PUBCHEM UPLOAD:
A STREAMLINED WAY TO SUBMIT CHEMICAL & BIOASSAY RESULTS

Benjamin A. Shoemaker
PubChem Data Size & Growth

Growth In PubChem Substances / Compounds

Total Count (millions)

Date
PubChem Data Size & Growth

Growth In PubChem Contributing Organizations

- Substance
- BioAssay
- BioAssay, RNAi

Total Count vs Deposit Date
PubChem Data Size & Growth

Growth In PubChem Substance Outcomes

Total
MLP
RNAi

Deposit Date
Nov-05 Sep-06 Jul-07 May-08 Mar-09 Jan-10 Nov-10 Sep-11 Jul-12 May-13

Total Count
10,000,000
1,000,000
100,000
10,000
1,000
100
10
1

Types of Experimental Data in PubChem

- **PubChem Substance:**
  - Chemical structures
  - Drug names/synonyms
  - Patents

- **PubChem BioAssay:**
  - 1 target, HTS screens
  - Dose-response, confirmatory
  - Summary, Probes
  - Counter screens

- Links to PubMed, Gene, Protein, OMIM, Biosystems
- Annotations/Ontologies
- ChEMBL
- siRNA screens
- Multi-target profiles
- Tox screens
- Curated (PDB, annotations)
Types of Submitters to PubChem

• Chemical supply companies
• Automated journal uploads (encourage assay journals)
• ChEMBL
• NIH Molecular Screening Centers
• siRNA screening consortium
• Database collections (gov’t, public, private)
• Individual research labs
PubChem Upload: Efficient Web Interface


• Support wide range user-defined data
• Useful to the submitter:
  • Take everything they have
  • Minimize hassle
• Useful to the PubChem user:
  • Make it easy to find relevant data
  • Make it possible to aggregate & compare results
Goals / Features for New Interface

• Fewer time-outs for complicated forms
• Form data saved more robustly
• Less scrolling
• Page reloads minimized
• Extensible interface
  • Easy to modify
  • Data access independent
• In-place edit without file reload
• Uniform spreadsheet use
Previous Web Interface Issues

- Web controls: HTML sent, page size, reload
Security needs

- Initial curator check of new submitter
- Login authentication
- Accredited updates/new additions per data source
- Multiple users on one submission account
- On-hold data up to one year
New Submitters: Make it easy!

- Get minimal information
- Convenient to start submission
- Get in and out quickly

Challenges:
- Want to verify identity
- Get good source keywords for PubChem search
- Sign Data Transfer Agreement
- Quickly transition from test to production
- Keep track of test submissions
New Submitter Info

PubChem Upload

Welcome to PubChem Upload (formerly known as the PubChem Deposition Gateway)

The PubChem Upload tool enables you to submit data to the PubChem Substance and BioAssay databases, including chemical structures, experimental biological activity results, annotations, siRNA data and more. It can also be used to update records that you previously submitted to PubChem. In either case, you have the option of uploading a file or completing web forms to enter your data. Once your submission is reviewed and published in PubChem, your data will be available for Open Access to the public around the world.

Getting Started

An account is required in order to submit data to PubChem. You are welcome to first create a test account, which allows you to upload data into a pending submissions database, then upgrade later to a full account, which allows you to commit data from the pending submissions database into the official PubChem database. A data transfer agreement is required when you choose to setup or convert to a full account, and is displayed as part of the setup wizard for a full account. We look forward to receiving your data.

We Appreciate Your Feedback

Please note that this is still a beta version of the interface, and we appreciate your patience and feedback while the project evolves. For feedback, questions, and comments please contact our helpdesk.
New Submitter Info

Do you wish to make your data public?

- Yes.
- No, just testing at this time.
New Submitter Info

Accept Data Transfer Agreement

AGREEMENT TO PARTICIPATE IN NLM PUBCHEM PROJECT

Made this 25 day of April 2013, by and between the National Library of Medicine, Department of Health and Human Services (hereinafter referred to as “NLM”) and Linus Pauling (hereinafter referred to as “Participant”).

