Open PHACTS
Open Pharmacological Space

Open PHACTS Data

With Acknowledgment to the Open PHACTS project consortium

Open PHACTS
Open Pharmacological Space

- Data sources presently in Open PHACTS selected by:
- Choosing data sources to answer business questions
- Selecting appropriate data sources based on licensing
- Filtering criteria of judged data quality
- **Chemistry Data**: ChEBI, ChEMBL, Drugbank aggregated through ChemSpider – initial focus on pharmacology
- **Biology Data**: SwissProt, GO/GOA and ENZYME
- Data underpins public version plus all related exemplars
- Data are intended to be updated quarterly for the near future
- Curation/annotation flow intended to be continuous

Data in Open PHACTS

Open PHACTS
Open Pharmacological Space

The process:

- Chemicals imported to ChemSpider and registered with CSIDs
- Registered compounds passed to ConceptWiki
- ConceptWiki passes data to Identifier Mapping Service
- RDF generated from ChemSpider for deposition to cache
- Linked Data Cache receives RDF from various original data sources – RDF is provided by data source provider, or is sourced from alternate provider, or is generated by Open PHACTS team

The Process

Open PHACTS
Open Pharmacological Space

RDF is well suited to describing complex data

- Open and supported by a growing body of tools and scientists
- Critical data sources are published as RDF
- Producers of RDF can enhance their RDF with information required to create Nanopublications
- We will support data providers to generate RDF as necessary

RDF

Open PHACTS
Open Pharmacological Space

- We will publish guidelines for producing RDF. These are community-endorsed best practices. For instance:
  - "Good URIs" (stable, opaque, dereferencable)
  - Attention to quantitative data
  - The use of semantic models for the dataset
- RDF will be used to produce and integrate nanopublications

- How many of your organizations already have skills and experience in generating RDF?

Making the RDF

Open PHACTS
Open Pharmacological Space

- Data licensing is hard! Licenses are commonly obscure…

- Open PHACTS hopes to accommodate both **open** and **license-restricted** data
- We will go live in September with data sets that have well-defined licensing
- We will develop a licensing strategy around public data to ensure OPS complies with the terms for use and distribution
- We will support restricted datasets within the platform in the future - licensed or private content.
- We are contracting expertise in data licensing to navigate…

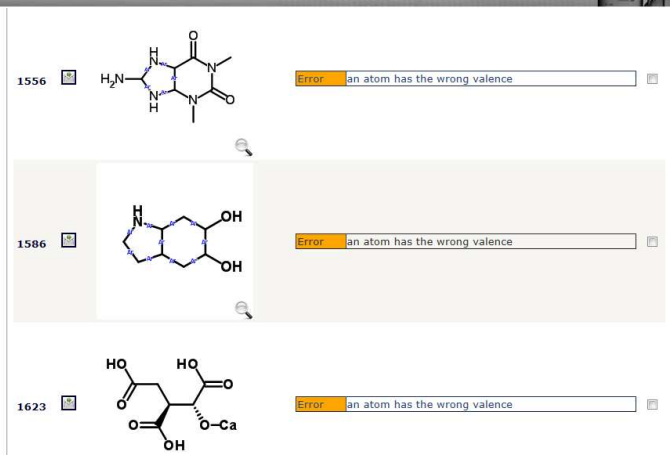
Data Licencing

Open PHACTS
Open Pharmacological Space

- ChemSpider registers compounds and provides services to perform structure/substructure/similarity searches
- Each data source providing chemistry is “registered”
  - This process assigns them a unique CSID
  - CSIDs, preferred name, synonyms, Inchi string & key and SMILES are published back into the cache
  - Chemical structure standardization rules are in development using FDA SRS standards. Validation included
  - ChemSpider provides validated synonyms to ConceptWiki for text-based searching/look-up

Chemistry Representation

Open PHACTS
Open Pharmacological Space



Curation and Annotation

Open PHACTS
Open Pharmacological Space

- Develop **systems** to validate and standardize data – our preference is to be proactive in data quality when possible
- Provide interactive ability to curate and annotate data from data sources
- Existing systems to be integrated and enhanced – ConceptWiki and ChemSpider **already** provides curation/annotation features
- Provide feedback to data suppliers regarding identified issues and help clean up existing data to contribute back to providers

- Our primary concern is participation – not the systems.

Curation and Annotation

Open PHACTS
Open Pharmacological Space

- Additional data sources will be identified based on project needs, data licensing, partner provisions

- What do you see as possibilities for integrating **YOUR** data?

- How might you want to use Open PHACTS data/technologies (recall that there will be Open APIs into the system)?

Future