in mice and/or rats. This data domain contains studies that evaluate the basic functional aspects of diverse immune system components including innate immune function and adaptive immune function (humoral mediated immunity, cell-mediated immunity).
In Life Observations	In-life observations are recorded during the exposure period of a study and describe individual animal's response to test article. Examples of endpoints: body weight measurements, clinical observations, food and water consumption, and body temperature measurements.
PCR	Toxicogenomics studies that employ the polymerase chain reaction (PCR) technique evaluate the biological response to a toxicological challenge at the level of individual genes or a small set of genes. PCR is often used to test specific gene-level hypotheses related to a toxicological challenge or validate results from larger scale microarray studies. Data are reported as the expression of a number of genes related to toxicology under different conditions.
Reproductive	Reproductive toxicity describes the adverse effect of test articles on fertility and fecundity. This data domain contains data relating to the fertility, gestation, and number of live offspring. Examples of endpoints: estrous cyclicity, sperm parameters, live pup counts per litter, and pup sex ratio.
Urinalysis	Urinalysis measures the physical, biochemical, and microscopic properties of urine. A urinalysis includes gross examination of urine for color and clarity and measurements of urine volume and specific gravity. This data domain contains data relating to levels of electrolytes, proteins, and enzymes. Examples of biochemical endpoints: levels of glucose, creatinine, protein, and sodium.
Microarray	Toxicogenomics studies that employ microarrays evaluate the biological response to a toxicological challenge at the level of the genome. Results from these studies are typically reported as gene and molecular pathway read outs, which can be further analyzed for their relationship to toxicity and disease outcomes. Datasets are provided as raw, probe-level files including CEL files (Affymetrix arrays) and TXT files (CodeLink or Agilent arrays).

3

b) Select an Institution

In most data domain folders, there will be a PDF file describing that domain and one or more institution folders. Again, click to select the desired institution.

Some exceptions to this format are:

- GAC Database: The available data for this domain is zipped files of gene and chromosome data from this database.
- Genetox-Drosophila: The only data available for this data domain is a text file with the conclusion for each study
- Genetox-Mouse Lymphoma: The only data available for this data domain is a text file with the conclusion for each study
- Genetox-Rodent Cytogenetics: The only data available for this data domain is a text file with the conclusion for each study
- Microarray: This data domain features data from both institutions and specific labs

c) Download Data by Data Type

i) Common File Types

In most institution folders, there are some or all of the following file types:

- Readme File: explanation of file types and data columns
- Zip file: experimental data in tab delimited text file(s)
- Sample file: small sample dataset of the tab delimited text file, provided as an example format for the full dataset
- Keys file: column header definitions for the zip and sample files.
- Conclusions file: tab delimited text file of the conclusion for all studies in the data domain.

We recommend first downloading the Sample file, before downloading the zipped data file. The sample file will include all of the data headings you will find in the full data file. If you do not understand the data headings, use the Keys file. Click to download any file.

ii) Sample and Zip File Column Headings

Most data domain files include at least the following column headings:

- STUDY_TITLE
- ACCESSION_NUMBER (a unique identification number for each individual CEBS study)

4

- ORGANIZATION_NAME
- FACTORS (study design variables tested in the study such as dose or time)
- CHEMICAL_NAME
- GROUP_NAME
- SUBJECT_NAME
- DOSE
- DOSE_UNIT

The remaining headings can be found in either the readme or keys file.

iii) Exception: Microarray

The exception to this format is the Microarray data domain. Within this data domain, the directories are organized as follows:

- CEBS Accession Number Directory: study data organized by CEBS accession number
- Platform Directory: array type(s) / platform(s) used in each study
- Raw Data Files: CEL or TXT files containing raw, probe-level data

Additionally, three microarray datasets are provided separately from the institution of origin:

- Elk River – contains data for the NTP toxicogenomics study of the chemicals from the West Virginia, Elk River chemical spill. This dataset has 3 test article folders and one old analysis folder. Each test article folder contains:
  - Zip Data File: experimental data files
  - Annotation File: details of data processing
  - Zip BMDExpress File: benchmark dose analysis tools

- MouseLiver – contains data from GEO for toxicogenomics studies of liver; dataset complied in 2012. Files provide study details and normalized data.
- DrugMatrix – contains data from NTP DrugMatrix, a comprehensive rat toxicogenomics dataset, organized by tissue and array type.

d) Download Raw Data by Individual Study

Additionally raw data files provided by depositors are available by selecting the [parent directory] link or navigating up one folder, ftp://157.98.192.110/ntp-cebs/.

These files are located in FileList and individualstudy folders. Both folders are organized by CEBS accession number, which can be found in CEBS when reviewing study data.

