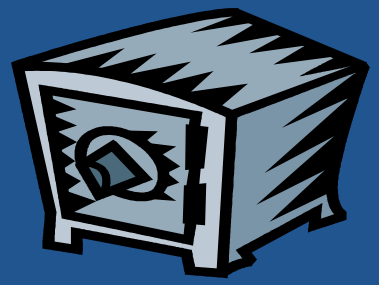
PubChem 2017
where are we now and where are we going?

Evan Bolton, Ph.D.
U.S. National Center for Biotechnology Information (NCBI)
Oct. 11, 2017

Global **In**gredient **A**rchival **S**ystem (GInAS) Meeting 2017 at USP

PubChem Resource

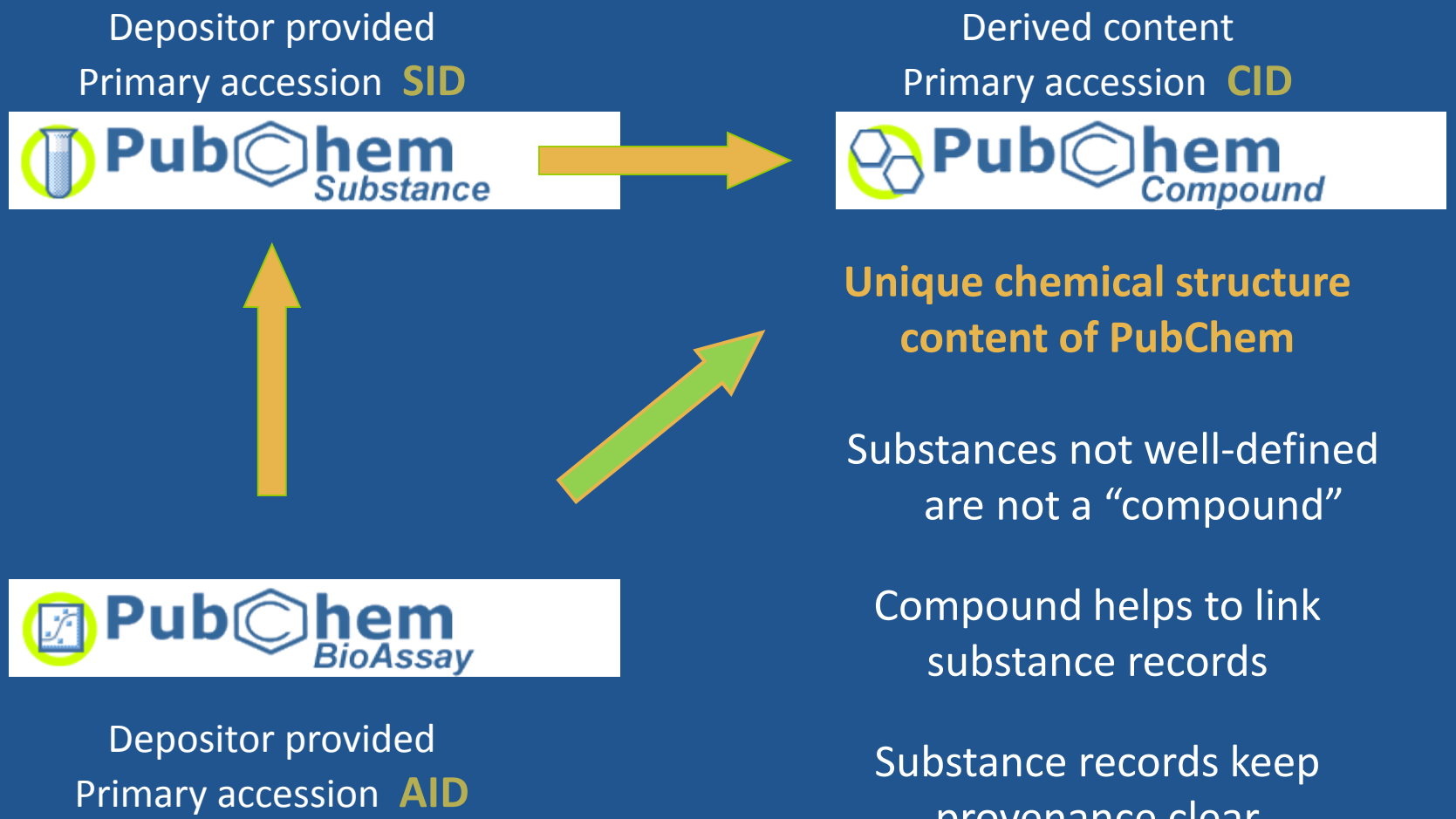
https://pubchem.ncbi.nlm.nih.gov/



What is PubChem?

- An open archive
  - anyone can contribute
    - biological experiments
    - chemical structures
    - (NCBI) cross-references
    - substance synonyms
    - link-back URLs
    - textual commentary
    - hierarchical annotation
  - records are versioned
  - voluntary data push
  - interconnects chemical biology resources
- A public resource
  - anyone can access
    - search, subset, select, ~~analyze~~, download
  - integrated
    - biomedical literature, sequences, pathways, patents, ontologies, etc.
  - programmatic layers
    - URL-based interfaces
    - NCBI Entrez Utilities
    - PUG [XML, VIEW, SOAP, REST, RDF]
    - PubChem Widgets
      [embed PubChem into your web pages]

PubChem as an archive



Unique chemical structure content of PubChem

Substances not well-defined are not a “compound”

Compound helps to link substance records

Substance records keep provenance clear

Why does a user come to PubChem?

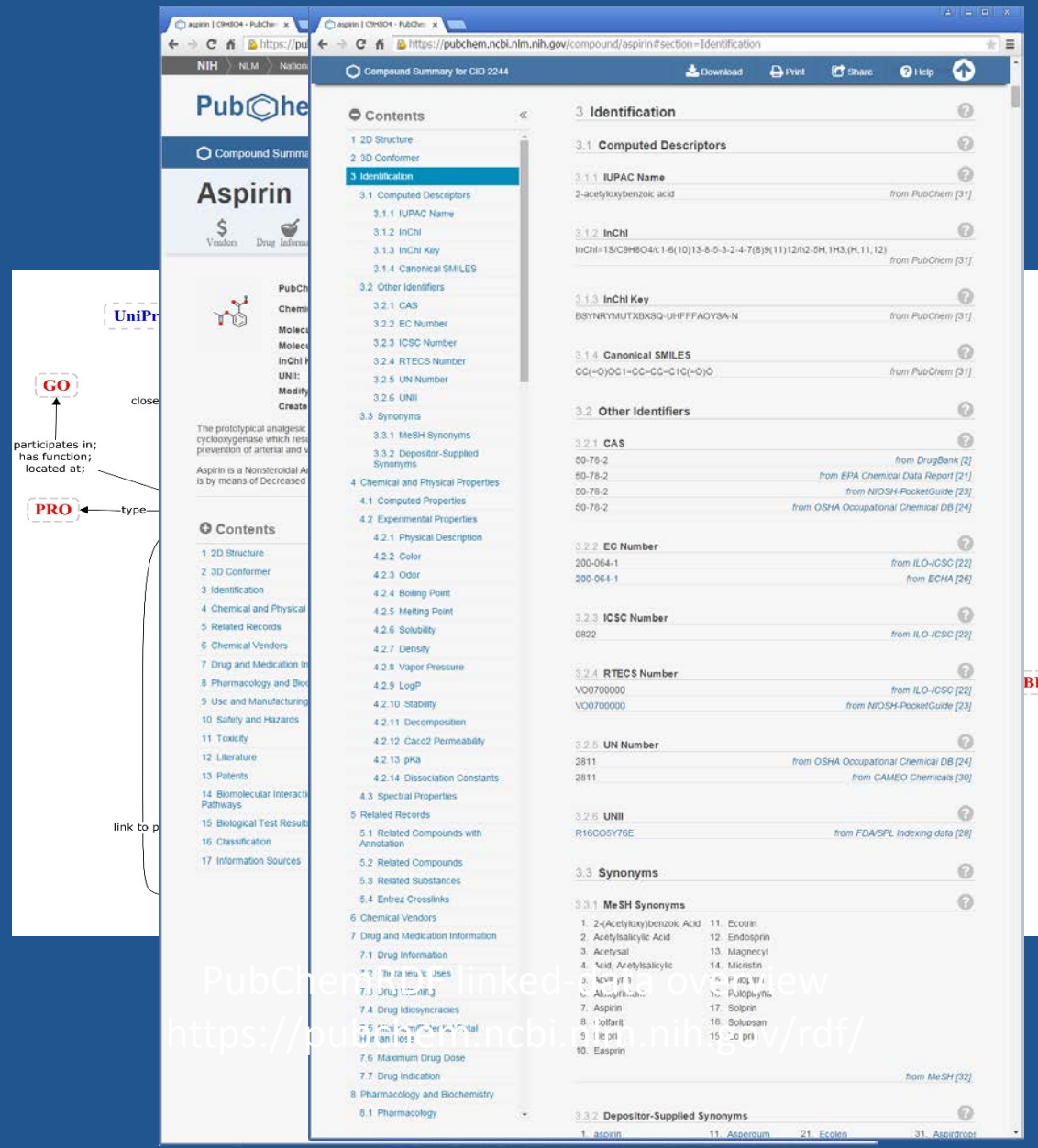
- Google/Yahoo/Baidu suggested it
- I want to buy molecule ‘X’
- Publications about molecule ‘X’
- Patents/Biological activities

Image credit: http://blogs.egu.eu/network/palaeoblog/2012/10/31/why-bother-communicating/

- What is known about the molecule?
  - Physical properties
  - Pharmacology
  - Biological activity
  - Safety information
  - Spectroscopy
  - Toxicity
  - Pathways
  - Etc.
- Deep questions query can decipher

PubChem data complexity

- Many links between large record collections
  - ~230M Substances <-> ~90M Compounds
  - ~90M Compounds <-> ~90M Compounds
  - ~230M Bioactivities <-> ~3M Substances
  - ~230M Bioactivities <-> ~2M Compounds
  - ~230M Bioactivities <-> ~1M BioAssays
  - ~10M PMIDs <-> ~100K Compounds
  - ~3M Patents <-> ~30M Substances
  - ~3M Patents <-> ~15M Compounds
- Sparse and dense data and/or linking
- New types of data, links, and metadata on a regular basis



PubChem RDF linked data overview
https://pubchem.ncbi.nlm.nih.gov/rdf/

Scientific information is a tale of two cities …

- Any (bioactivity) data system is a balancing act between two factions
  - Data creators
  - Data users
- Each has their own needs
  - Creators want to capture all details necessary to reproduce experiment
  - Users want ease of navigation and comparison between experiments
- PubChem tries to please both but caters to the users

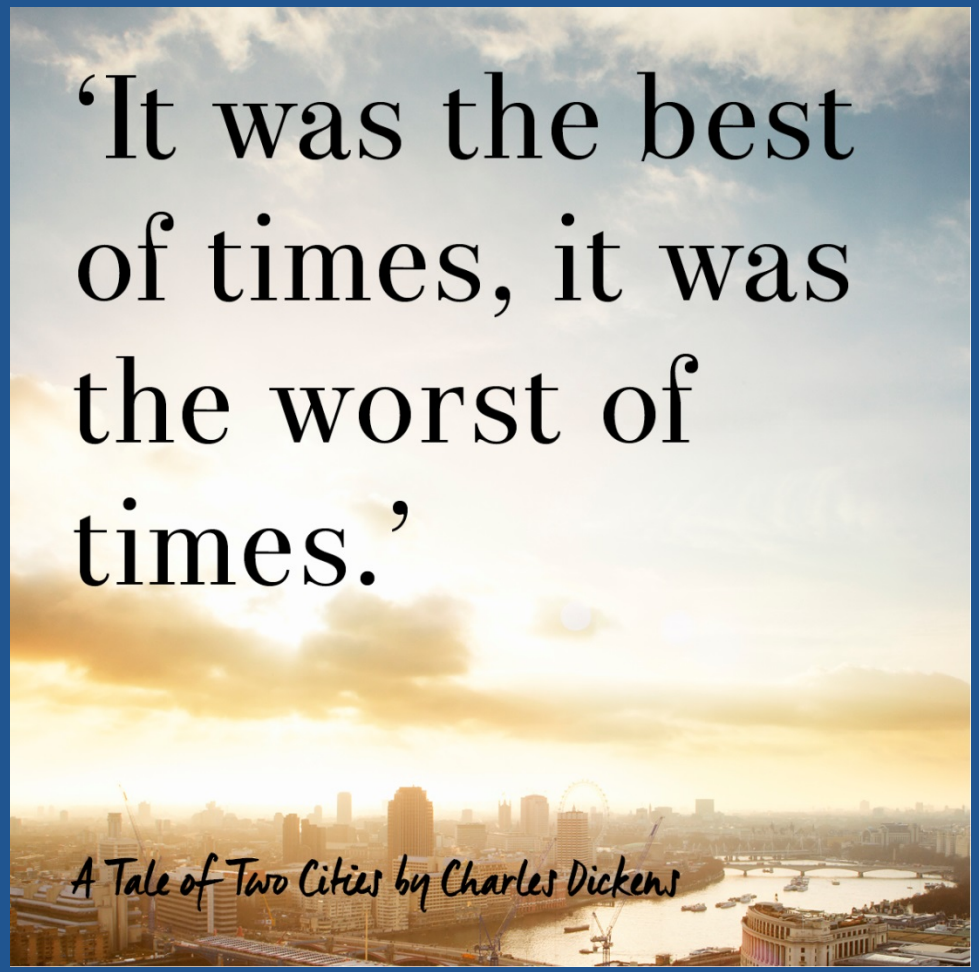
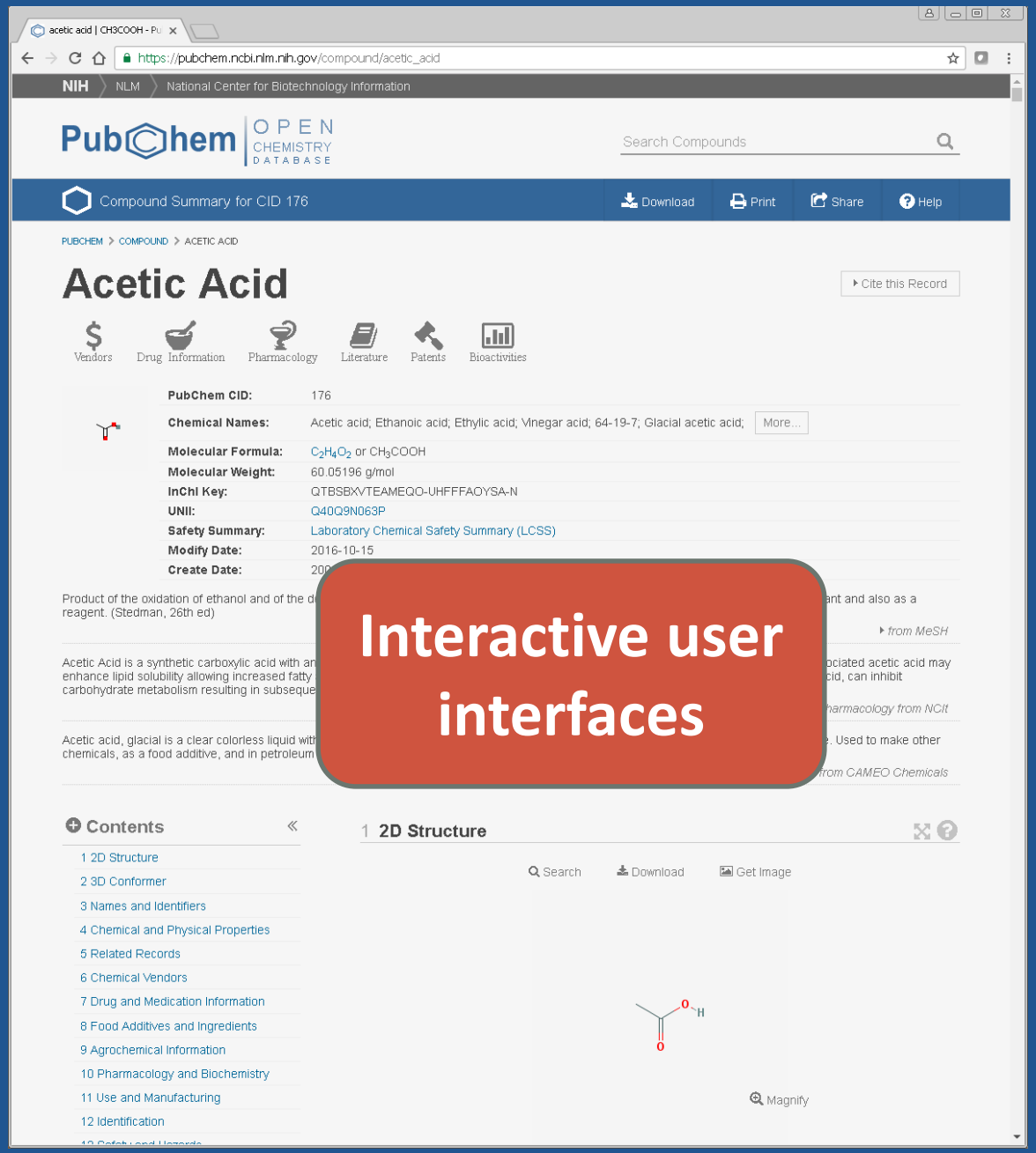


Image credit:
http://redonline.cdnds.net/main/thumbs/18557/a_tale_of_two_cities_best_first_lines_from_books_redonline.co.uk-2.jpg

PubChem as a public interface

Machine accessible data



Interactive user interfaces

PubChem Programmatic Services
PUG REST

http://pubchem.ncbi.nlm.nih.gov/rest/pug
prolog

/compound/name/vioxx
input

/property/InChI
operation

/TXT
output

https://pubchem.ncbi.nlm.nih.gov/pug_rest/PUG_REST_Tutorial.html

Scientific information has many CAVEATS!

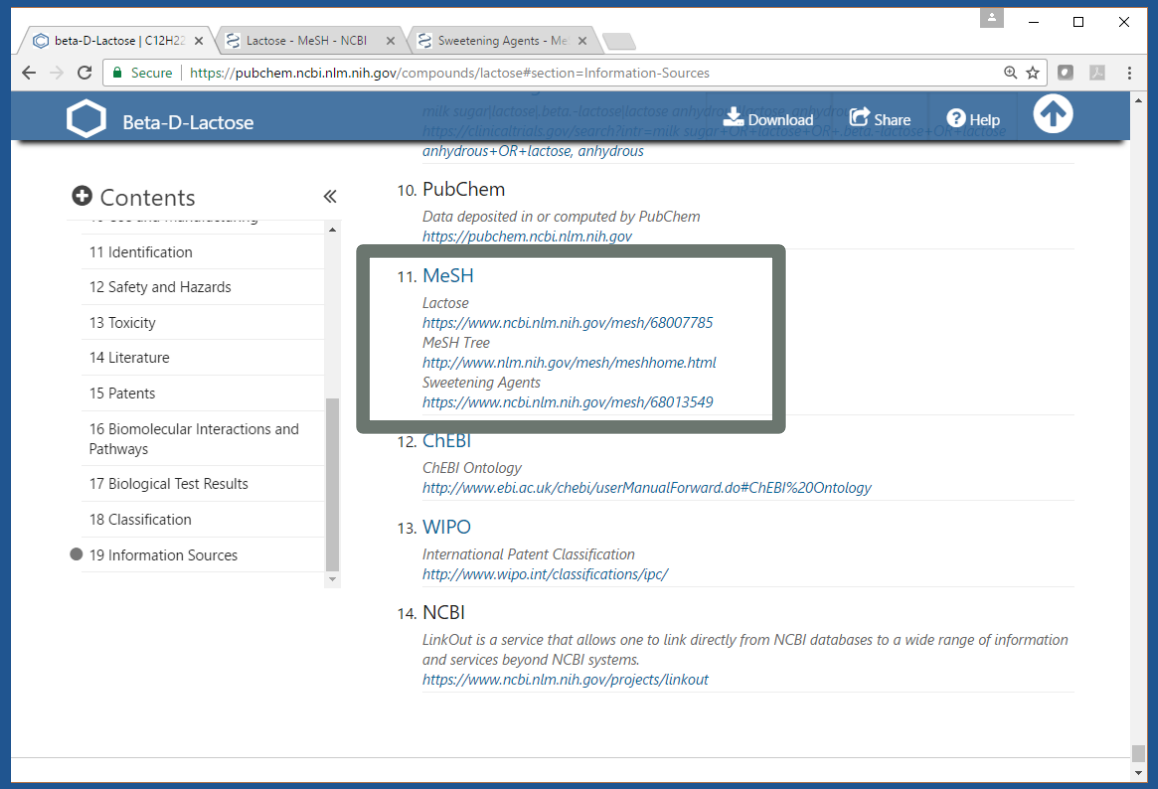
- Chemical information is a bit of a mess and can be rather nuanced
  - Names, names, and more names (+300M in PubChem)
    - Some standard names are not open and cannot be used/verified without $$$
  - Name/structure associations vary by use case (many overlapping)
    - Acetic acid vs. Acetic acid tri-hydrate
    - Formaldehyde: (gas) vs. Formalin (liquid , 40% formaldehyde w/ water)
    - Sulfuric acid: SO_3 (gas) vs. H_2SO_4 (liquid)
    - Glucose: L/D, ring open/closed (p vs. f), alpha/beta/both vs. monohydrate
    - Large corpus in the 'wild' .. data source dependent nuances
- One needs to verify with annotation source(s) as to what is meant
  - i.e., is this the form of the chemical I care about?