WHEREAS, the NLM was established by statute in order to assist the advancement of medical and related sciences, and to aid the dissemination and exchange of scientific and other information important to the progress of medicine and to the public health, and, in carrying out this purpose is authorized by statute to publish and make available its indexes and bibliographical listings, and to engage in other activities in furtherance of NLM’s purpose (sections 381 and 382 of the Public Health Service Act (42 U.S.C. 275, 275a)); and

WHEREAS, one of the goals of the National Information Infrastructure is to improve national health care through the delivery of information services;

I accept the Data Transfer Agreement displayed above.
Starting a submission

Substance Upload: Fill in Form

Do you have an SID (PubChem Substance ID) or a CID (PubChem Compound ID) that you would like to start from?

- No
- PubChem SID
- PubChem CID

Hint

PubChem most likely already contains a substance with an exactly the same or similar structure to the one you are looking to upload. If you want to use an existing substance as a template, you can search PubChem Substance and then enter an SID here. Or, if you prefer, you can search PubChem Compound and then enter CID here.

If you enter an SID of the substance you own, you will also be given options to update it.

For more info, please refer to our brief substance upload tutorial.
Substance Upload: Modify Substance (SID 131485344)

Substance Name * cannot be changed in modify mode
CPD-13380

Synonyms
CPD-13380
dTDP-2,6-dideoxy-D-glycero-hex-2-enose-4-ulose

Structure
### PubChem Upload

**These substances have been previously uploaded to PubChem by you**

Select a row for further options.

**Showing rows 1 to 10 of 927.**

<table>
<thead>
<tr>
<th>N</th>
<th>Depiction</th>
<th>SID</th>
<th>REGID</th>
<th>SYNONYM</th>
<th>Creation Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Image" /></td>
<td>131485348</td>
<td>CPD-13530</td>
<td>4-carboxy-2-hydroxy-6-methoxy-6-oxohepta-2,4-dienoate CHMOD</td>
<td>2012/01/10</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2" alt="Image" /></td>
<td>131485347</td>
<td>CPD-13498</td>
<td>gamma-Glu-Cys gamma-Glu-Cys-beta-Ala CFD-13498</td>
<td>2012/01/10</td>
</tr>
</tbody>
</table>

[View in PubChem] [Use as Template] [Modify] [Revoke]
Substance Form

[Image of a chemical structure with SMILES notation: O(C1=C-C-C1C(O[H]=O)O(-)=O)C]
Substance Tags

**Substance Upload: Fill in Form**

<table>
<thead>
<tr>
<th>Tab</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name and Structure</td>
<td></td>
</tr>
<tr>
<td>URLs</td>
<td></td>
</tr>
<tr>
<td>Dates</td>
<td></td>
</tr>
<tr>
<td>Comments</td>
<td></td>
</tr>
<tr>
<td>Cross-References</td>
<td></td>
</tr>
</tbody>
</table>

**PubChem ID**

None...click to add

**GenBank ID**

None...click to add

**OMIM ID**

None...click to add

**Taxonomy ID**

None...click to add
Assay Wizard

Assay Upload: Form

Do you have an AID (PubChem Assay ID) that you would like to start from?

- [No]
- [PubChem Assay ID] 652068

Hint

PubChem might already contain an assay similar to an assay you are looking to upload. If you would like to use an existing assay as a template, you can search PubChem Assay for a suitable assay and then enter the AID here.

If you enter an AID of the assay you own, you will also be given options to update it.

For more info, please refer to our brief assay upload tutorial.

Back Next
New Assay Submission

PubChem Upload

ASSAY SUBMISSION 15311 (Created)

New Assay: AUTO: name 20130426143214
DSN: 1064
Submitted: 04/28/2013 14:32:14

Please use the form below to enter/edit assay description (annotations) and data (test results), and then click Validate button when ready. You can also use File Upload button to upload all or part of the assay.

Description | Data | Options | History

External Assay Registry ID (a unique assay identifier in your database) *
AUTO: 20130426143214

Assay Name (for display purposes) *
AUTO: name 20130426143214

Assay Description
None...click to add

Assay Protocol (procedure used to generate the assay)
None...click to add
Input from PubChem List

### PubChem Upload

#### Welcome Substances Assays Account Settings

#### New Submission

#### Pending In PubChem

These assays have been previously uploaded to PubChem by you. Select a row for further options. Showing rows 1 to 10 of 1386.