2) Additional Instructions/Information

1) 'CEBS Home' icon returns the user back to the CEBS Home page
2) Downloading and displaying data in Excel
   - Select the Download Data Icon
   - The next screen displays available locations to save data
   - Once a location has been determined, select Save
   - To access the data, go to the location where the data was saved
   - You will notice that the data is saved in text format
   - Open in Notepad or your favorite text editor
   - Select all by choosing Ctrl + A (PC) or Command + A (Mac). This function will highlight the entire document.
   - Right click your mouse and select copy
   - Open Excel
   - Right click your mouse and select paste
   - Save as an Excel document
   - To add filtering options, navigate to the Data Tab in Excel and select Filter
3) Flash movie running flex code for CEBS indicates a software platform is being used to create graphics or texts for CEBS
4) CEBS data may be cited by navigating to the CEBS Support page, and selecting the Citing CEBS document under FAQs. Here, instructions can be found on citing NTP Data as well as non-NTP data in CEBS.

3) Example Section

Below are a couple examples of questions that can be answered using the FTP site to download data.

a) Example: Immunology Data for 3 Test Articles

For this example, we will answer this question: How can I download the individual animal immunology data for three test articles? Our test articles of interest are: n,n-Dimethyl-p-Toluidine, M-Nitrotoluene, and o-Nitrotoluene. To answer this question, a user could locate the immunology studies for each test article, download the individual animal data for each study, and then combine the data manually into one worksheet. An easier way is to use the FTP site to download all individual animal immunology data and then filter the worksheet in Excel to show only the test articles of interest.

1) From the CEBS homepage (http://tools.niehs.nih.gov/cebs3/ui/) select the Download Data button. This will take you to the FTP Server website.
2) Because we are interested in Immunology data, click to select the IMMUNOLOGY data domain folder.

6

3) Select the NTP folder

4) Because we are interested in all immunology data, select the IMMUNOLOGY_NTP.zip file to download. This will download a compressed file containing all NTP individual animal immunology data.

5) Save this file and then use your file explorer on your computer to extract the compressed file (unzip) to your desired location.

6) Open the extracted folder and open the IMMUNOLOGY_NTP.txt file in Excel.
    i. Because this file does not contain CASRN's, you can do this by alternate (right) clicking on the file, and then selecting Open With and then Excel. See [Additional Instructions](#) for other download and opening options.

7) Once the file is open in Excel, save it to your desired location

8) Use Excel to filter the Chemical column to show only the desired three chemicals
    i. Select a cell in row 1 of the worksheet
    ii. In the Excel menu, navigate to the Data tab, and select the Filter icon. Dropdown menu arrows should appear in all heading cells in the first row.
    iii. Select the dropdown filter menu in the Chemical column (Column A). Click the top Select all option, to uncheck all chemicals. Then scroll through the list to check the chemicals of interest: n,n-Dimethyl-p-Toluidine, M-Nitrotoluene, and o-Nitrotoluene.
    iv. The list will now be filtered to show all immunology data for those three test articles.
    Note: *This filter will not include the data from the control groups. Filter by Accession number to include control groups.*

b) Example: All PCE Micronucleus Data

For this example we will answer the question: How can I find all the genetic toxicity individual animal data for the number of micronucleated bone marrow polychromatic erythrocytes?

1) From the CEBS homepage (http://tools.niehs.nih.gov/cebs3/ui/) select the Download Data button. This will take you to the FTP Server website.

2) Because we are interested in micronucleus data, click to select the GENETOX_Micronucleus data domain folder.

3) Select the NTP folder

4) Because we are interested in all Micronucleus data, select the GENETOX_Micronucleus_NTP.zip file to download. This will download a compressed file containing all NTP individual animal Micronucleus.

5) Save this file and then use your file explorer on your computer to extract the compressed file (unzip) to your desired location.

6) Open the extracted file and open the GENETOX_Micronucleus_NTP.txt file in Excel. See Additional Instructions for full download and opening options.

7) Once the file is open in Excel, save it to your desired location

8) Locate the column for the desired endpoint: the number of micronucleated polychromatic erythrocytes. You may need to open and use the keys.txt file to do this. This endpoint is column AU: BMPCEMNK

9) Use Excel to filter the BMPCEMNK column to show only the desired three chemicals

    i. Select a cell in row 1 of the worksheet

    ii. In the Excel menu, navigate to the Data tab, and select the Filter icon. Dropdown menu arrows should appear in all heading cells in the first row.

    iii. Select the dropdown filter menu for the column BMPCEMNK. In the list of values, scroll down to the bottom to find (Blanks). Click to uncheck this value

    iv. Only rows that contain a value for the BMPCEMNK column will be displayed.