PubChem aggregates content by chemical structure

- Helps users to locate desired content
  - Other websites
  - Categorized annotation
  - Related chemicals (w/ annotation)
- Helps to bridge text world and chemical structure world
- More data means more errors
  - Separation of wheat from chaff important and never-ending battle
- Focus is on researchers (+educators)

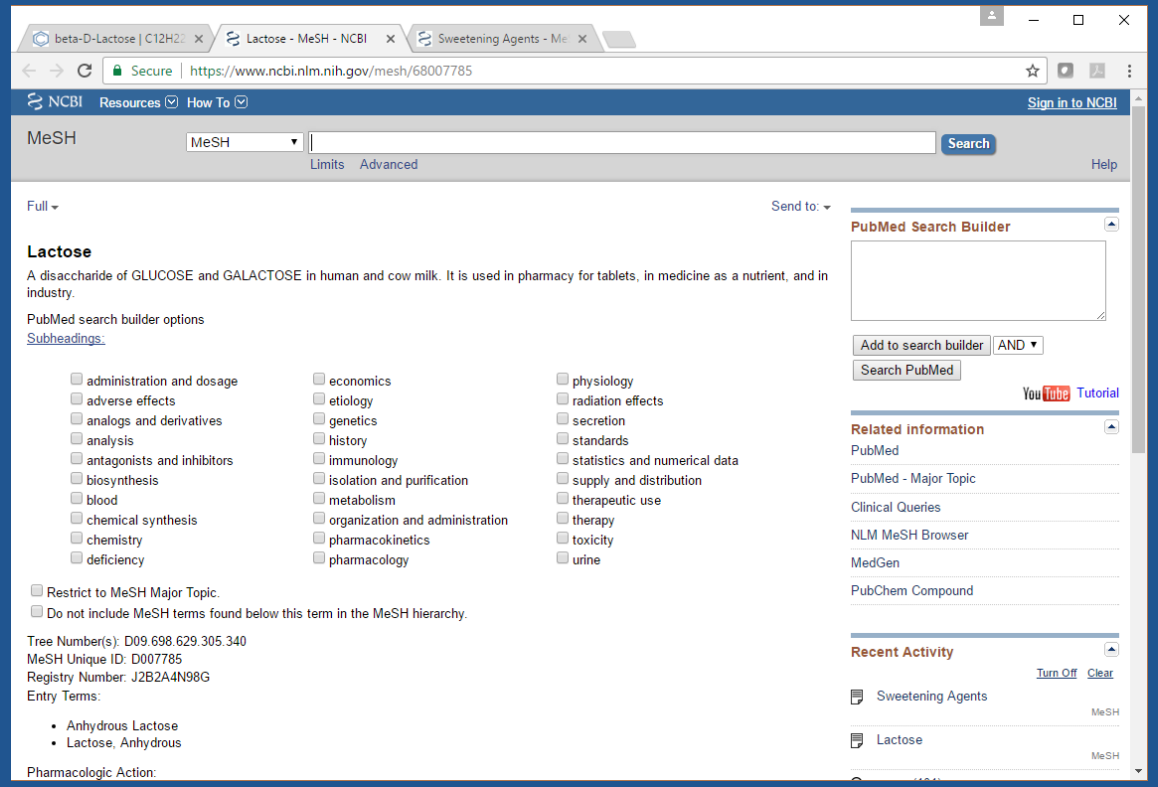
- PubChem does not curate data
  - Authoritative and curated sources are very important to PubChem
  - Algorithm consistency checking
    - Not sure who is right
    - Detects errors in curated content
- Concepts change on context
  - Chemical names mean different things to different scientists
  - Can be whole classes of things
  - Much to do here…

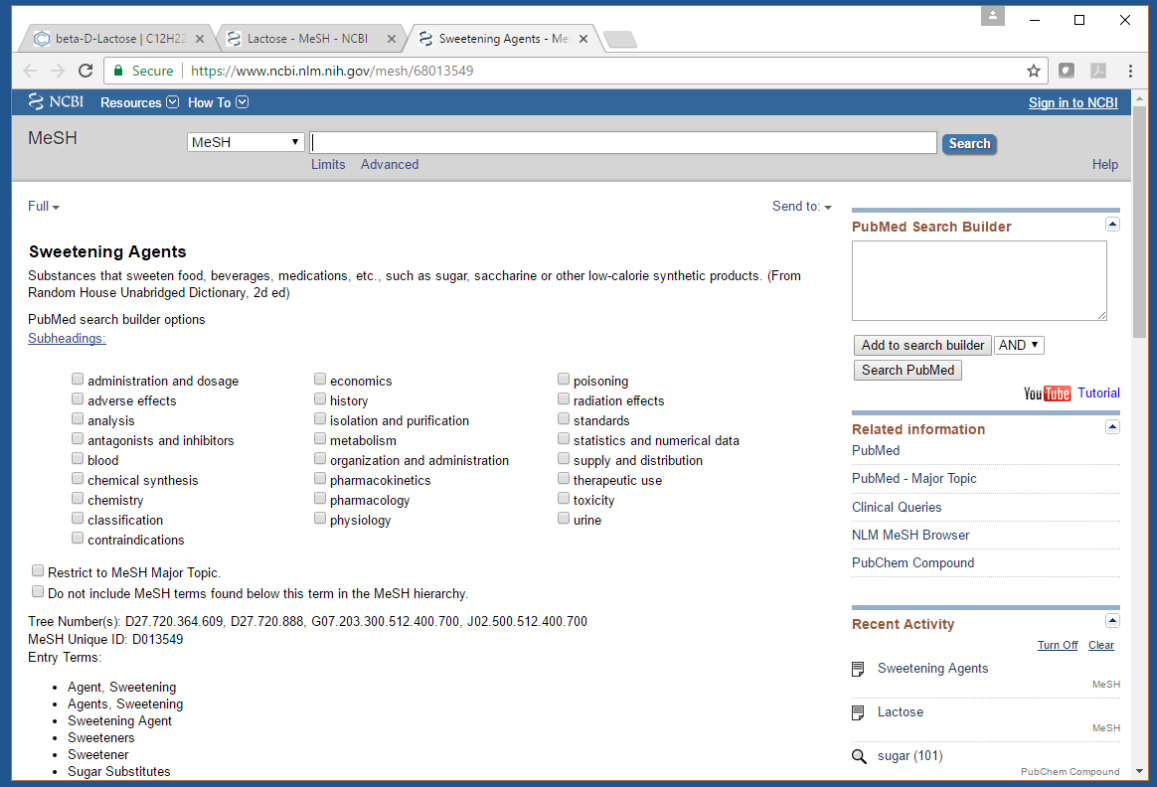
Semantics and concepts



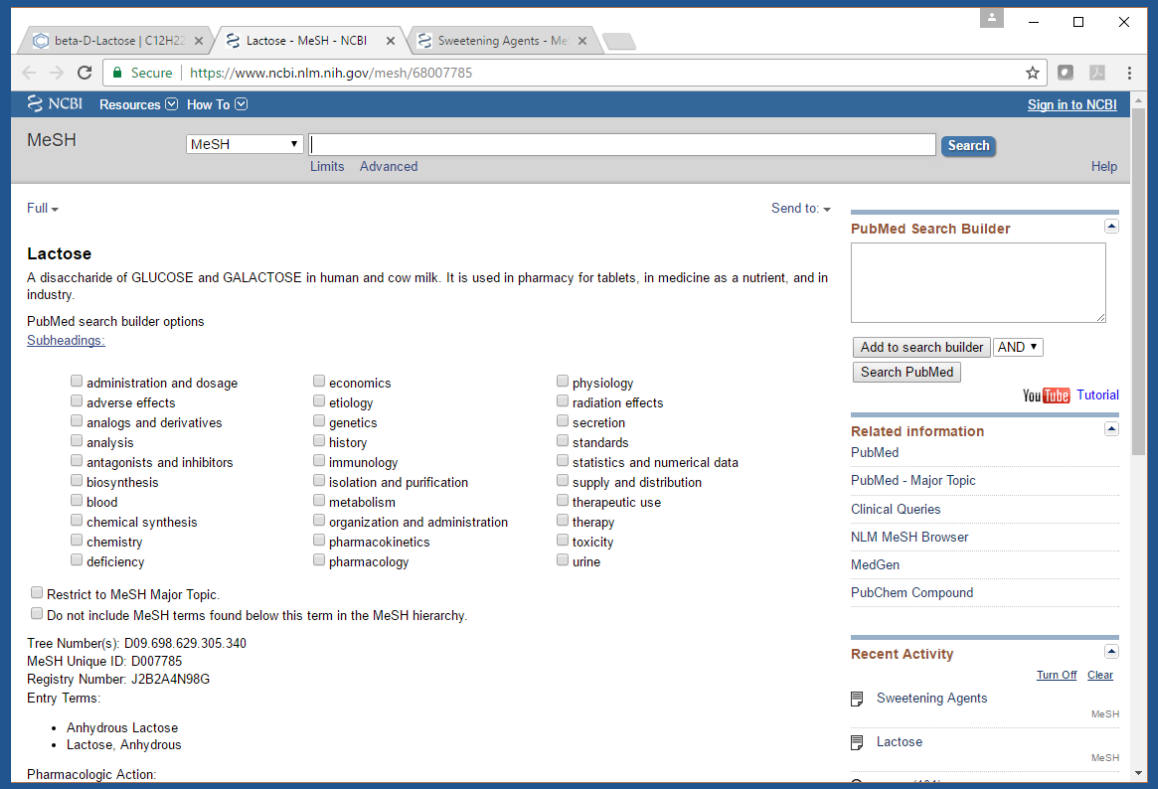
- Why so many different links to MeSH?
  - Annotations and names are wild
  - Cause PubChem many headaches

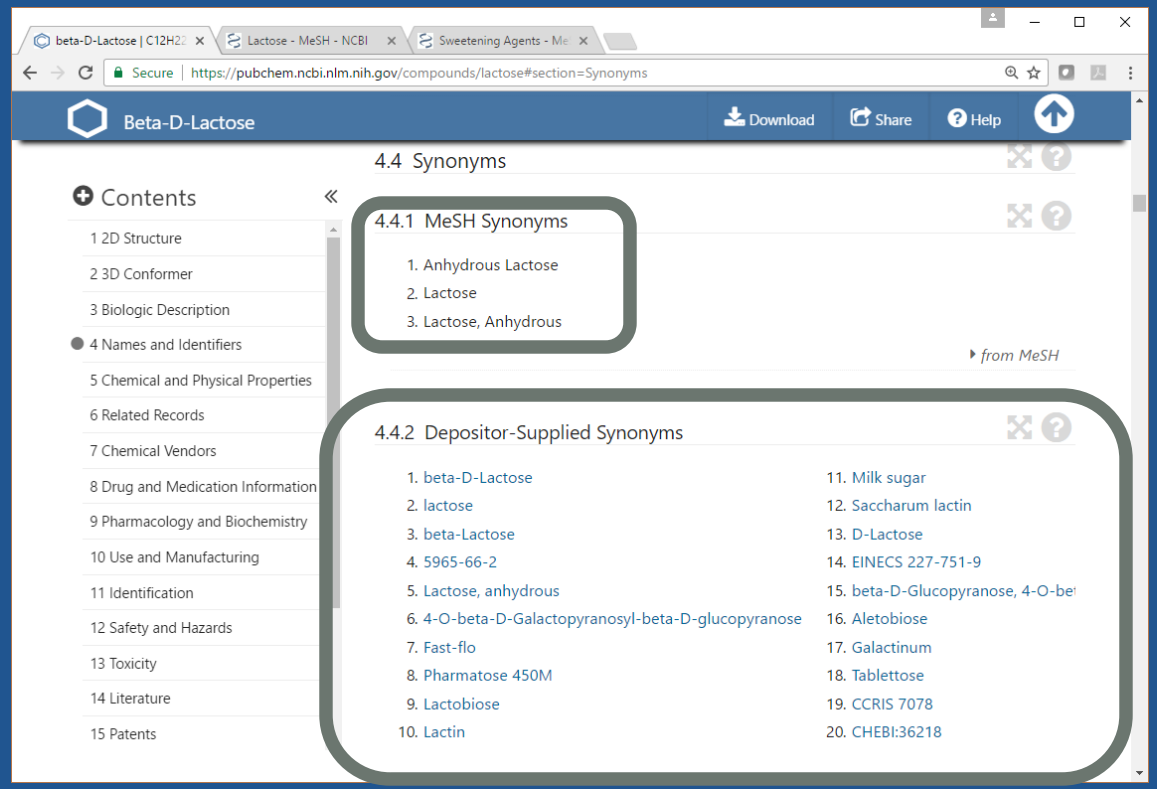
Semantics and concepts



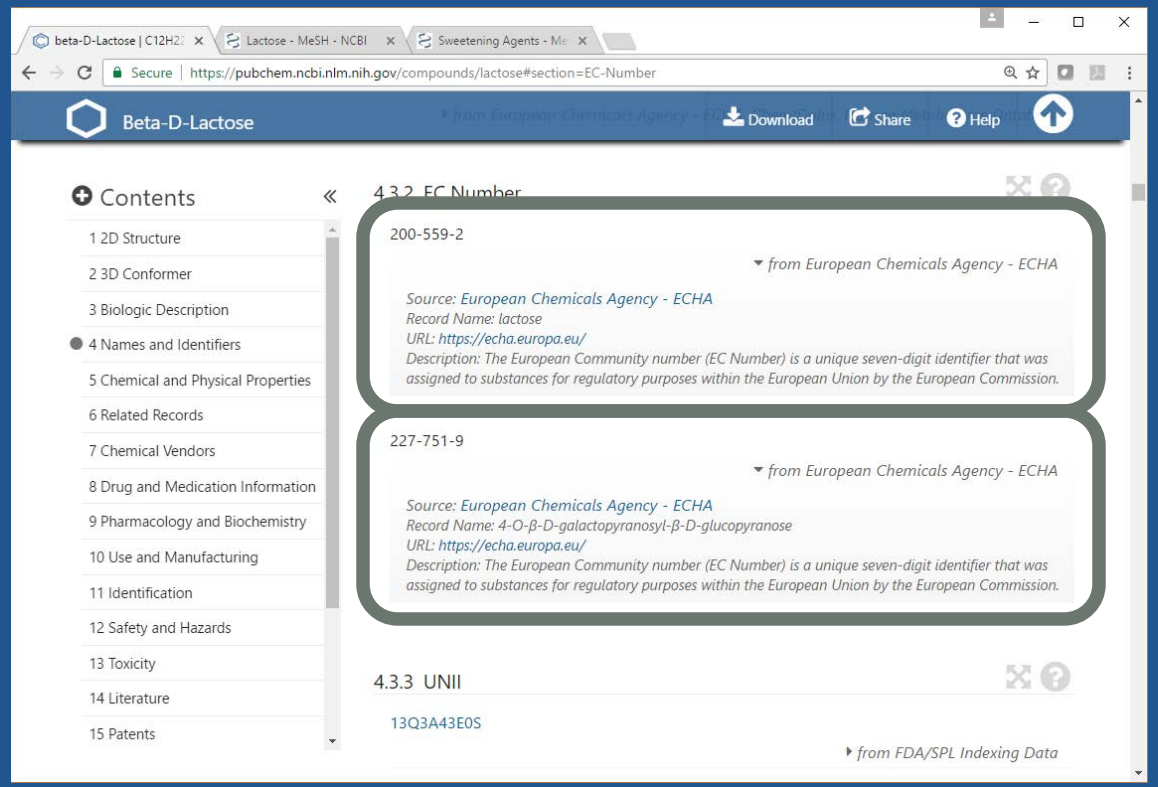


Semantics and concepts





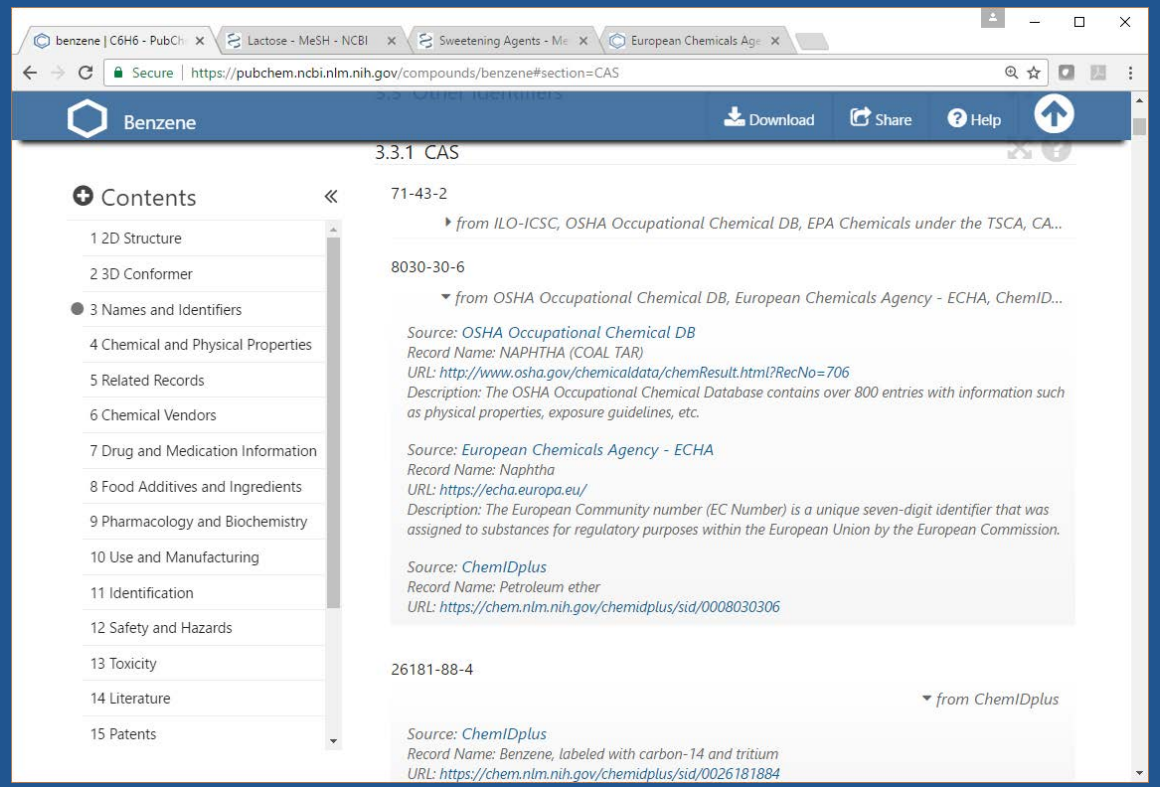
Semantics and concepts



- Authoritative sources have different use cases for the same chemical placed into different concepts
- How can we divide up the annotation variation?
- How can we better represent appropriate content to our users?