<table>
<thead>
<tr>
<th>N</th>
<th>AID</th>
<th>Name</th>
<th>Creation Date</th>
<th>Hold Until Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>666953</td>
<td>Summary of a probe development effort to identify inhibitors of COUP-TFII (NR2F2)</td>
<td>2013/04/24</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>652038</td>
<td>Counterscreen for molecules that bind (CAG RNA repeats: fluorescent based biochemical counterscreen assay for inhibitors of the DNA-based (CAG/GTC) TO-PRO-1 dye complex</td>
<td>2013/02/28</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>651821</td>
<td>Fluorescence-based biochemical primary high throughput screening assay to identify molecules that bind (CAG RNA repeats</td>
<td>2012/11/28</td>
<td></td>
</tr>
</tbody>
</table>

---

Write to Helpdesk | Disclaimer | Privacy statement | Accessibility | Data Citation Guidelines  
National Center for Biotechnology Information | NLM | NIH | IHS
New Assay Submission

[Image of the NCBI PubChem assay submission page]

**Assay Submission 15311 (Created)**

New Assay: AUTO: name 20130426143214
DSN: 1084
Submitted: 04/28/2013 14:32:14

Please use the form below to enter/edit assay description (annotations) and data (test results), and then click Validate button when ready. You can also use File Upload button to upload all or part of the assay.

**Description**

External Assay Registry ID (a unique assay identifier in your database) *
AUTO:20130426143214

Assay Name (for display purposes) *
AUTO: name 20130426143214

**Assay Description**

Here is the complete description of my assay...

**Assay Protocol (procedure used to generate the assay)**

None... click to add
Edit Multi-Value Fields

The database type of XRef identifier supplied. This type column must appear as the first column of the spreadsheet.

<table>
<thead>
<tr>
<th>REQUIRED_HEADER</th>
<th>XREF_TYPE</th>
<th>XREF_VALUE</th>
<th>XREF_ANNOTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PUBCHEM_AID</td>
<td>851821</td>
<td>Primary screen (nCAG) inhibitors</td>
</tr>
<tr>
<td>2</td>
<td>PUBCHEM_AID</td>
<td>851826</td>
<td>Summary (CAG) inhibitors</td>
</tr>
<tr>
<td>3</td>
<td>PUBCHEM_NCBIGENE_ID</td>
<td>3064</td>
<td>HTT huntingfish</td>
</tr>
<tr>
<td>4</td>
<td>PUBCHEM_EXT_DATASOURCE_URL</td>
<td><a href="http://mipan.fiorida%E7%94%9F%E6%B6%AF">http://mipan.fiorida生涯</a></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PUBCHEM_NCBICIMM_ID</td>
<td>143100</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>PUBCHEM_NCBITAXONOMY_ID</td>
<td>9006</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>PUBCHEM_NCBIPUSHMED_ID</td>
<td>17417037</td>
<td>Reference 1</td>
</tr>
<tr>
<td>8</td>
<td>PUBCHEM_NCBIPUSHMED_ID</td>
<td>8937828</td>
<td>Reference 2</td>
</tr>
<tr>
<td>9</td>
<td>PUBCHEM_NCBIPUSHMED_ID</td>
<td>12747635</td>
<td>Reference 3</td>
</tr>
<tr>
<td>10</td>
<td>PUBCHEM_NCBIPUSHMED_ID</td>
<td>15229312</td>
<td>Reference 4</td>
</tr>
</tbody>
</table>
Context-dependent Help

**Summary (t(CAG) Repeat)**

**XREF_ANNOTATION**

An explanatory text describing the relevance of this cross-referenced item to the assay.

<table>
<thead>
<tr>
<th>REQUIRED_HEADER</th>
<th>XREF_TYPE</th>
<th>XREF_VALUE</th>
<th>XREF_ANNOTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PUBCHEM_AID</td>
<td>051821</td>
<td>Primary screen (t(CAG) inhibitors in singlicate)</td>
</tr>
<tr>
<td>2</td>
<td>PUBCHEM_AID</td>
<td>881625</td>
<td>Summary (t(CAG) inhibitors)</td>
</tr>
<tr>
<td>3</td>
<td>PUBCHEM_NCBI_GENE_ID</td>
<td>3664</td>
<td>HTT huntingtin</td>
</tr>
<tr>
<td>4</td>
<td>PUBCHEM_EXT_DATASOURCE_URL</td>
<td><a href="http://milp.moffitt.cancer.org">http://milp.moffitt.cancer.org</a></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PUBCHEM_NCBI_OMIM_ID</td>
<td>143100</td>
<td>Reference 1</td>
</tr>
<tr>
<td>6</td>
<td>PUBCHEM_NCBI_TAXONOMY_ID</td>
<td>9606</td>
<td>Reference 2</td>
</tr>
<tr>
<td>7</td>
<td>PUBCHEM_NCBI_PUBMED_ID</td>
<td>17417937</td>
<td>Reference 3</td>
</tr>
<tr>
<td>8</td>
<td>PUBCHEM_NCBI_PUBMED_ID</td>
<td>8327828</td>
<td>Reference 4</td>
</tr>
<tr>
<td>9</td>
<td>PUBCHEM_NCBI_PUBMED_ID</td>
<td>12747995</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>PUBCHEM_NCBI_PUBMED_ID</td>
<td>16223312</td>
<td></td>
</tr>
</tbody>
</table>
Assay Data: Result Definitions