Slowly moving towards Concepts-based approach

- Modifying MeSH (to make a PubChem MeSH) by:
  - Expanding names
  - Adding concepts
- Often many concepts per chemical structures [E.g., Benzene, Coal Naptha]
- Use concepts to map annotation
- Give concepts own page to separate out their annotations



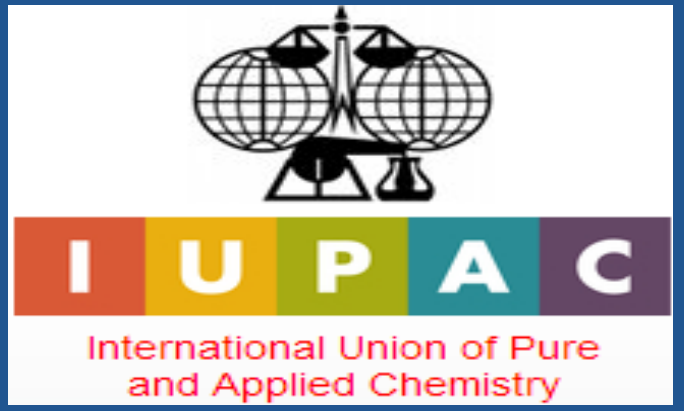
Semantics, terminologies, and standards

- PubChem is increasingly involved with standards
  - IUPAC (International Union of Pure and Applied Chemistry) InChI
  - Chemical structure guidelines, best practices, and normalization approaches
  - Terminology projects (IUPAC Gold Book)
- Computer understanding requires terminologies and semantic approaches
  - Toxicology information, health and safety, experimental procedures, …

- To make progress in representing scientific information/knowledge, a fair bit of work will be necessary
  - “It takes a village” - a communal effort .. what can the ‘village’ do?

Chemical information standards gaining traction

- InChI ← an acronym
  - International Chemical Identifier
- Standard
  - Initially created by NIST
  - Under auspices of IUPAC
  - Open source, non-proprietary
- Algorithm
  - Normalizes chemical representation
  - Includes ‘hashed’ form called an InChIKey

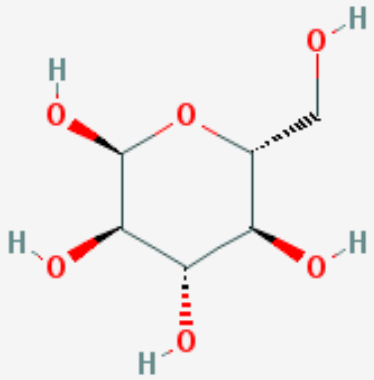


InChI is a string

Version Type
Chemical formula
Connectivity
Charge&Proton
Stereochemical
Other (e.g., Isotopic)

InChI=1S/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1

"layered" line notation



alpha-D-Glucose

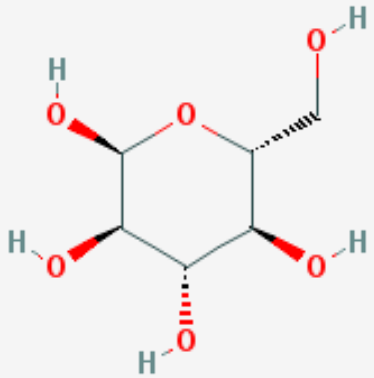
InChIKey is a “hashed” InChI

- Fixed length and search engine friendly InChI
- May allow for ‘secure’ lookup of a chemical

Chemical formula
& Connectivity
Stereochemical &
Hydrogens &
Other (e.g., Isotopic)
Type Version
Charge

WQZGKKKJIJFFOK-DVKNGEFBSA-N

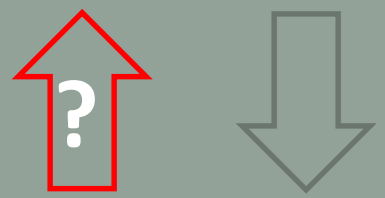
“layered” line notation



alpha-D-Glucose

InChIKey can be a ‘secret’

InChI=1S/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1

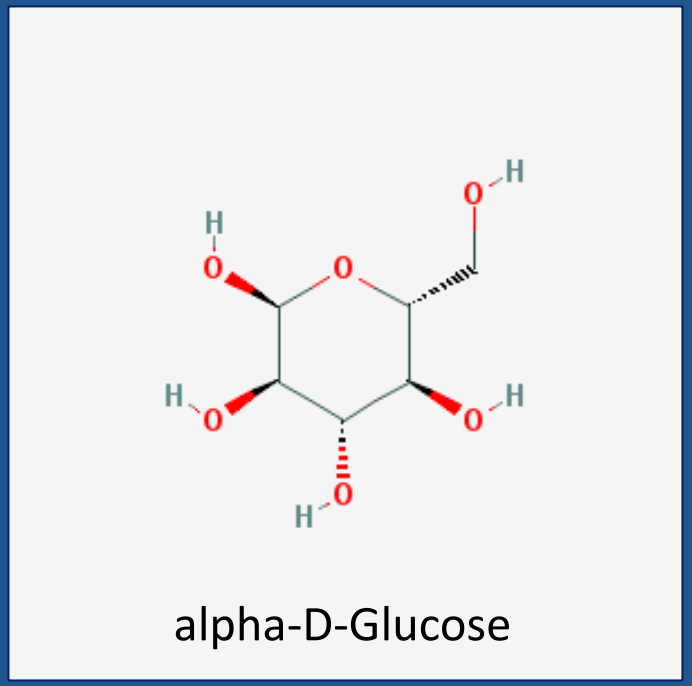


WQZGKKKJIJFFOK-DVKNGEFBSA-N

There is no chemical information in an InChIKey … in that, if you do not know the InChI, you cannot convert the InChIKey back into a chemical structure

What about SMILES?

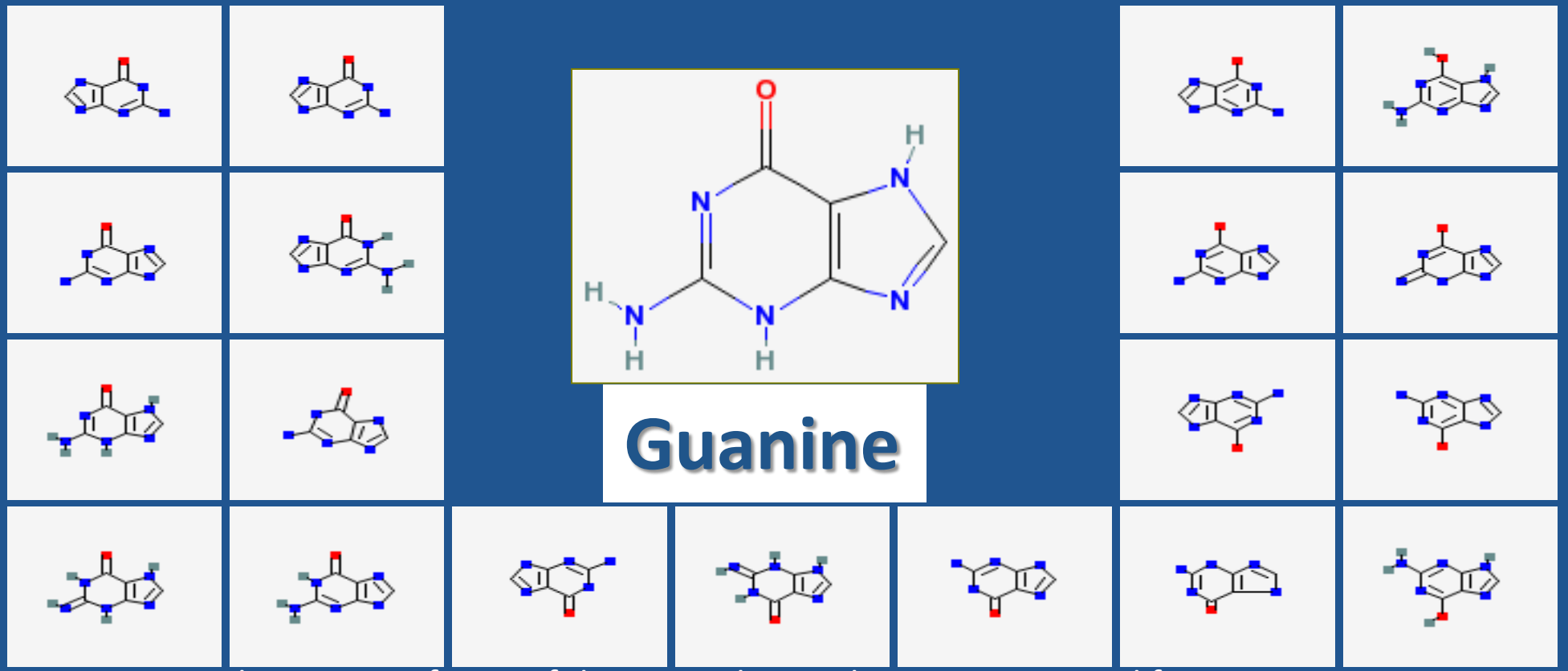
C([C@@H]1[C@H]([C@@H]([C@H]([C@H](O1)O)O)O)O)O



alpha-D-Glucose

- SMILES is a popular “industry standard” line notation
  - But not a published standard
- Every vendor has its own implementation
  - Differences in aromaticity models can lead to structure corruption
- Cannot reliably compare strings
  - Different software packages can make different strings for same structure
- No structure normalization
  - Different structural representations can yield different strings

Different equivalent (tautomer) forms have different SMILES but same InChI



Tautomers and resonance forms of the same chemical structure are prolific

Map your structures to PubChem

- PubChem FTP site provides a complete tab-delimited file containing all ~90M chemicals
  - PubChem Compound CID, InChI, InChIKey

  ftp://ftp.ncbi.nlm.nih.gov/pubchem/Compound/Extras/CID-InChI-Key.gz
- Easy to get mapping between your structures and PubChem
  - Just compute InChI/Key and then do string comparison (or DB join)

4.5GB file
ftp://ftp.ncbi.nlm.nih.gov/pubchem/Compound/Extras/CID-InChI-Key.gz

How does PubChem use InChI/Key?

- Deposition structure format
- Computed descriptor of chemicals
- Text-based chemical structure lookup
- Structure export format
- Structure search input
- Standardization service input/output
- Programmatic service input/output
- Semantic linked data

Image credit:
https://farscapedevelopment.files.wordpress.com/2012/07/eureka.gif

PubChem is going though some …

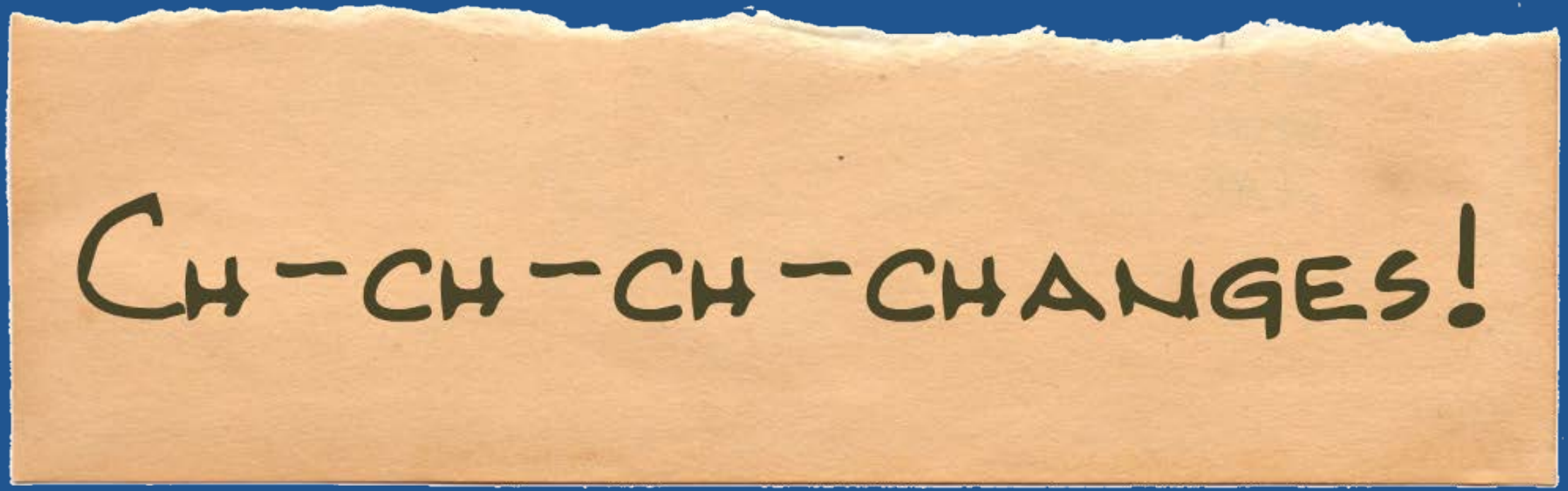
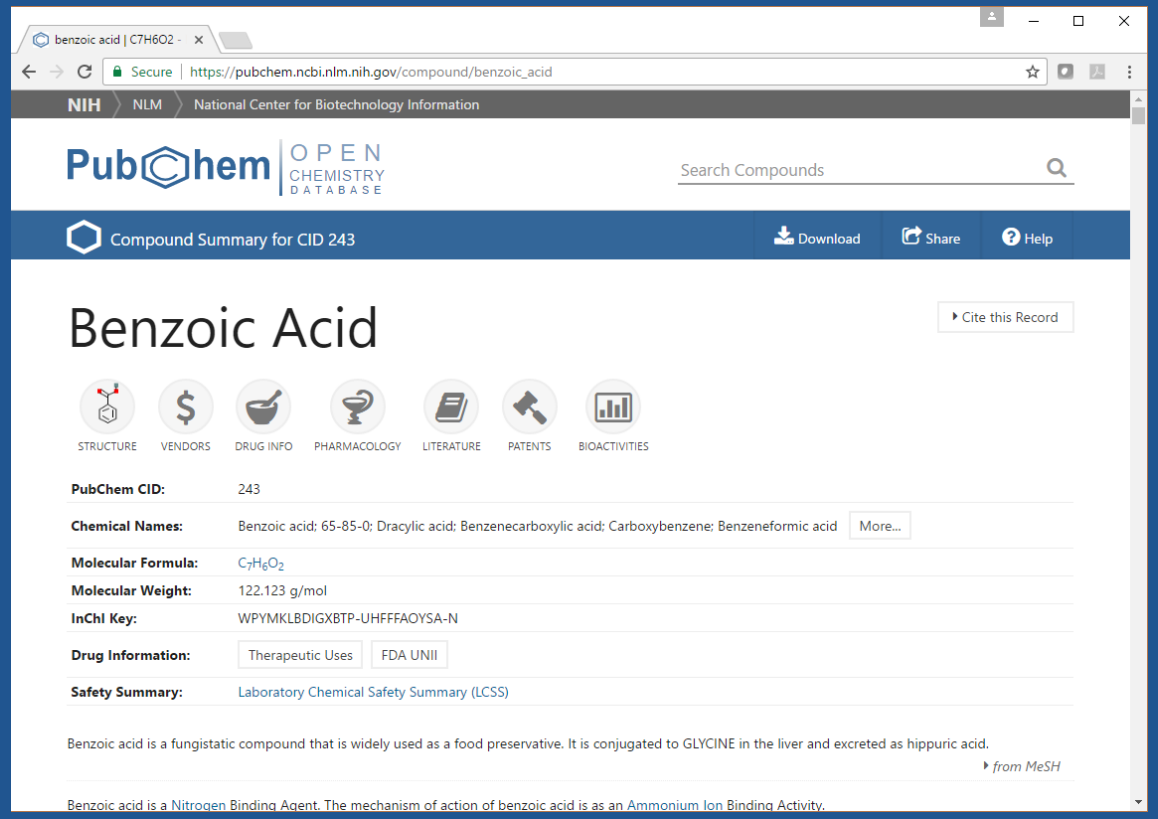


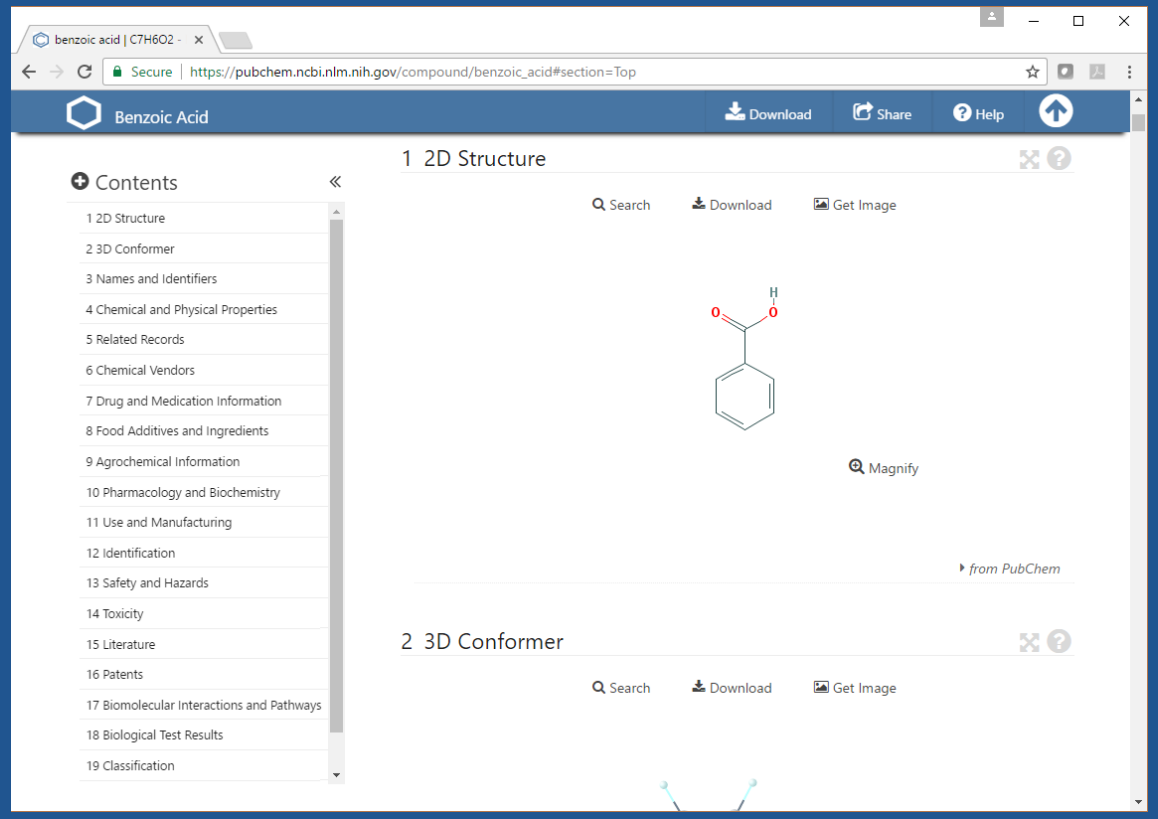
Image credit:
http://www.fortemusiceducation.com/wp-content/uploads/2016/06/Ch-ch-ch-changes.png

Technology revamp continues



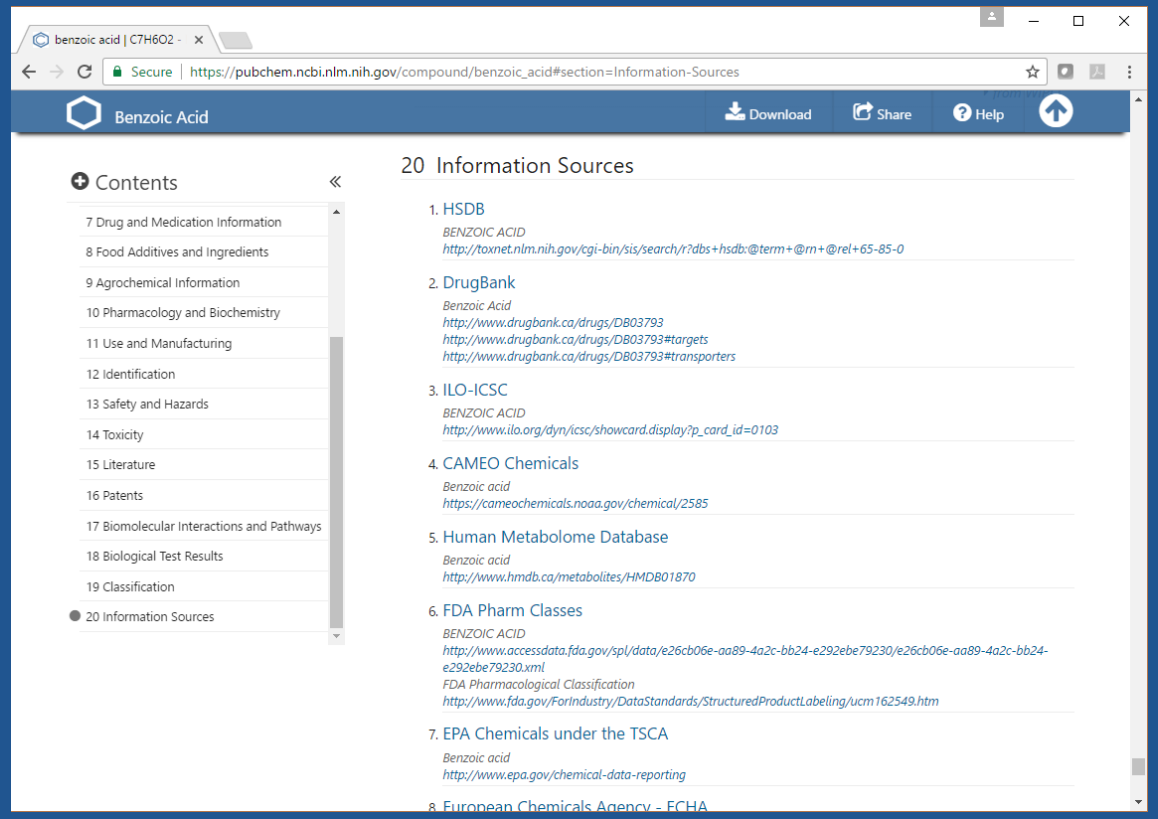
- Continue to refine, improve, add, modify, tweak, remove, expand …

Technology revamp continues



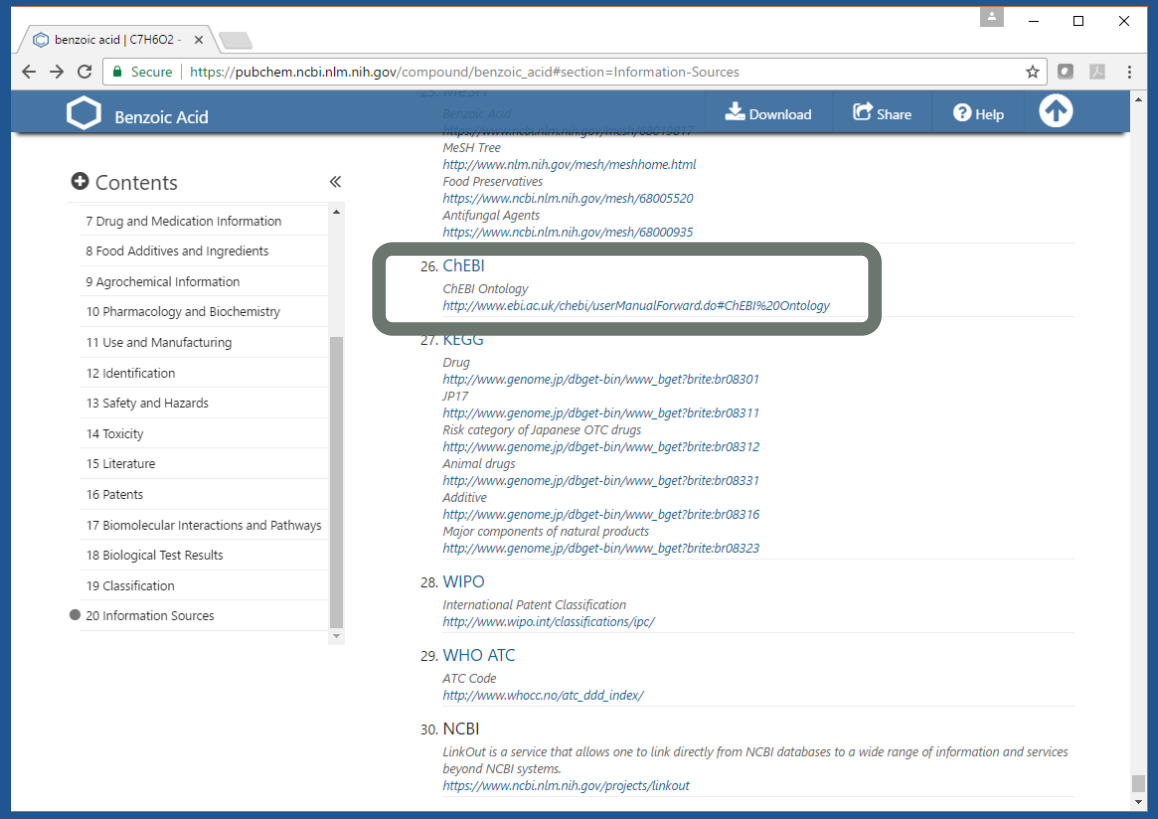
- Continue to refine, improve, add, modify, tweak, remove, expand …
- Adding new major sections as needed .. sometimes with fission/fusion events

Technology revamp continues



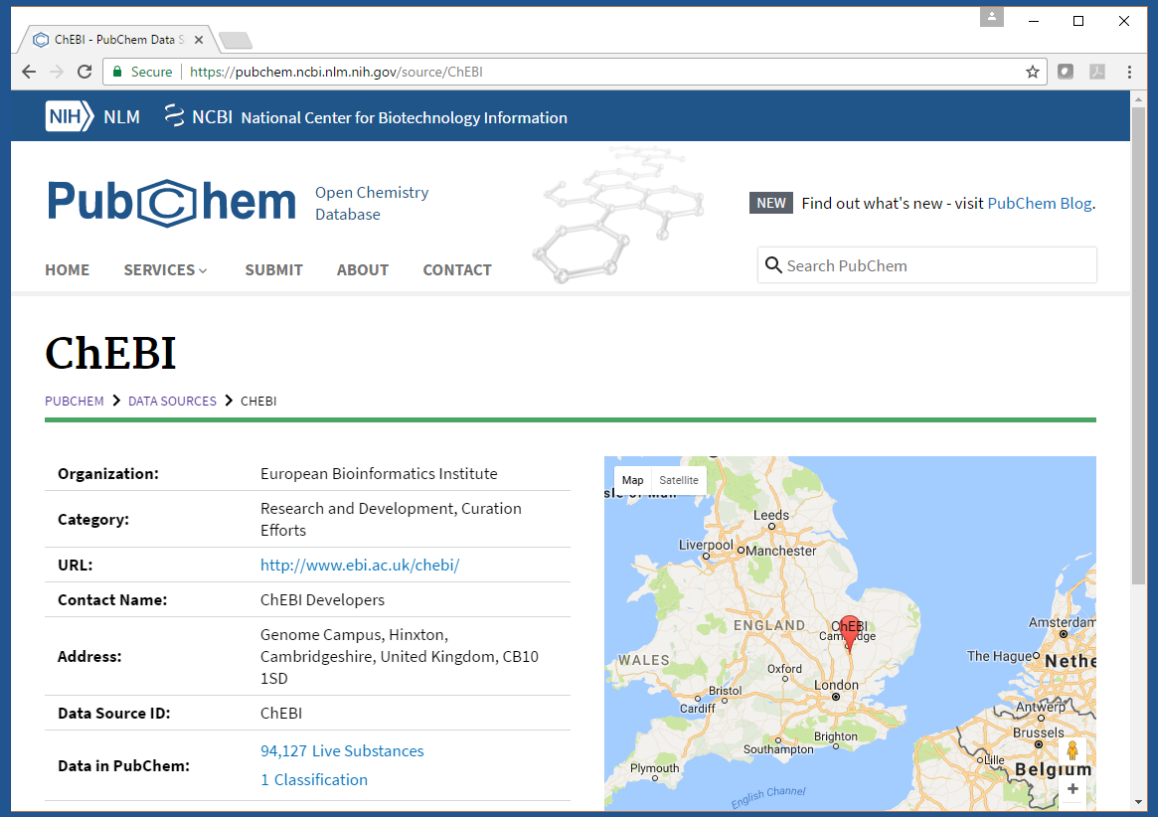
- Continue to refine, improve, add, modify, tweak, remove, expand …
- Adding new major sections as needed .. sometimes with fission/fusion events
- Information Sources section continues to expand with new data sources but also to enhance linking

Technology revamp continues



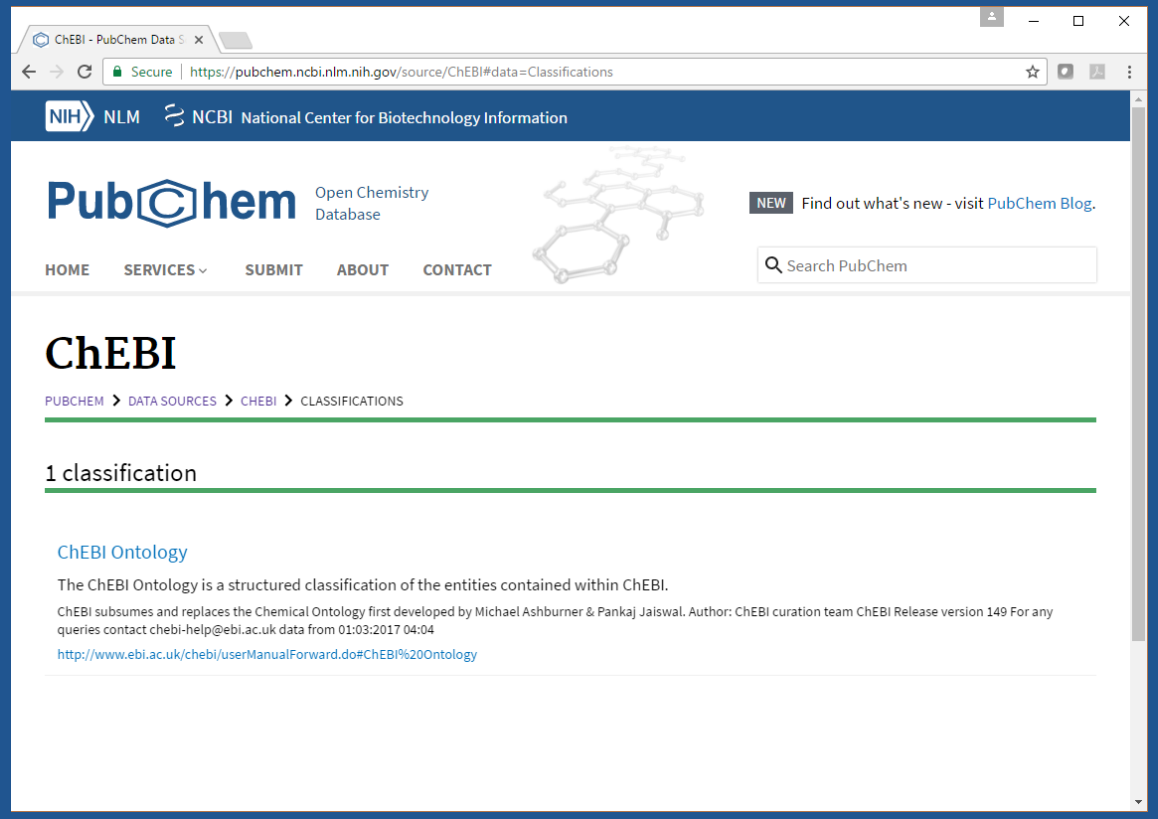
- Continue to refine, improve, add, modify, tweak, remove, expand …
- Adding new major sections as needed .. sometimes with fission/fusion events
- Information Sources section continues to expand with new data sources but also to enhance linking
- PubChem highlights quite a number of information sources

Technology revamp continues



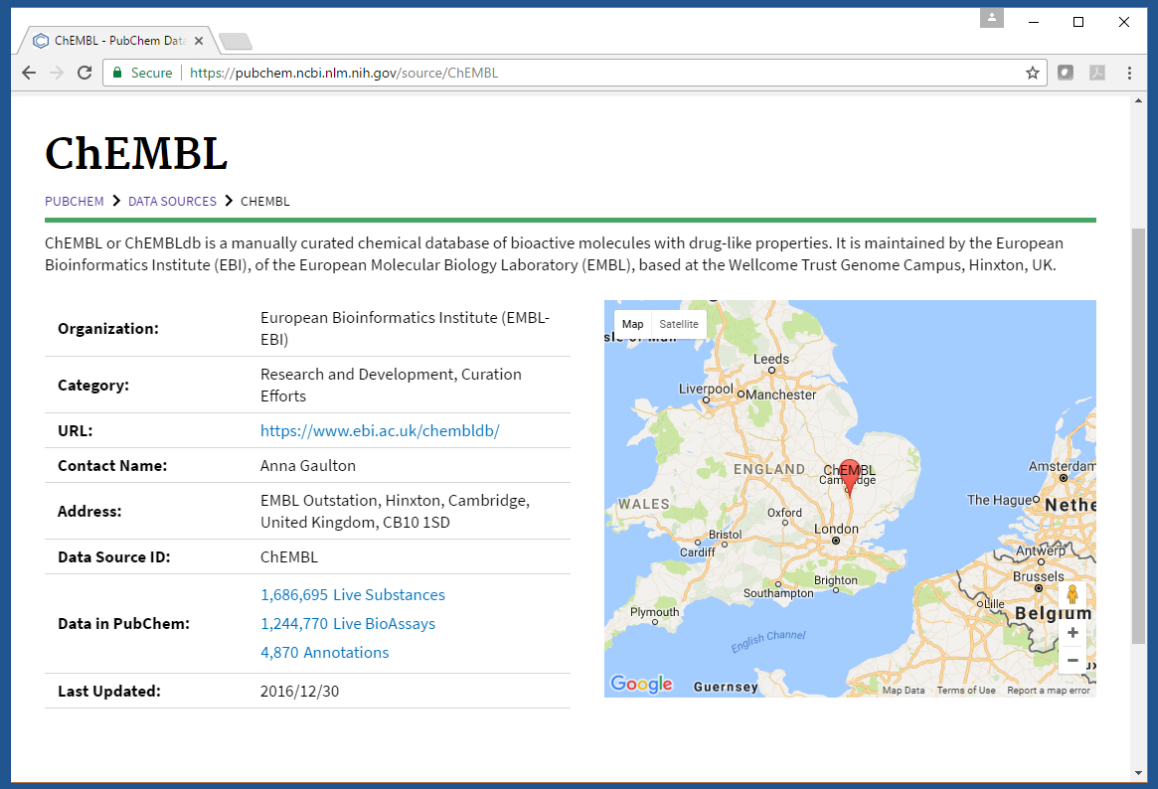
- New Data Source pages released in the last year

Technology revamp continues



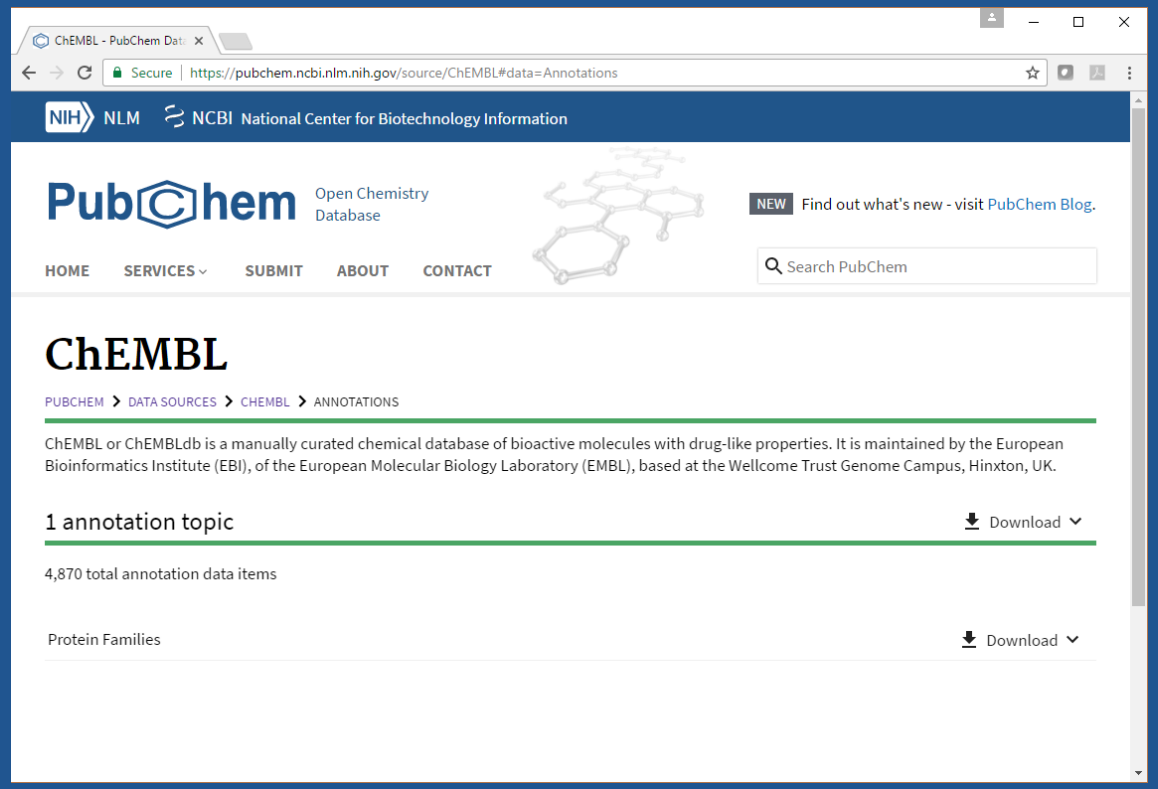
- New Data Source pages released in the last year
- Provides links to all annotation and classification content from a source

Technology revamp continues



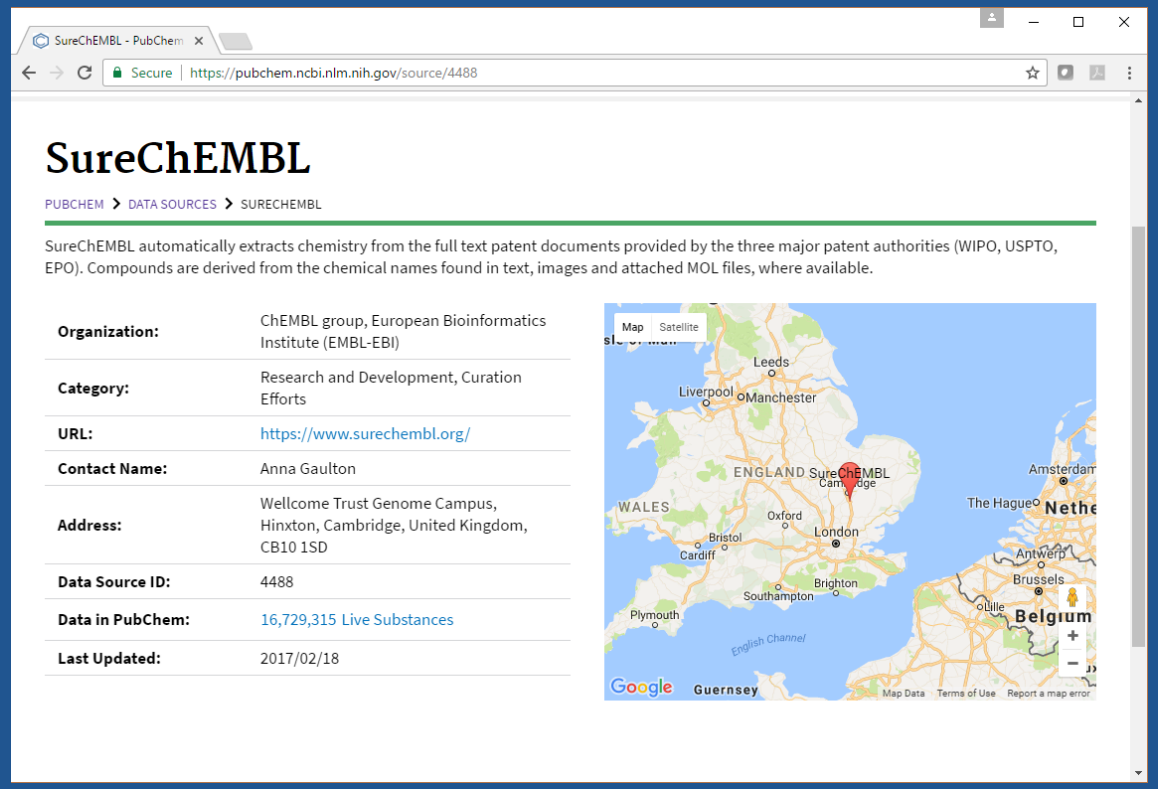
- New Data Source pages released in the last year
- Provides links to all annotation and classification content from a source