![Assay Data Form](image)

**Result Column Definitions**

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUBCHEM_RESULT_TAG</td>
<td>FLOAT</td>
<td>Average inhibition at 3 uM</td>
</tr>
<tr>
<td>RESULT_TYPE</td>
<td></td>
<td>Standard deviation of inhibition at 3 uM</td>
</tr>
<tr>
<td>RESULT_DESCR</td>
<td>PERCENT</td>
<td>Standard deviation of inhibition at 3 uM</td>
</tr>
<tr>
<td>RESULT_UNIT</td>
<td></td>
<td>Inhibition at 3 uM [1]</td>
</tr>
<tr>
<td>RESULT_ATTR_CONC_MICROMOL</td>
<td>PERCENT</td>
<td>Inhibition at 3 uM [2]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inhibition at 3 uM [3]</td>
</tr>
</tbody>
</table>
Assay Data: Result Definitions

**PERCENT**

This is not a standard result column provided by the assay submitter: Average Inhibition at 3 uM

**RESULT_UNIT**

Various units are available to better define the measurement of a given result column.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUBCHEM_RESULT_TAG</td>
<td>Average Inhibition at 3 uM</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>RESULT_TYPE</td>
<td>FLOAT</td>
<td>Inhibition at 3 uM [1]</td>
</tr>
<tr>
<td>RESULT.Descr</td>
<td>Normalized percent inhibition of the primary screen at a compound concentration of 5.00 micromolar Standard deviation of inhibition at 3 uM Normalized percent inhibition</td>
<td>FLOAT</td>
</tr>
<tr>
<td>RESULT.Unit</td>
<td>PERCENT</td>
<td>NONE</td>
</tr>
<tr>
<td>RESULT_ATTR_CONC_MICROMOL</td>
<td>NONE</td>
<td>PERCENT</td>
</tr>
<tr>
<td>RESULT_CNSC_RESPONSE.Series_ID</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>RESULT.IS_ACTIVE_CONCENTRATION</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
## Review Pending Submissions

**PubChem Upload**

### Review or modify pending assay submissions
Select a row for further options.

<table>
<thead>
<tr>
<th>ID</th>
<th>Action</th>
<th>Status</th>
<th>RgID</th>
<th>Date</th>
<th>Curator</th>
</tr>
</thead>
<tbody>
<tr>
<td>15311</td>
<td>New Assay: Other</td>
<td></td>
<td>AUTO:20130426143214 AUTO:20130425143214</td>
<td>2013/04/26</td>
<td>14:32</td>
</tr>
<tr>
<td>15310</td>
<td>New Assay: Primary Screening</td>
<td>TEMPLATE:20130426143012 (D(CAG)_INH_FLIN...)</td>
<td>2013/04/26</td>
<td>14:30</td>
<td></td>
</tr>
</tbody>
</table>

[View/Edit Submission]
Preview in PubChem: Substance

Note: Preview record

acetosal - Substance Summary

Table of Contents
Synonyms
Substance Information

Synonyms
acetosal
acetyl/salicylic acid
aspirin
my-registry-id-1

Substance Information
Preview_SID 157375021
Data Source:
Depositor: 1084
External ID: my-registry-id-1
**Comment**

Due to the increasing size of the MLPCN compound library, this assay may have been run as two or more separate campaigns, each campaign testing a unique set of compounds. All data reported were normalized on a per-plate basis. Possible artifacts of this assay can include, but are not limited to: dust or lint located in or on wells of the microtiter plate, and compounds that modulate well fluorescence. All test compound concentrations reported above and below are nominal; the specific test concentration(s) for a particular compound may vary based upon the actual sample provided by the MLSMR. The MLSMR was not able to provide all compounds selected for testing in this assay.