Technology revamp continues



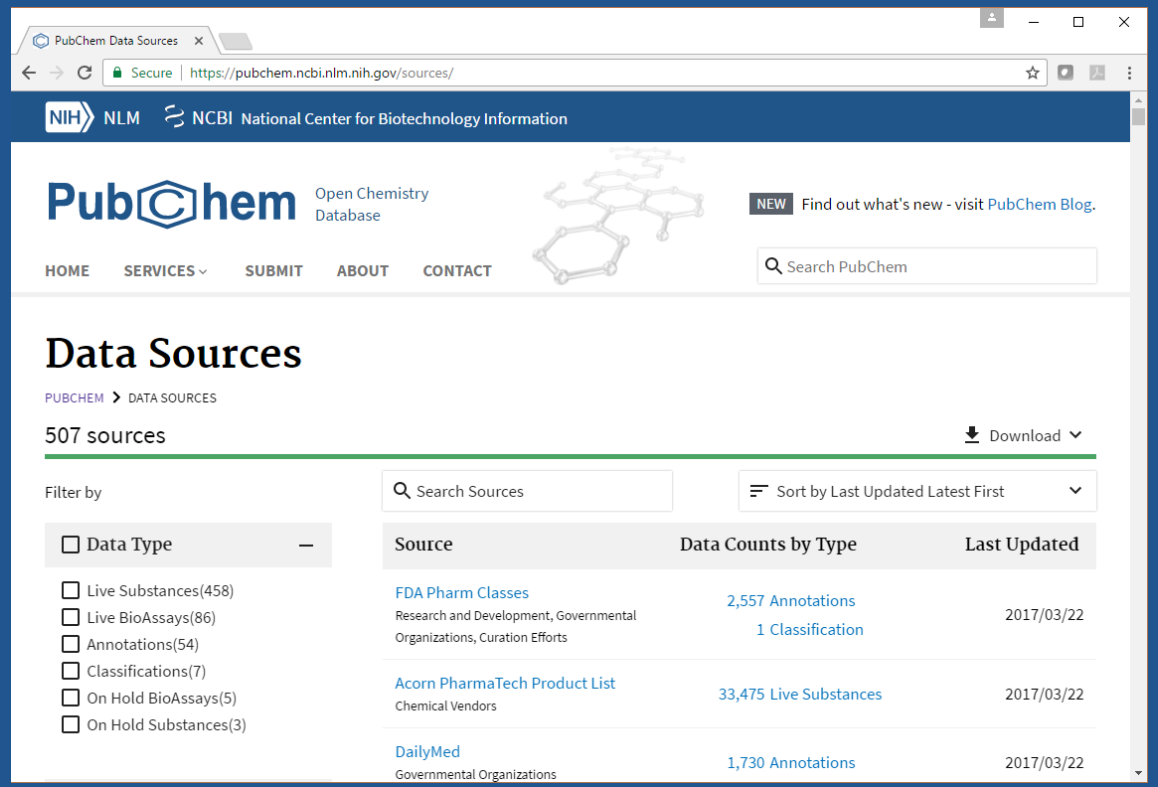
- New Data Source pages released in the last year
- Provides links to all annotation and classification content from a source

Technology revamp continues



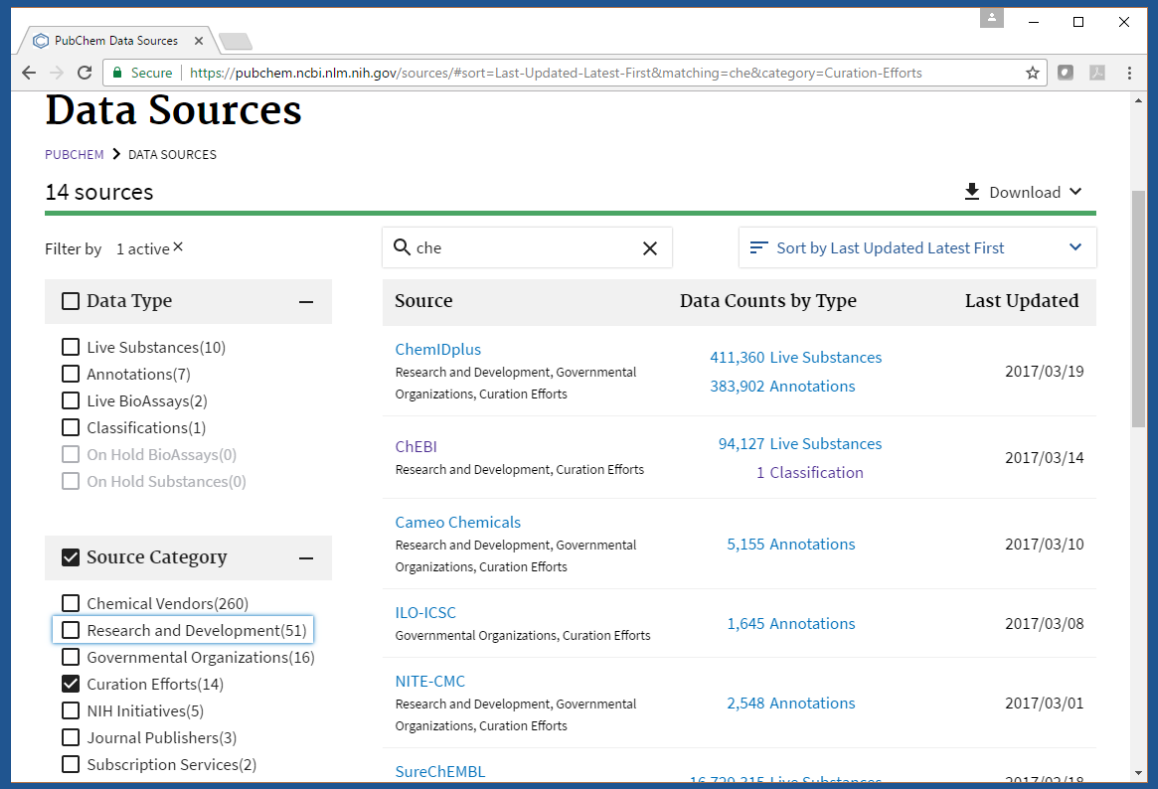
- New Data Source pages released in the last year
- Provides links to all annotation and classification content from a source
- SureChEMBL is available too!

Technology revamp continues



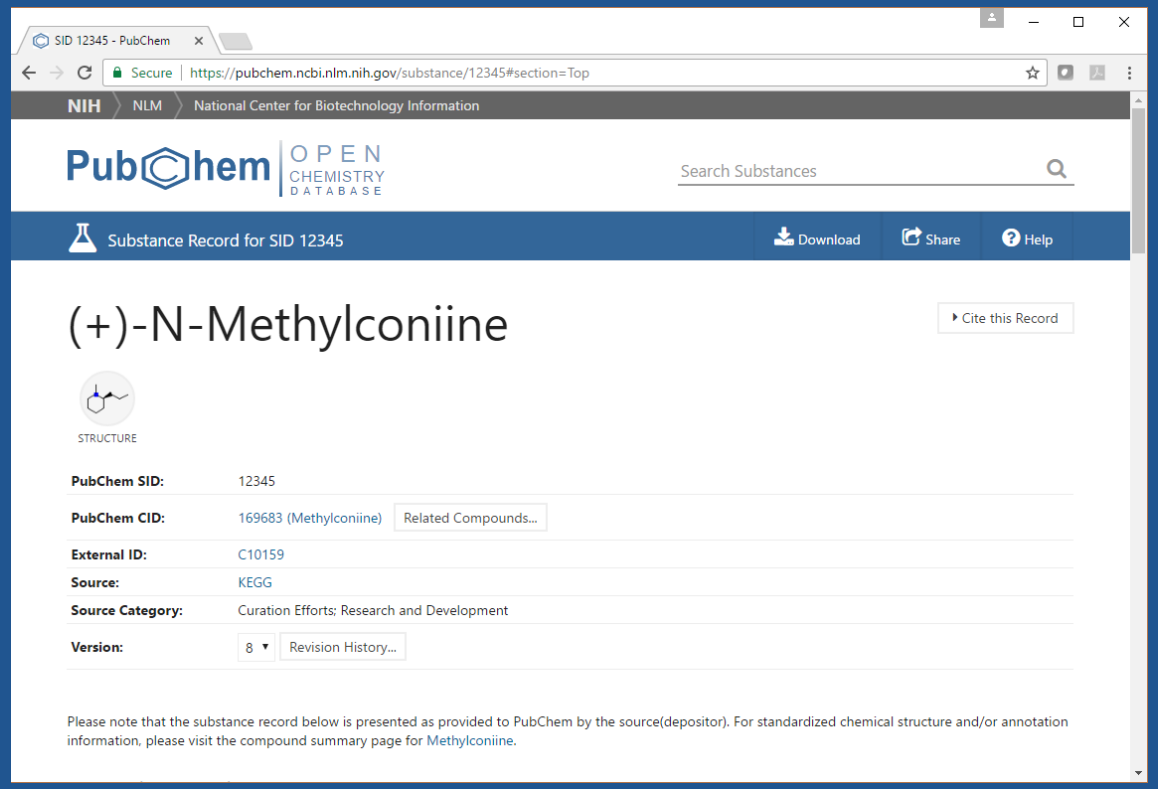
- Data Sources page gives a summary of all data sources

Technology revamp continues



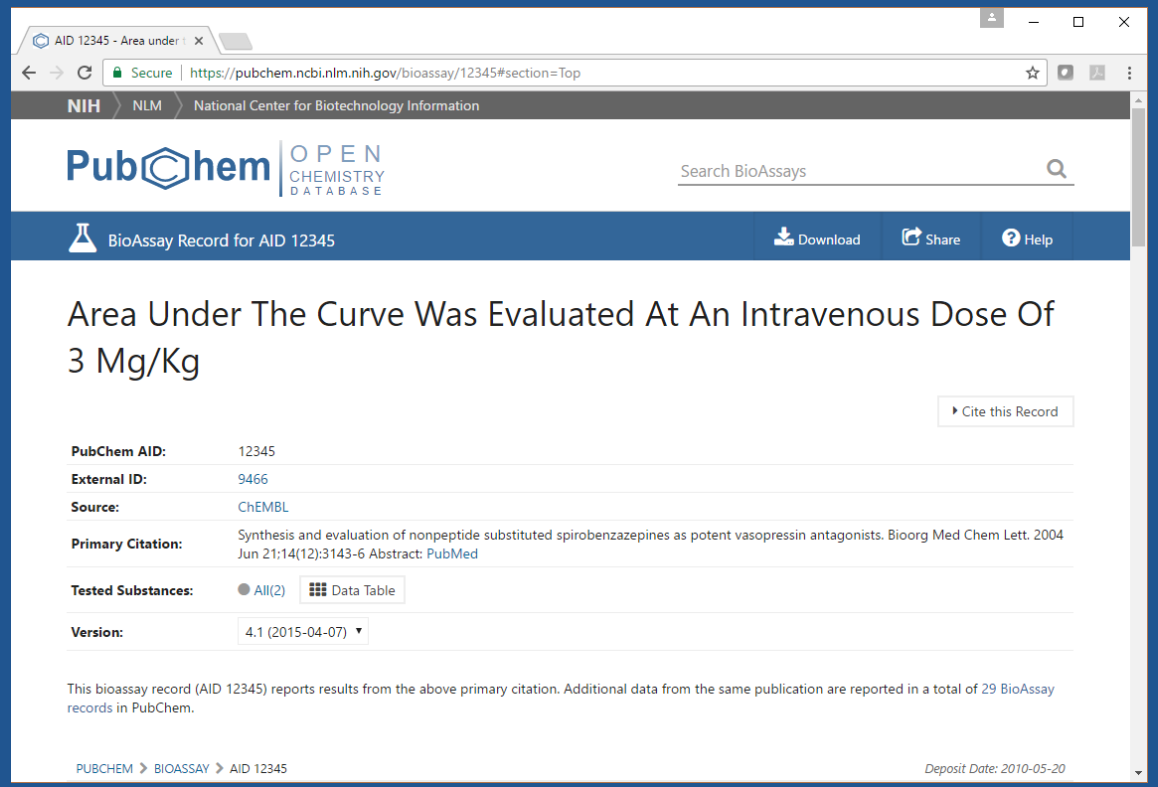
- Data Sources page gives a summary of all data sources
- One can use faceting to subset by various means
- Uses same/similar approach as Compound Summary page (JSON data delivery followed by rendering in browser .. display is data driven)

Technology revamp continues



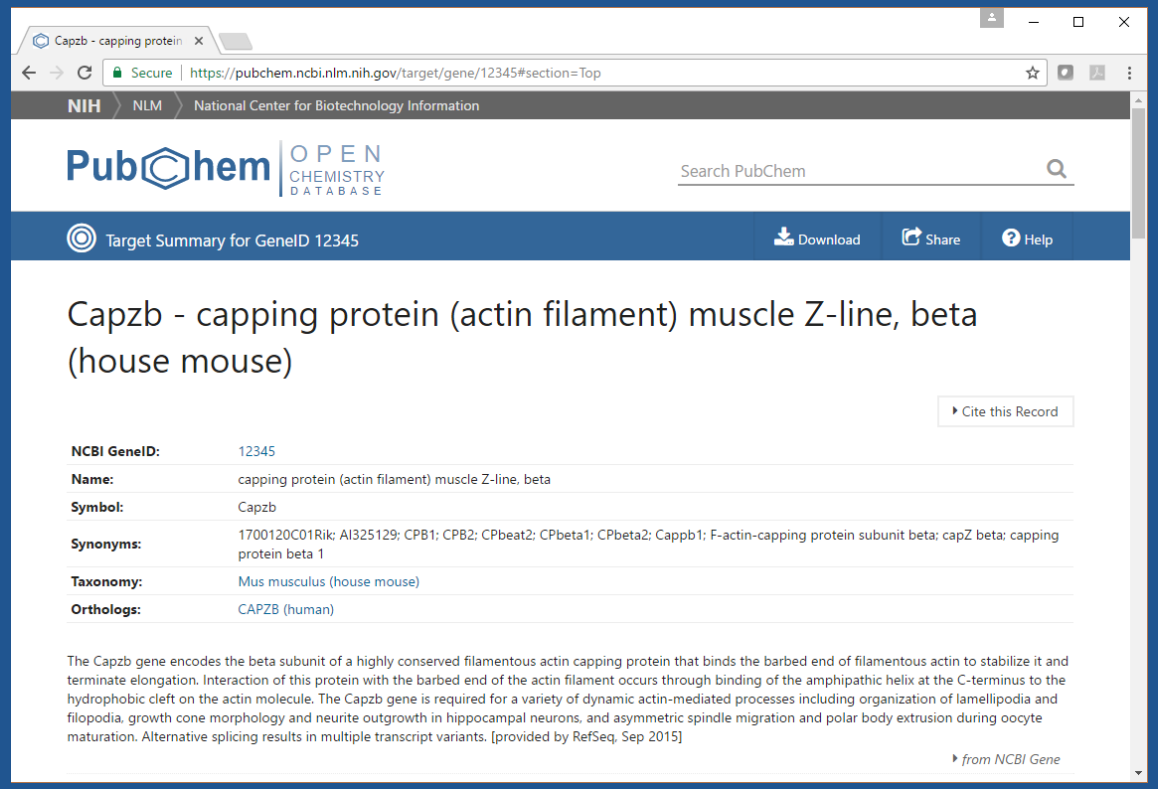
- Substance Record page uses same/similar approach as Compound Summary page

Technology revamp continues



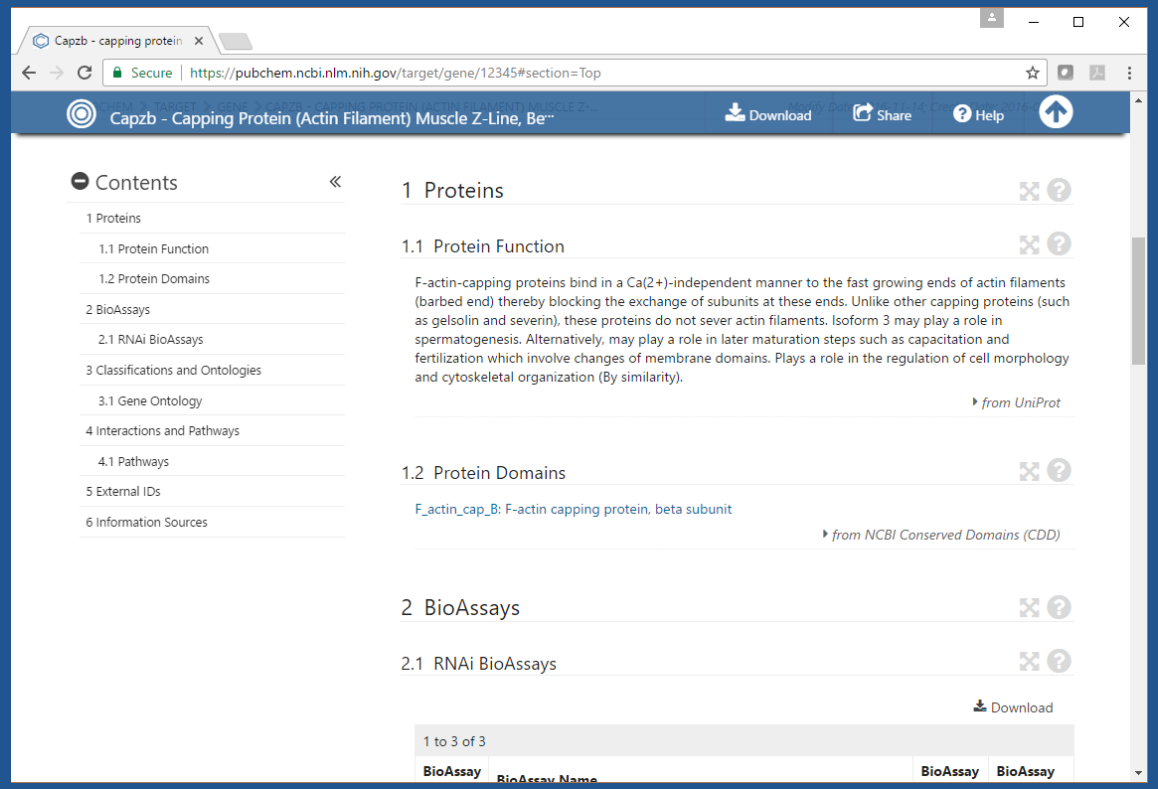
- BioAssay Record page uses same/similar approach as Compound Summary page