**Result Definitions**

<table>
<thead>
<tr>
<th>TID</th>
<th>Name</th>
<th>Description</th>
<th>Histogram</th>
<th>Type</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Average Inhibition at 3 uM (3μM**)</td>
<td>Normalized percent inhibition of the primary screen at a compound concentration of 3.00 micromolar.</td>
<td></td>
<td>Float</td>
<td>%</td>
</tr>
<tr>
<td>2</td>
<td>Standard Deviation</td>
<td>Standard deviation of inhibition at 3 uM</td>
<td></td>
<td>Float</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Inhibition at 3 uM [1] (3μM**)</td>
<td>Normalized percent inhibition of the primary screen at a compound concentration of 3.00 micromolar, replicate [1]</td>
<td></td>
<td>Float</td>
<td>%</td>
</tr>
<tr>
<td>4</td>
<td>Inhibition at 3 uM [2] (3μM**)</td>
<td>Normalized percent inhibition of the primary screen at a compound concentration of 3.00 micromolar, replicate [2]</td>
<td></td>
<td>Float</td>
<td>%</td>
</tr>
<tr>
<td>5</td>
<td>Inhibition at 3 uM [3] (3μM**)</td>
<td>Normalized percent inhibition of the primary screen at a compound concentration of 3.00 micromolar, replicate [3]</td>
<td></td>
<td>Float</td>
<td>%</td>
</tr>
</tbody>
</table>

**Additional Information**

**Grant Number:** 1 R03 DA033977-010
PubChem Upload Layout

- jQuery JavaScript renders all web HTML on client side
- Server APIs (C++) send & receive data from/to the database
- Client (jQuery Ajax) calls ‘masked’ server cgi
- RESTful communication between client & server
- Security primarily handled by server cgi

- Client responsible for web navigation, logic & state
- Server responsible for data queries, not state
Example client/server exchanges

1. Display table of user’s pending submissions
   - Send cookie, usr_id, request type to cgi
   - Receive status code, JSON table of data
   - Render display, make additional calls

2. Edit data within a submission
   - Send cookie, sub_id, request type, field, text
   - Receive status code
   - Render display (grey-out ‘Save’ button)
Read Data from Server

- Make request to server with arguments (& cookie):
  - action=get_subs
  - sub_id=16536, pgsise=10, startrec=21

- Receive response from server in JSON:

```json
{
  "response": [
    {
      "name": "return",
      "headers": ["code"],
      "data": ["pass"]
    },
    {
      "name": "subst_list",
      "headers": ["Record", "SID", "Version", "tid", "RegId", "Comments"],
      "data": ["21", "26701783", "8", "157215778", "54922.2", "PDB Accession Code 105M..."
    ]
  ]
}
```
Read Data from Server

- Parsed data from JSON
- Server response includes error code

<table>
<thead>
<tr>
<th>Record</th>
<th>SID</th>
<th>Version</th>
<th>tid</th>
<th>RegId</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>26701783</td>
<td>8</td>
<td>157215778</td>
<td>54922.2</td>
<td>PDB Accession Code 105M Sperm Whale Myoglobin N-Butyl Isocyanide At Ph</td>
</tr>
<tr>
<td>22</td>
<td>26701784</td>
<td>8</td>
<td>157215781</td>
<td>54922.3</td>
<td>PDB Accession Code 105M Sperm Whale Myoglobin N-Butyl Isocyanide At Ph 9.0 Oxygen Transport</td>
</tr>
<tr>
<td>23</td>
<td>26701785</td>
<td>8</td>
<td>157215784</td>
<td>54922.4</td>
<td>PDB Accession Code 105M Sperm Whale Myoglobin N-Butyl Isocyanide At Ph 9.0 Oxygen</td>
</tr>
</tbody>
</table>
Read Data from Server

• Make request to server with arguments (& cookie):
  • action=get_subs
  • sub_id=16536, pgsiz=10, startrec=21

• Receive response from server in JSON:

{ "response": [{ "name": "return", "headers": ["code"], "data": ["pass"] }, { "name": "subst_list", "headers": ["Record", "SID", "Version", "tid", "RegId", "Comments"], "data": ["21", "26701783", "8", "157215778", "54922.2", "PDB Accession Code 105M…"] }]

• Simple error code
• Data transfer is minimal
• Data on server can be arbitrarily large
Example client/server exchanges

1. Display table of user’s pending submissions
   • Send cookie, usr_id, request type to cgi
   • Receive status code, JSON table of data
   • Render display, make additional calls

2. **Edit data within a submission**
   • Send cookie, sub_id, request type, field, text
   • Receive status code
   • Render display (grey-out ‘Save’ button)
Edit/Save Data to Server

This is not a standard PubChem header, and will be treated as a result definition provided by the assay submitter: replicate1_readout_0_hr

<table>
<thead>
<tr>
<th>Hit Confirmation</th>
<th>Phenotype</th>
<th>replicate1_readout_0_hr</th>
<th>replicate1_readout_2_hr</th>
<th>replicate1_readout_4_hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result Type</td>
<td>STRING</td>
<td>FLOAT</td>
<td>FLOAT</td>
<td>FLOAT</td>
</tr>
<tr>
<td>Result Descr</td>
<td>Confirmed by individual siRNA in Bmax1_dLucBiological effect type</td>
<td>bioluminescence measurement</td>
<td>bioluminescence measurement</td>
<td>bioluminescence measurement</td>
</tr>
<tr>
<td>Result Unit</td>
<td>NONE</td>
<td>PERCENT</td>
<td>PERCENT</td>
<td>PERCENT</td>
</tr>
<tr>
<td>ResultAttrConc_Micromol</td>
<td>NONE</td>
<td>PERCENT</td>
<td>PERCENT</td>
<td>PERCENT</td>
</tr>
<tr>
<td>ResultCondResponseSeriesID</td>
<td>Result is Active Concentration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>andf</td>
<td>n</td>
<td>0.4</td>
<td>0.46</td>
</tr>
<tr>
<td>2</td>
<td>want</td>
<td>differentiated</td>
<td>23.1</td>
<td>0.45</td>
</tr>
<tr>
<td>3</td>
<td>sdfsdfs</td>
<td>sdfsdf</td>
<td>0.6</td>
<td>123</td>
</tr>
</tbody>
</table>

23.1
Edit/Save Data to Server

- Make request to server with args and data (& cookie):
  - action=replace_datarows
  - startrow=2, nrows=1
  - csv="2,wert,differentiated,23.1,345.8,43"

- Receive response from server in JSON:

- Render display

<table>
<thead>
<tr>
<th>Assay Data: please use the spreadsheet below to type, paste (Ctrl+v), or upload a file, and then close...</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td><img src="image-url" alt="Image" /></td>
</tr>
<tr>
<td><img src="image-url" alt="Image" /></td>
</tr>
</tbody>
</table>
Full Multi-Sheet Spreadsheet Support

- CSV (single sheet), Excel & OpenOffice
- Substance & BioAssay submissions
- Data upload & download
- Consistent with new web data editing

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PUBCHEM_EXT_DATASOURCE_REGID</td>
<td>PUBCHEM_SUBSTANCE_SYNONYM</td>
<td>PUBCHEM_EXT_DATASOURCE_SMILES</td>
<td>PUBCHEM_SUBSTANCE_COMMENT</td>
</tr>
<tr>
<td>2</td>
<td>my_sub1</td>
<td>GLUCOSE</td>
<td>C(C1C(C(C(O1)O)O)O)O)O)O</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>my_sub2</td>
<td></td>
<td>CCOC1=CC=CC=C1NC(=O)C2=CC3=CC=CC=C3C</td>
<td>This is a comment.</td>
</tr>
<tr>
<td>4</td>
<td>my_sub3</td>
<td>benzene</td>
<td>C1=CC=CC=C1</td>
<td></td>
</tr>
</tbody>
</table>
## Multi-Sheet Spreadsheet Example