Technology revamp continues



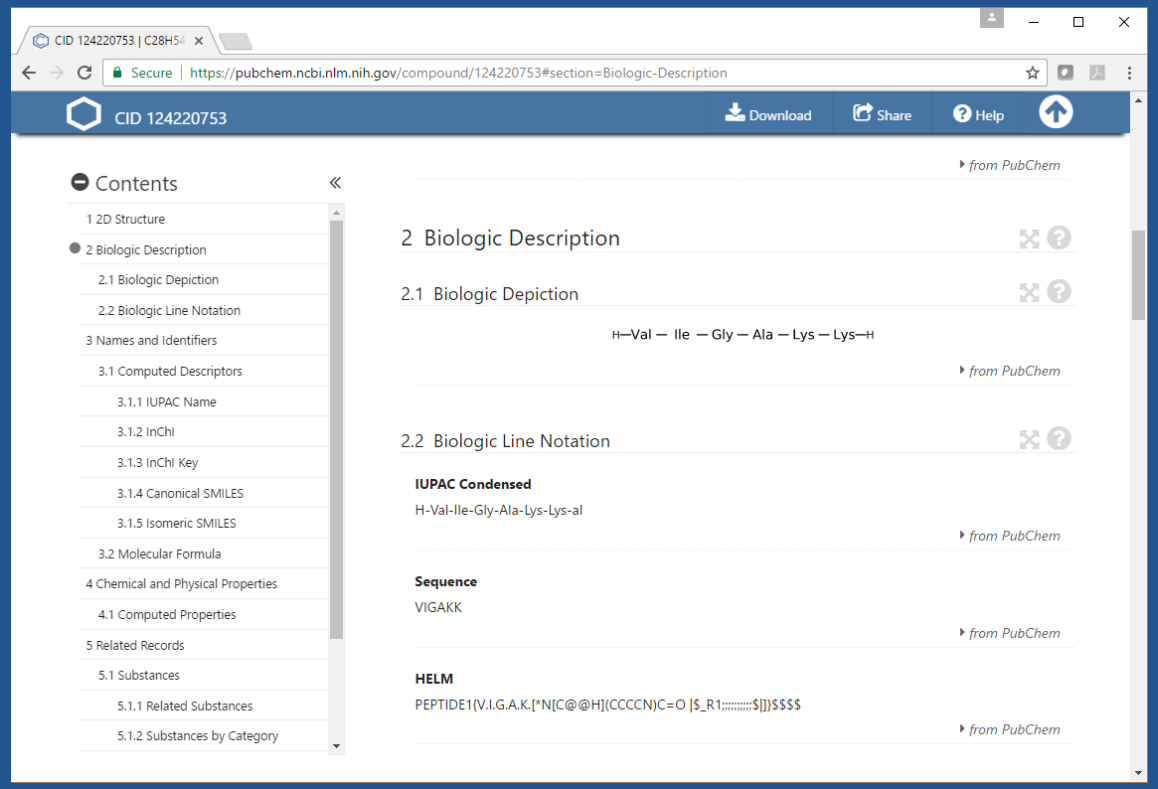
- Gene Target Summary page uses same/similar approach as Compound Summary page

Technology revamp continues



- Gene Target Summary page uses same/similar approach as Compound Summary page
- Recently announced
- Summarizes chemical and bioactivity content using a Gene Target concept for grouping purposes
- Protein Target page to be released
- Working towards other grouping paradigms

Technology revamp continues



- While delayed by more than a year (for various reasons), progress continues in earnest on Biologics
- Currently a separate section, with naming and descriptor strings
- Annotating amino-acid and saccharides that can be completely described (0.5M out of 10M)

Towards full support of “Biologics”

- Coverage of amino acids, nucleic acids, saccharides, and lipids
  - Standards when mixing or modifying .. naming and line notations need cleanup
    - Lipids absent .. but LipidMaps/ChEBI already did a bit of work here
    - Better line notation approaches needed when chemical modifications are made
  - Many, many ‘biopolymer’ containing structures are in PubChem (10M?)
- Going beyond 1000 atoms/bonds [← current structure limit in PubChem]
  - Mastering what is in PubChem first
    - Can we: define what is a biologic? define PTMs? handle small molecule → biologic?
  - Working towards a super-atom approach .. [e.g., Ala = one super atom]
  - Helping lay the ground work with standards …

Symbol Nomenclature for (Graphical Representations of) Glycans (SNFG)

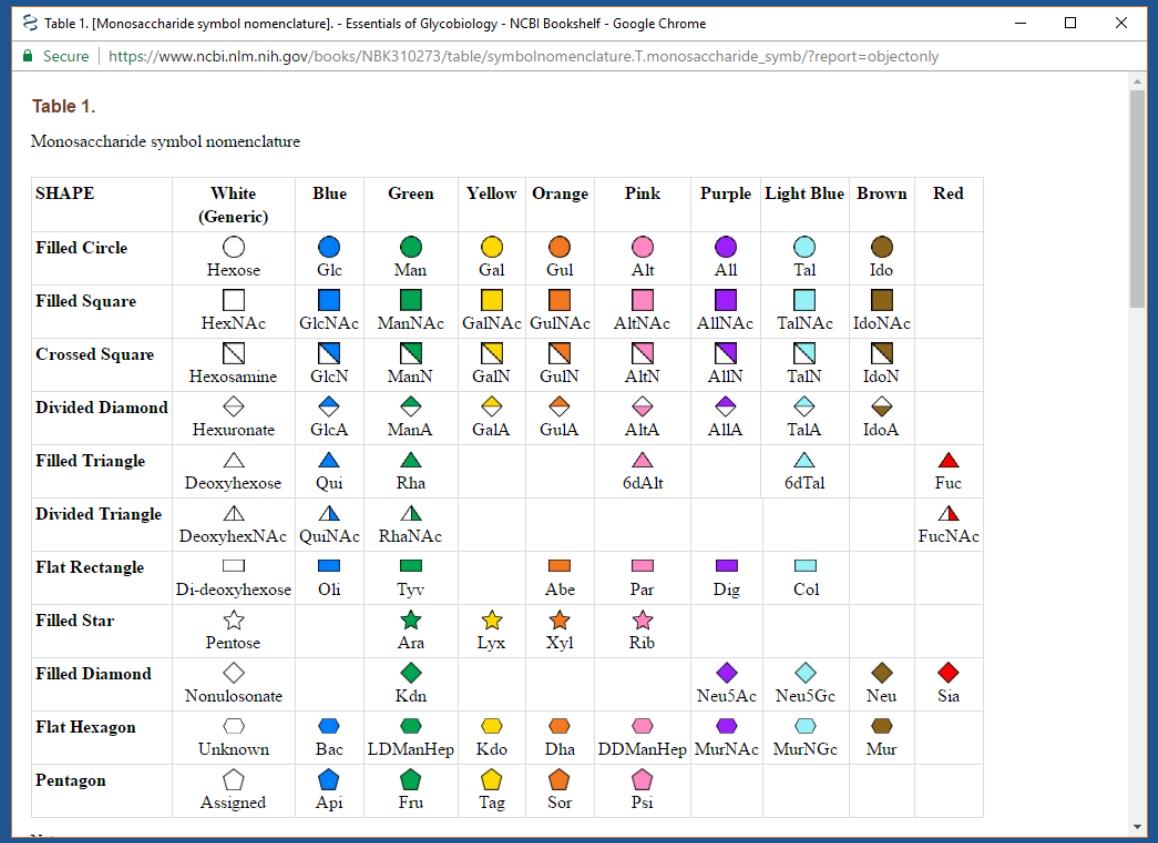


Table 1. [Monosaccharide symbol nomenclature]. - Essentials of Glycobiology - NCBI Bookshelf - Google Chrome

Secure | https://www.ncbi.nlm.nih.gov/books/NBK310273/table/symbolnomenclature.T.monosaccharide_symb/?report=objectonly

Table 1.

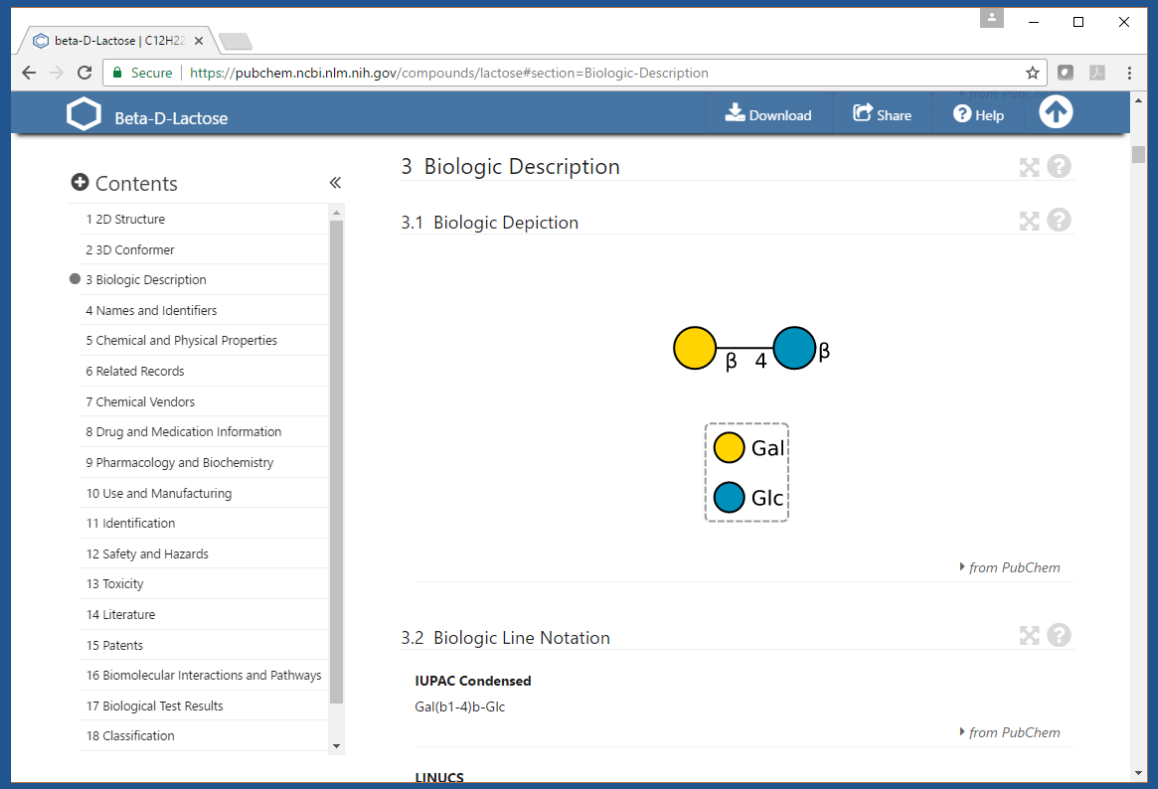
Monosaccharide symbol nomenclature

SHAPE	White (Generic)	Blue	Green	Yellow	Orange	Pink	Purple	Light Blue	Brown	Red
Filled Circle	Hexose	Glc	Man	Gal	Gul	Alt	All	Tal	Ido	
Filled Square	HexNAc	GlcNAc	ManNAc	GalNAc	GulNAc	AltNAc	AllNAc	TalNAc	IdoNAc	
Crossed Square	Hexosamine	GlcN	ManN	GalN	GulN	AltN	AllN	TalN	IdoN	
Divided Diamond	Hexuronate	GlcA	ManA	GalA	GulA	AltA	AllA	TalA	IdoA	
Filled Triangle	Deoxyhexose	Qui	Rha			6dAlt		6dTal		Fuc
Divided Triangle	DeoxyhexNAc	QuiNAc	RhaNAc							FucNAc
Flat Rectangle	Di-deoxyhexose	Oli	Tyv		Abe	Par	Dig	Col		
Filled Star	Pentose		Ara	Lyx	Xyl	Rib				
Filled Diamond	Nonulosonate		Kdn				Neu5Ac	Neu5Gc	Neu	Sia
Flat Hexagon	Unknown	Bac	LDManHep	Kdo	Dha	DDManHep	MurNAc	MurNGc	Mur	
Pentagon	Assigned	Api	Fru	Tag	Sor	Psi				

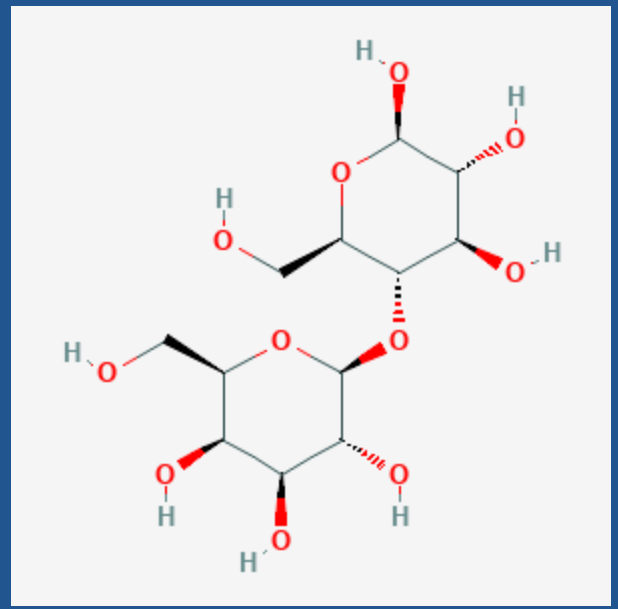
- Glycans use symbolic approach
- How to combine these with small molecule changes?
- How to combine these with other biopolymers? (glycoproteins, glycolipids, etc.)
- Helping to organize the glycobiology community (@ Biohackathon)
- PubChem helped launch the Glycans website at NCBI

https://www.ncbi.nlm.nih.gov/glycans/

Lactose

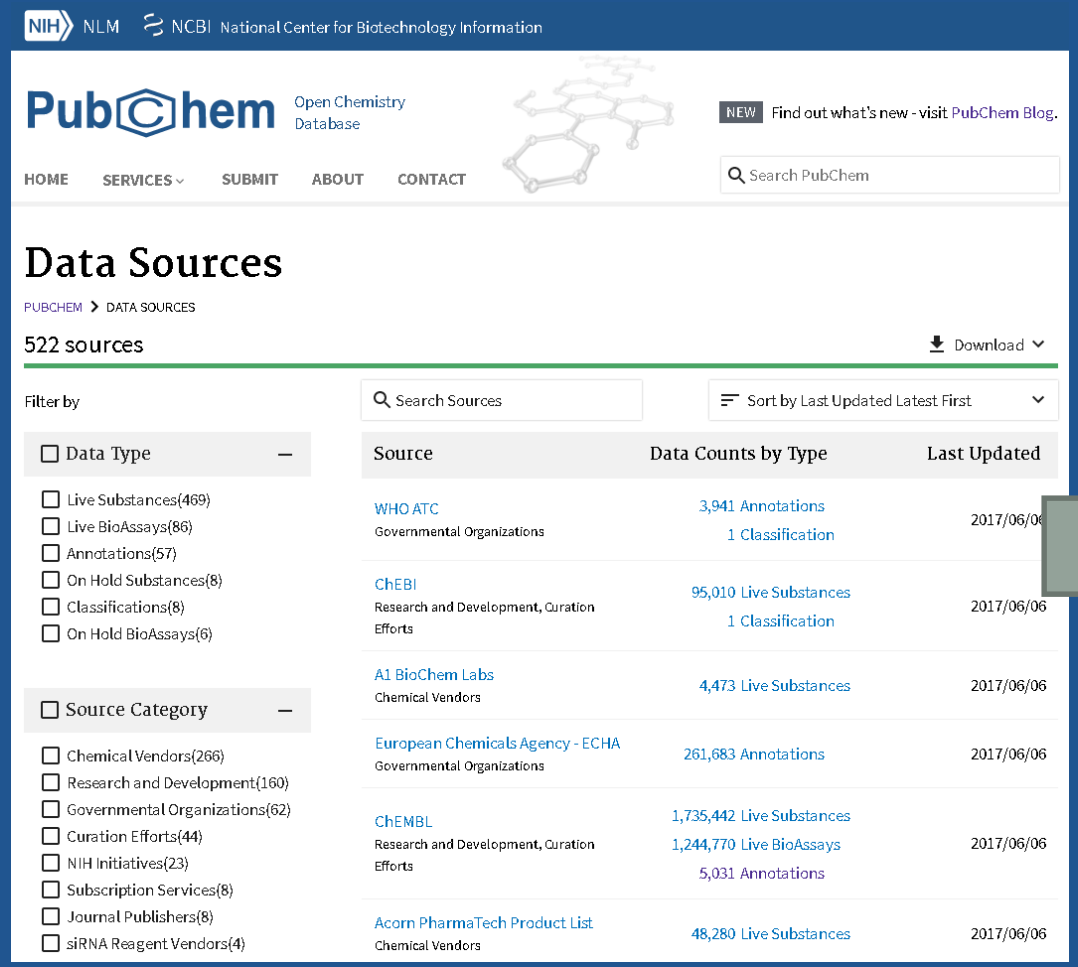


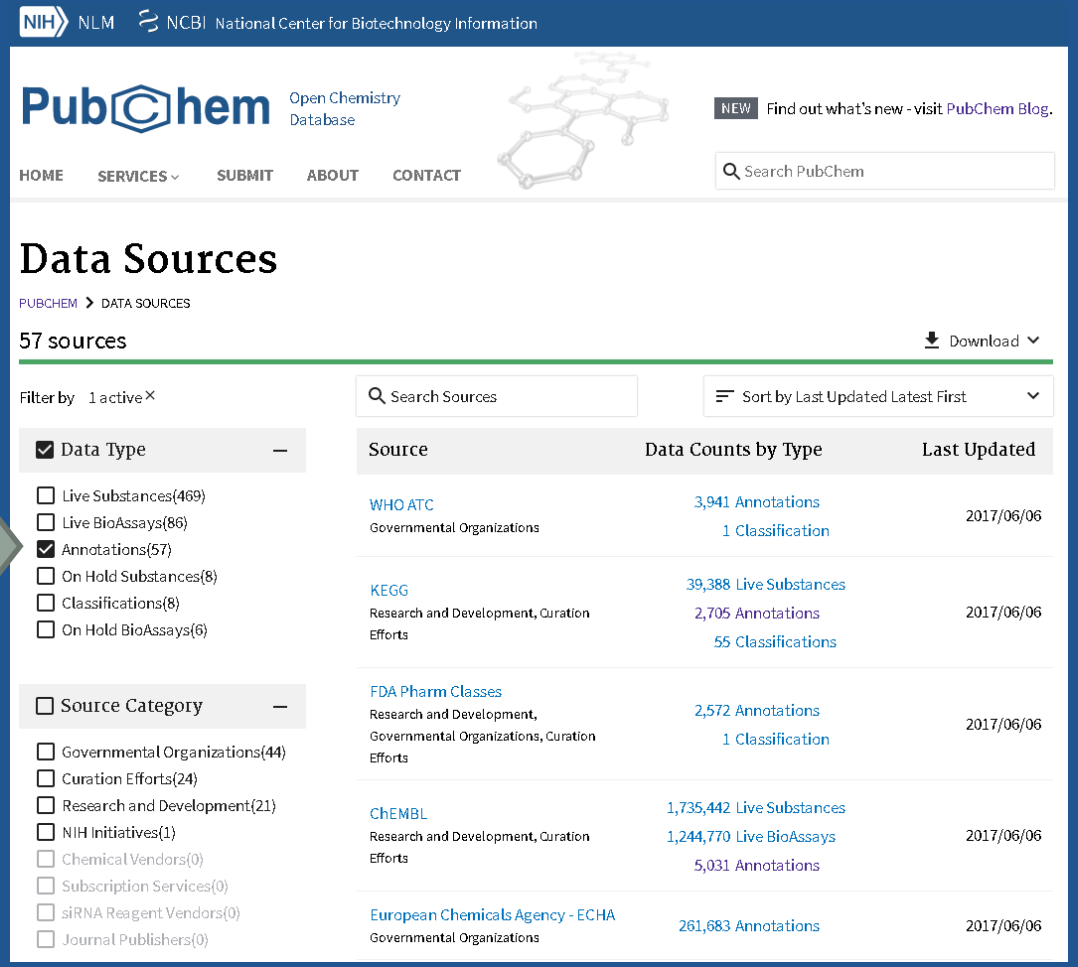
- Symbolic representation in action .. makes for pretty pictures that are way easier to interpret .. if you know what you are looking at …



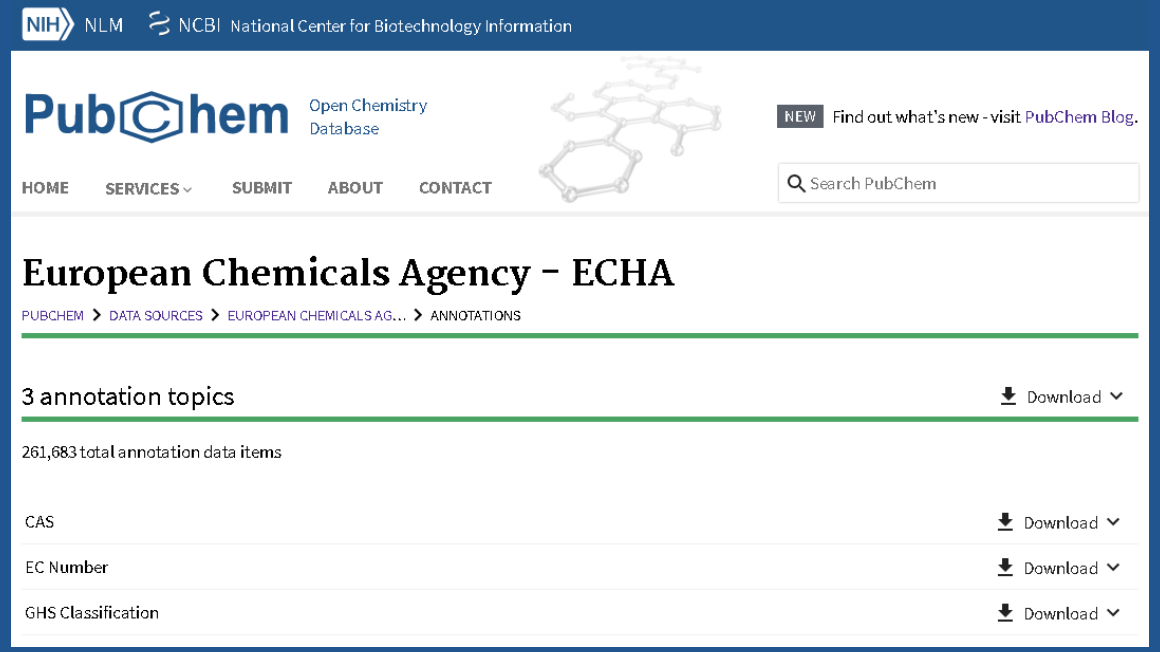
PubChem Data Sources

https://pubchem.ncbi.nlm.nih.gov/sources/



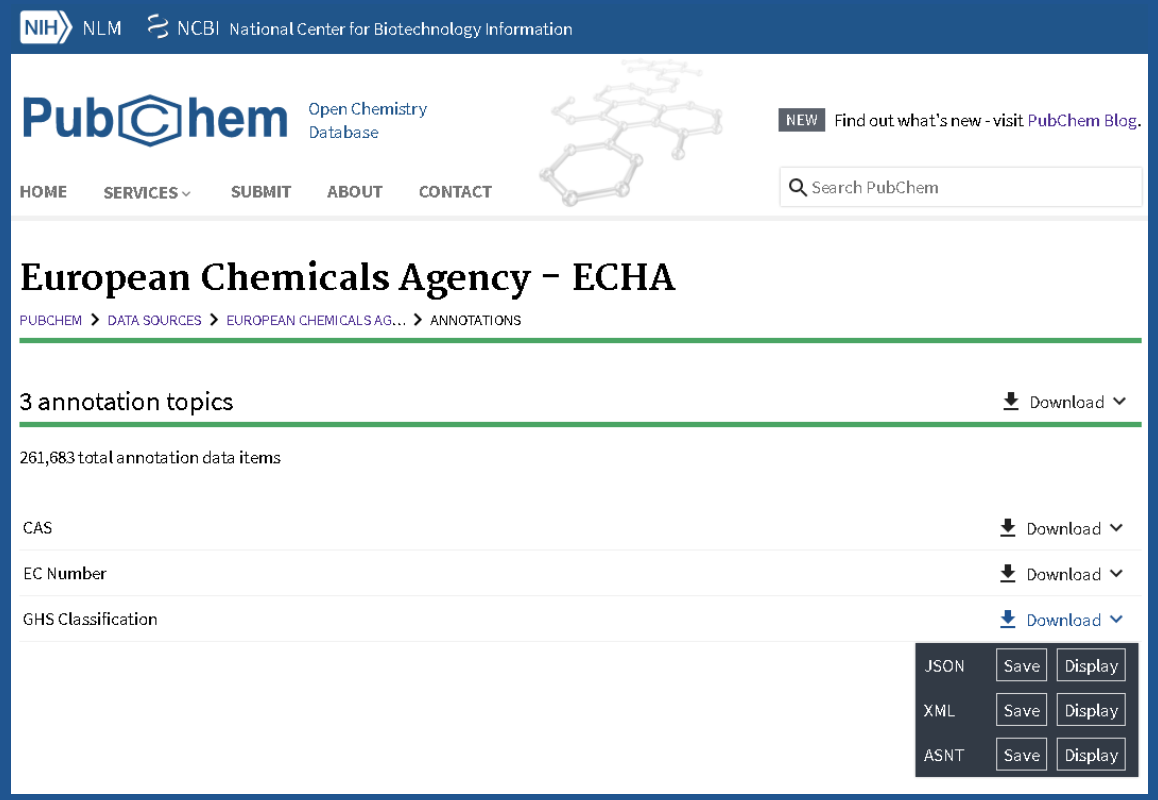


PubChem Data Sources – what data comes from whom?



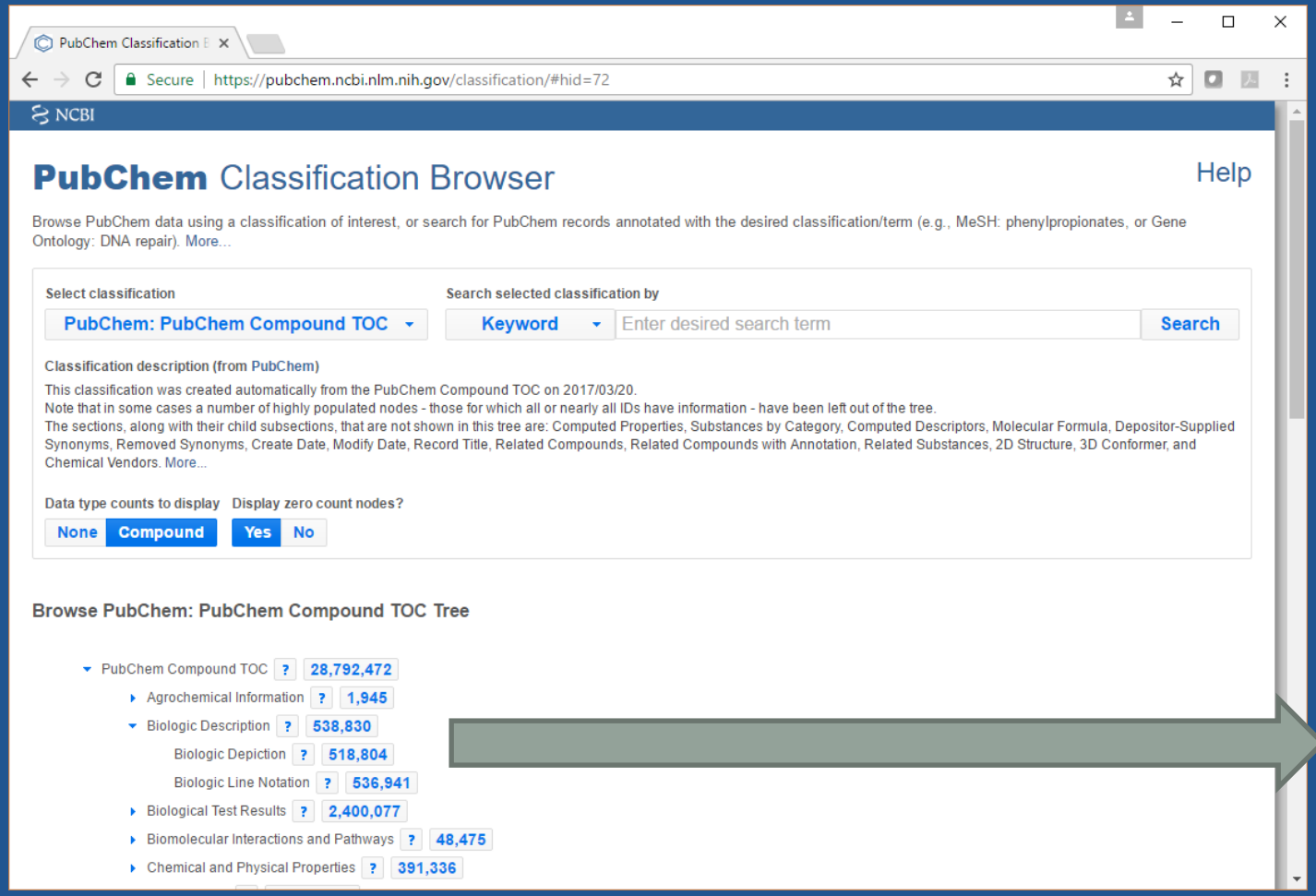
https://pubchem.ncbi.nlm.nih.gov/source/ECHA

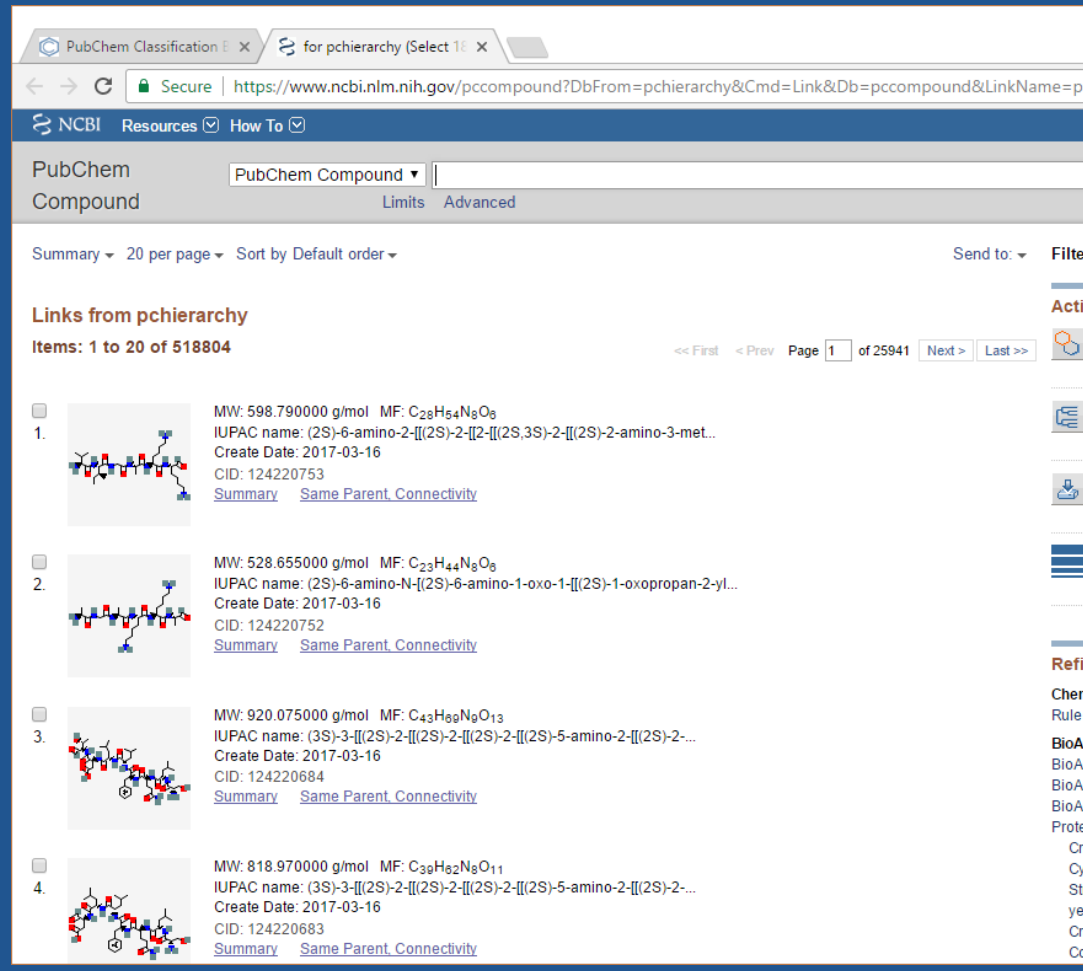
PubChem Data Sources – data downloadable




```
{
  "Annotations": {
    "Annotation": [
      {
        "SourceName": "EU REGULATION (EC) No 1272/2008",
        "SourceID": "001-001-00-9",
        "Name": "hydrogen",
        "Description": "Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on
classification, labelling and packaging of substances and mixtures. Ref: OJ L 353, 31.12.2008, p. 1-1355 Special edition in Croatian:
Chapter 13 Volume 020 P. 3-1357",
        "URL": "http://ec.europa.eu/growth/sectors/chemicals/classification-labelling/index_en.htm",
        "LinkToPubChemBy": {
          "CID": [
            783
          ]
        },
        "Data": [
          {
            "TOCHeading": "GHS Classification",
            "Name": "GHS Classification",
            "Value": {
              "String": [
                "<div class=\"pc-thumbnail-container\"><img src=\"/images/ghs/GHS02.svg\" title=\"GHS02: Flammables\"></div><b>Signal:
</b> <span class=\"fred\">Danger</span><br><div class=\"ghs-hazards\"><b>GHS Hazard Statements</b><br>H220: Extremely flammable gas
[<span class=\"fred\">Danger</span> Flammable gases - Category 1]<br></div><br><div class=\"ghs-precautionary\"><b>Precautionary
Statement Codes</b><br>P210, P377, P381, and P403<br>(The corresponding statement to each P-code can be found <a
href=\"https://pubchem.ncbi.nlm.nih.gov/ghs/#_prec\">here</a>.)</div>"
              ]
            }
          }
        ]
      },
      {
        "SourceName": "EU REGULATION (EC) No 1272/2008",
        "SourceID": "001-002-00-4",
        "Name": "aluminium lithium hydride",
        "Description": "Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on
classification, labelling and packaging of substances and mixtures. Ref: OJ L 353, 31.12.2008, p. 1-1355 Special edition in Croatian:
Chapter 13 Volume 020 P. 3-1357",
        "URL": "http://ec.europa.eu/growth/sectors/chemicals/classification-labelling/index_en.htm",
        "LinkToPubChemBy": {
          "CID": [
            21226445
          ]
        },
        "Data": [
          {
            "TOCHeading": "GHS Classification",
            "Name": "GHS Classification",
            "Value": {
              "String": [
                "<div class=\"pc-thumbnail-container\"><img src=\"/images/ghs/GHS02.svg\" title=\"GHS02: Flammables\"><img
src=\"/images/ghs/GHS05.svg\" title=\"GHS05: Corrosives\"></div><b>Signal:</b> <span class=\"fred\">Danger</span><br><div class=\"ghs-
hazards\"><b>GHS Hazard Statements</b><br>H260: In contact with <a class=\"pubchem-internal-link CID-962\"
href=\"https://pubchem.ncbi.nlm.nih.gov/compound/water\">water</a> releases flammable gases which may ignite spontaneously [<span
class=\"fred\">Danger</span> Substances And Mixtures Which, In Contact With <a class=\"pubchem-internal-link CID-962\"
href=\"https://pubchem.ncbi.nlm.nih.gov/compound/Water\">Water</a>, Emit Flammable Gases - Category 1]<br>H314: Causes severe skin burns
```

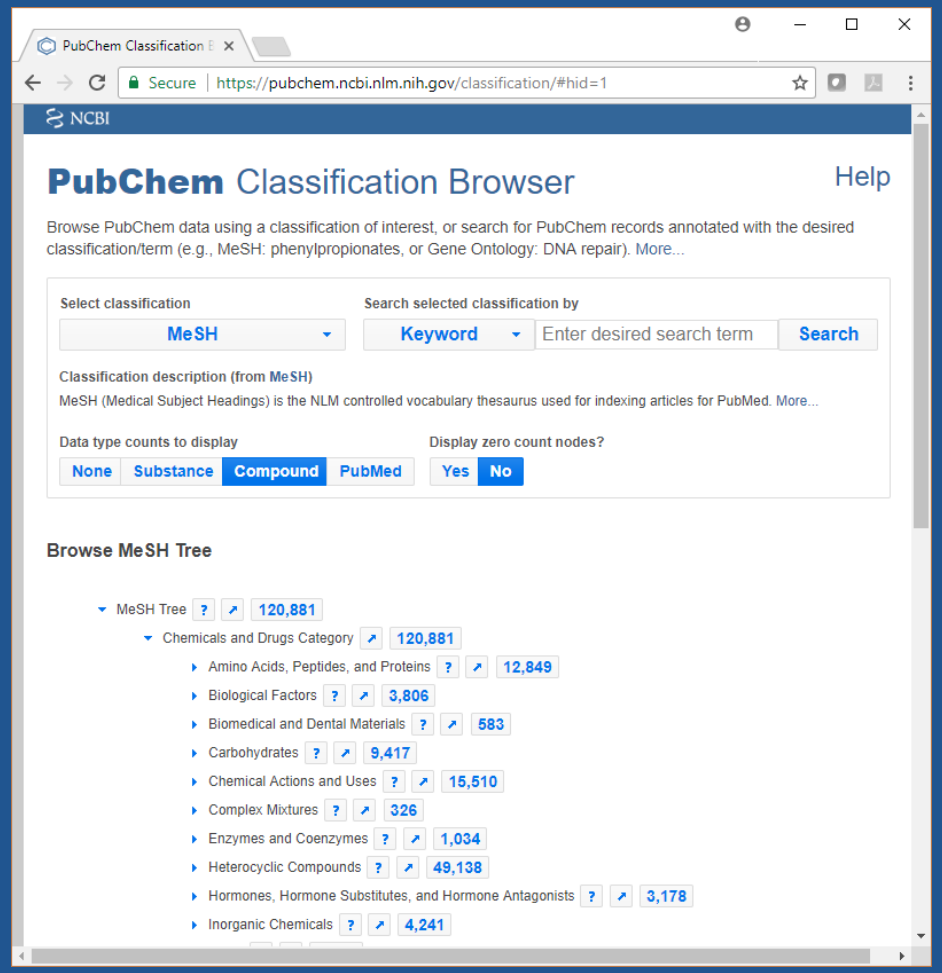
Classification Browser enables finding annotations