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>XREF_TYPE</td>
<td>XREF_VALUE</td>
<td>XREF_ANNOTATION</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>PUBCHEM_AID</td>
<td>641 Screening assay</td>
</tr>
<tr>
<td>3</td>
<td>PUBCHEM_AID</td>
<td>858 Confirmatory assay</td>
</tr>
<tr>
<td>4</td>
<td>PUBCHEM_NCBI_TAXONOMY_ID</td>
<td>9606 human sequences</td>
</tr>
<tr>
<td>5</td>
<td>PUBCHEM_EXT_DATASOURCE_URL</td>
<td><a href="http://www.yoursite.com">http://www.yoursite.com</a></td>
</tr>
<tr>
<td>6</td>
<td>PUBCHEM_NCBI_GENE_ID</td>
<td>1812 DRD1 dopamine receptor D1 [ Homo sapiens ]</td>
</tr>
<tr>
<td>7</td>
<td>PUBCHEM_NCBI_PUBMED_ID</td>
<td>15974933 AllostERIC modulation of dopamine receptors, Schetz JA, Mini Rev Med</td>
</tr>
<tr>
<td>8</td>
<td>PUBCHEM_NCBI_PUBMED_ID</td>
<td>12021390 Dopamine receptor signaling, Neve KA, Seamans JK, J Recept Signal</td>
</tr>
</tbody>
</table>
Large Submissions

- **Substances**
  - ~10s of millions per submission
  - Separate processing to SGE farm
  - Reworked database/backend handling with chunks of data

- **Assays**
  - Data rows often approach 500K in one assay
  - Result definitions (data columns) in the 1,000s
  - Panel Assays group columns per gene/protein target
  - Reworked access/update for preview, data table & download
Batch Id Lookup Tools

PubChem Upload TOOLS

Gene Symbol to Gene ID Conversion Tool

Given a list of gene symbols, this tool will return a corresponding list of gene id's.

Note, however, that a gene symbol can be used to represent the same gene in several different species. Therefore, in order to make a unique mapping, we also require a taxonomy id. There are two ways to provide taxonomy ids:

1. Specify a single taxonomy id, in which case it will be assumed that all genes belong to the same species.
2. Specify a list of taxonomy ids - one for each gene symbol.

<table>
<thead>
<tr>
<th>Gene Symbol (one per line)</th>
<th>TaxonomyID (one per line)</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLOCK</td>
<td>9605</td>
<td>9675</td>
</tr>
<tr>
<td>A2M</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>CBL</td>
<td></td>
<td>967</td>
</tr>
<tr>
<td>NFAT5</td>
<td></td>
<td>10725</td>
</tr>
</tbody>
</table>

Look Up
Batch Id Lookup Tools

PubChem Upload TOOLS

Nucleotide Accession to Gene ID Conversion Tool

Given a list of nucleotide accessions, this tool will return a corresponding list of gene ID's.

<table>
<thead>
<tr>
<th>Nucleotide Accession (one per line)</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>NW_013620.2</td>
<td>18360</td>
</tr>
<tr>
<td>NW_001008504</td>
<td>4988</td>
</tr>
<tr>
<td>NW_007425.3</td>
<td>11596</td>
</tr>
</tbody>
</table>

Look Up
Curator/Power-user controls

Review or modify pending assay submissions
Select a row for further options

<table>
<thead>
<tr>
<th>ID</th>
<th>Submitter</th>
<th>Action</th>
<th>Status</th>
<th>RegID</th>
<th>Date</th>
<th>Curator</th>
<th>IS FTP</th>
<th>AutoOK</th>
</tr>
</thead>
<tbody>
<tr>
<td>15311</td>
<td>Linus Pauling (1084)</td>
<td>New Assay</td>
<td>Other</td>
<td>My-Unique-Assay-1... Atherosclerosis Modulators of C1 Receptors</td>
<td>2013/04/27</td>
<td>Ben Shoemaker</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>15251</td>
<td>Paul Kung (Burnham Center for Chemical Gen...)</td>
<td>New Assay</td>
<td>Primary Screening 680954 (1.1)</td>
<td>SBCCG-691-Artemis-P... uHTS identification of small molecule inhibitors of Artemis ...</td>
<td>2013/04/27</td>
<td>Ben Shoemaker</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>
PubChem Upload Feedback

• “I used it today and it looks really good. Works very smoothly and pages load a lot faster compared to the old one. And it’s great that the SD file can now be edited directly in the interface, saves me some downloading/deleting/re-uploading.” –PubChem Submitter/Journal Editor

• “By the way, tour new website looks great! Very clean look and feel.” –PubChem Submitter

• We welcome your feedback:

• pubchem-deposit-help@ncbi.nlm.nih.gov
Acknowledgments

• Asta Gindulyte
• Steve Bryant
• Yanli Wang
• Evan Bolton
• Paul Thiessen
• Siqian He
• Jeff Zhang
• Tugba Suzek
• Jiyao Wang
• PubChem Team