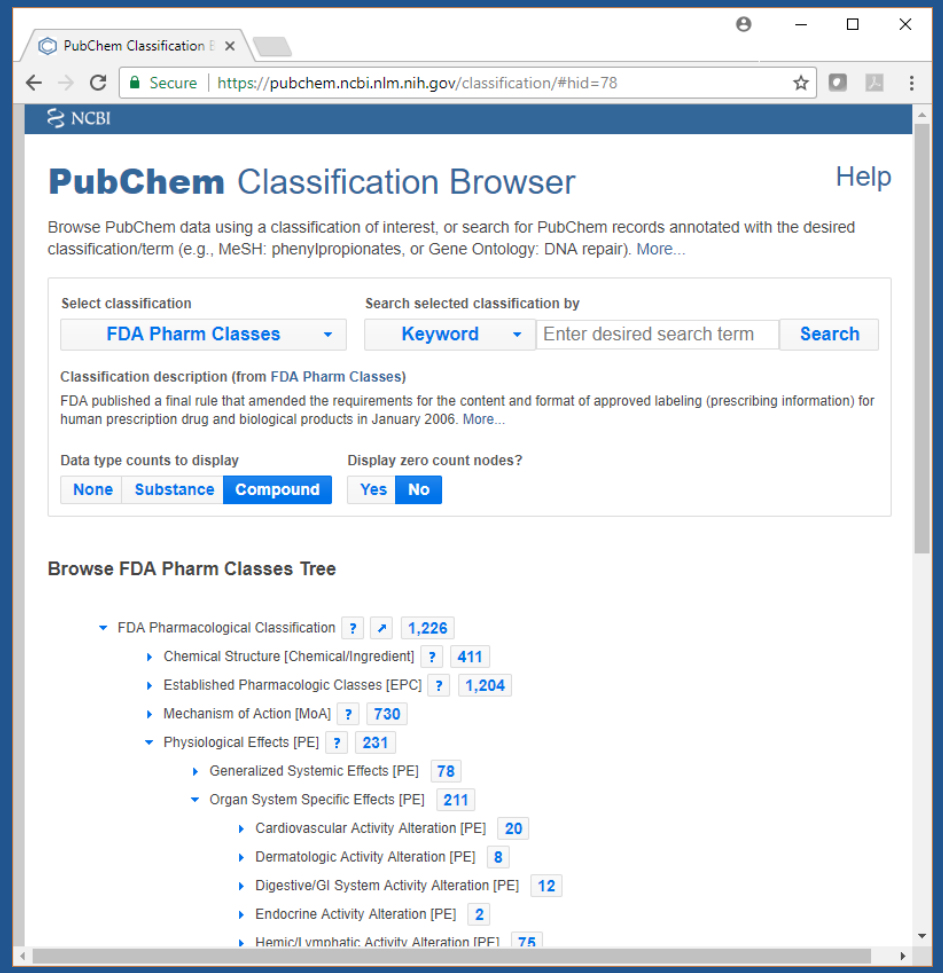




Give me all “Biologics” found in PubChem

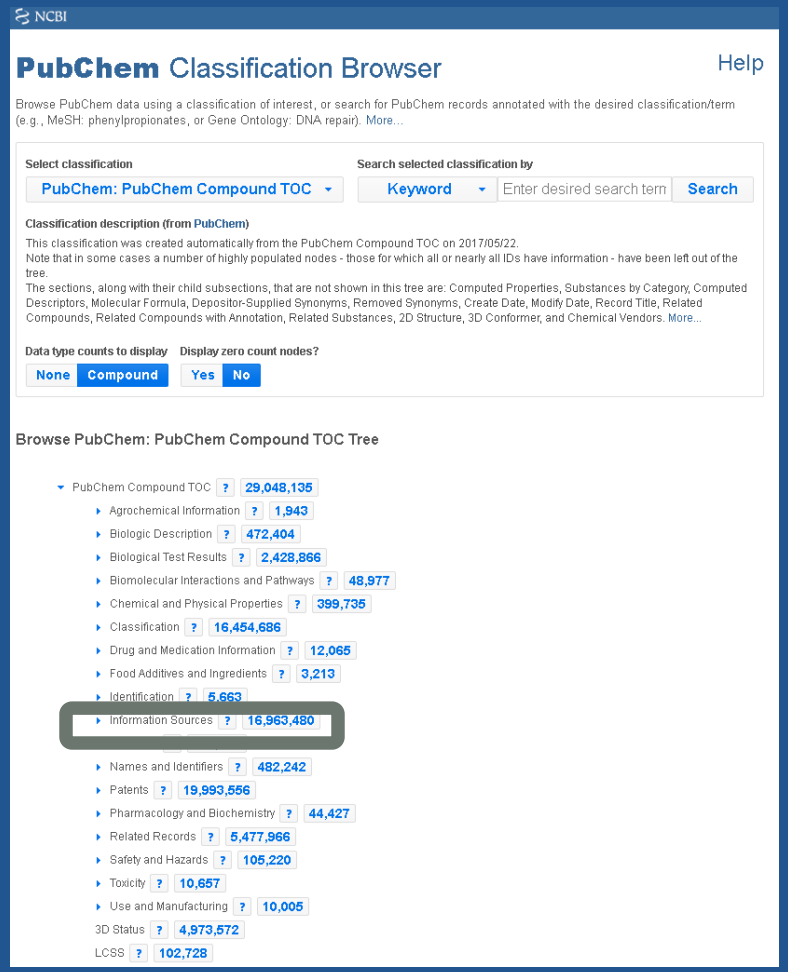
Classification Trees are a form of hierarchical annotation

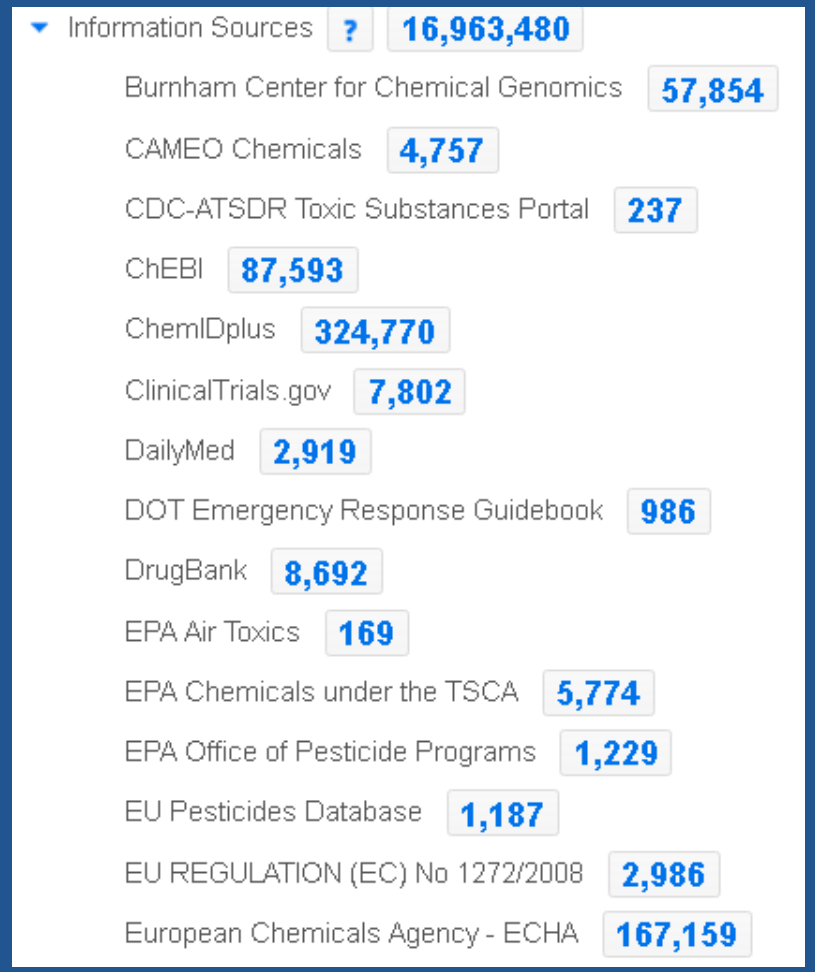




Classification Browser – PubChem Compound TOC

https://pubchem.ncbi.nlm.nih.gov/classification/#hid=72

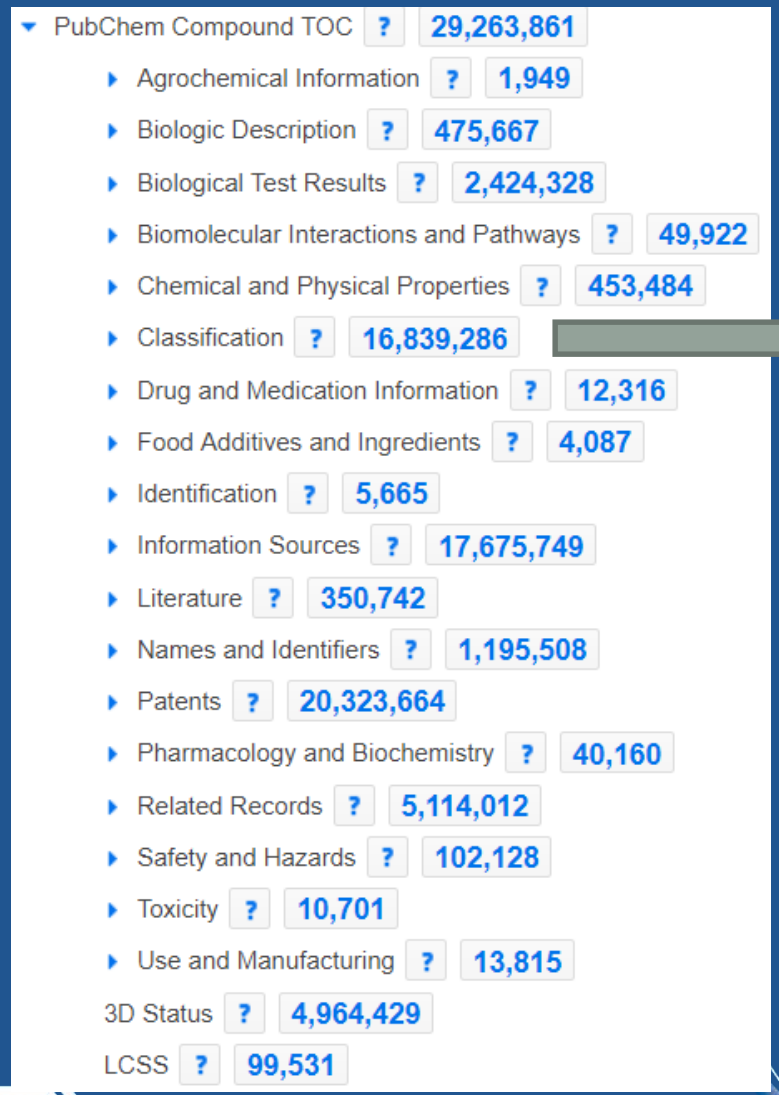


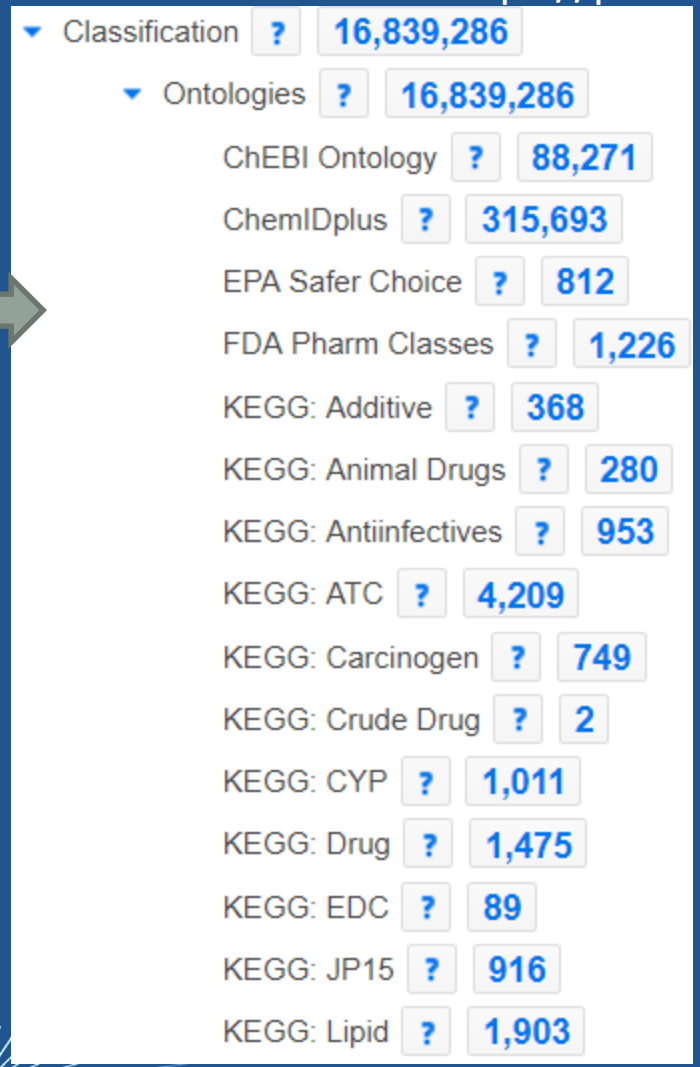


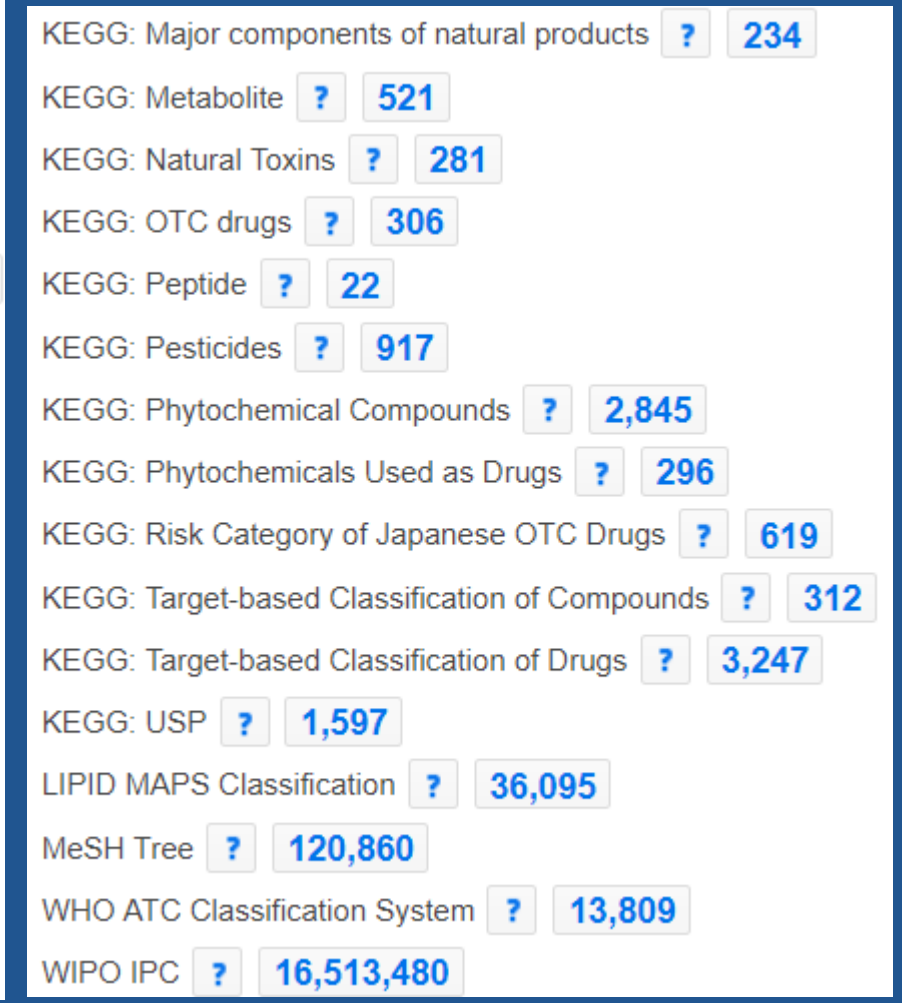
And many more….

Classifications include the Compound Table of Contents (TOC)

https://pubchem.ncbi.nlm.nih.gov/classification/#hid=72







Many annotation information sources

https://pubchem.ncbi.nlm.nih.gov/classification/#hid=72

- Information Sources ? 17,675,749
  - Burnham Center for Chemical Genomics 57,854
  - CAMEO Chemicals 4,746
  - CDC-ATSDR Toxic Substances Portal 180
  - ChEBI 88,271
  - ChemIDplus 360,516
  - ClinicalTrials.gov 7,854
  - DailyMed 3,036
  - DOT Emergency Response Guidebook 992
  - DrugBank 8,704
  - DTP/NCI 260,248
  - EPA Air Toxics 170
  - EPA Chemicals under the TSCA 5,775
  - EPA DSStox 681,956
  - EPA Office of Pesticide Programs 1,229
  - EPA Safer Choice 812
  - EU Food Improvement Agents 3,254
  - EU Pesticides Database 1,188
  - EU REGULATION (EC) No 1272/2008 2,820
  - European Chemicals Agency - ECHA 163,363
  - FAO/WHO Food Additive Evaluations - JECFA 2,504
  - FDA Medication Guides 601
  - FDA Orange Book 4,467
  - FDA Pharm Classes 1,227
  - FDA/SPL Indexing Data 61,895
  - Flavor & Extract Manufacturers Association - FEMA 2,781
  - HSDB 8,256
  - Human Metabolome Database (HMDB) 20,705
  - ILO-ICSC 1,720
  - KEGG 11,876
  - LIPID MAPS 36,095
  - LiverTox 1,591
  - MeSH 120,931
  - NCBI 11,685
  - NCI Investigational Drugs 212
  - NCIt 5,907
  - NIOSH Manual of Analytical Methods 572
  - NIST 262,518
  - NITE-CMC 2,875
  - NJDOH RTK Hazardous Substance List 2,289
  - OSHA Chemical Sampling Information 1,479
  - OSHA Occupational Chemical DB 842
  - PubMed Health 2,737
  - Safe Work Australia - HCIS 3,399
  - SpectraBase 84,376
  - The Cambridge Structural Database 60,727
  - The National Institute for Occupational Safety and Health - NIOSH 2,258
  - UN Globally Harmonized System of Classification and Labelling of Chemicals (GHS) 101,562
  - USDA Pesticide Data Program 530
  - USGS Columbia Environmental Research Center 380
  - WHO ATC 13,812
  - Wikipedia 168,048
  - WIPO 16,513,480

Many chemicals, many properties (Overall +500 types of information)

- Experimental Properties ? 160,573
  - Acid Value ? 37
  - Auto-Ignition ? 990
  - Boiling Point ? 3,803
  - Caco2 Permeability ? 120
  - Color ? 6,807
  - Corrosivity ? 766
  - Decomposition ? 4,217
  - Density ? 4,219
  - Dissociation Constants ? 2,136
  - Flash Point ? 2,138
  - Heat of Combustion ? 572
  - Heat of Vaporization ? 719
  - Hydrophobicity ? 21
  - Ionization Potential ? 323
  - Isoelectric Point ? 17
  - Kovats Retention Index ? 79,656
  - LogP ? 8,040
  - LogS ? 245
  - Melting Point ? 11,676
  - Odor ? 3,060
  - Odor Threshold ? 468
  - Optical Rotation ? 184
  - pH ? 1,062
  - Physical Description ? 19,848
  - pKa ? 547
  - Polymerization ? 184
  - Refractive Index ? 11
  - Relative Evaporation Rate ? 120
  - Solubility ? 68,276
  - Stability ? 3,540
  - Surface Tension ? 584
  - Taste ? 1,064
  - Vapor Density ? 1,366
  - Vapor Pressure ? 4,877
  - Viscosity ? 741

- Literature ? 350,742
  - Depositor Provided PubMed Citations ? 260,864
  - General References ? 2,220
  - Metabolite References ? 12,430
  - NLM Curated PubMed Citations ? 116,087
  - Synthesis References ? 4,088

- Spectral Properties ? 309,210
  - C13-NMR ? 39,636
  - GC-MS ? 234,509
  - H1-NMR ? 46,946
  - Infrared Spectra ? 65,238
  - Mass ? 953
  - MS-MS ? 16,370
  - Raman ? 10,831
  - UV ? 14,032
  - 1D NMR ? 1,632
  - 2D NMR ? 1,024
  - EI-MS ? 1,007
  - GC ? 2
  - HPLC ? 202

Much more to do… and lots in the pipeline

- Dramatic quality improvements
  - Classification of chemical names (e.g., to hide/remove 'bad' names)
  - Cross-validation of billions/trillions of links (e.g., PubTator, LeadMine)
- Construct chemical concept map (e.g., to capture all chemical concepts in use)
- Improve integration with patents (e.g., USPTO, EPO, WIPO, JPO, KPO, CPO)
- Reinvent PubChem search (e.g., handle natural language questions)
- Consider more chemical biology use cases (e.g., spectra, reactions, expand annotations)
- And many, many more things we should be doing… especially with text

PubChem Crew …

Evan Bolton

Jie Chen

Tiejun Cheng

Gang Fu

Asta Gindulyte

Jane He

Siqian He

Sunghwan Kim

Ben Shoemaker

Paul Thiessen

Bo Yu

Leonid Zaslavsky

Jian Zhang

Special thanks to the NCBI Help Desk, especially Rana Morris, and past PubChem group members.

Special thanks

- PubChemRDF Collaborators
- Chemical Health and Safety collaborators
  - Especially: Leah McEwen (Cornell U.), Ralph Stuart (Keene State College)
- Software collaborators
  - NextMove Software (Roger Sayle, Noel O'Boyle, John May) and Daniel Lowe
  - Xemistry GmbH (Wolf D. Ihlenfeldt)
  - OpenEye Scientific Software
- All PubChem Contributors and Collaborators

- PubChem is supported by the Intramural Research Program of the NIH, National Library of Medicine.

Have any questions?