

Chem³ to search SAR space... and much more

Jean-Marc Neefs

Computational Chemistry, Janssen Beerse

We need Chem³

- To *Innovate* in small molecules, we must also know all that are already 'out there'
- We can not waste time *searching* for data
- *Basic (SAR) searching* must be easy, fast, comprehensive, relevant, and produce up to date results for further analysis

We need a 'Pharmacology Google'

Challenges

- Searching data beyond proprietary information requires some level of integration
- Integration must happen in hardware, software, processes, mindsets
- One option is to build something entirely new
- Another option is to integrate into existing software familiar to end user chemists and biologists

Opportunities

- **Open PHACTS** is a innovative integration platform
- 3DX is the workhorse software for chemists and biologists
 - data visualisation and analysis, pharmacology, supported by our IT organisation
- Pipeline Pilot (BioVia) is the workhorse software computational chemists
 - Computational workflows - cfr Knime, also supported by our IT organisation
 - Remember Edgar's presentation

Chem³ is here

In 3DX and in Pipeline Pilot

The screenshot displays the 3DX software interface. The menu bar includes File, Edit, Select, View, Data, Chemistry, Query, Request, Reports, Sequence, Abcd2, PDS, Tools, and Help. The 'Abcd2' menu is open, showing 'Database' and 'Research' options. The 'Database' submenu is expanded, listing 'EZ SPARQL', 'EMolecules', 'Abcd2 (Target, Compound)', and 'ABCD and ChEMBL Compound Act...'. A large white arrow points from the 'EZ SPARQL' menu item to the data table below. Another large white arrow points from the 'Database' menu item to the 'EZ SPARQL' menu item.

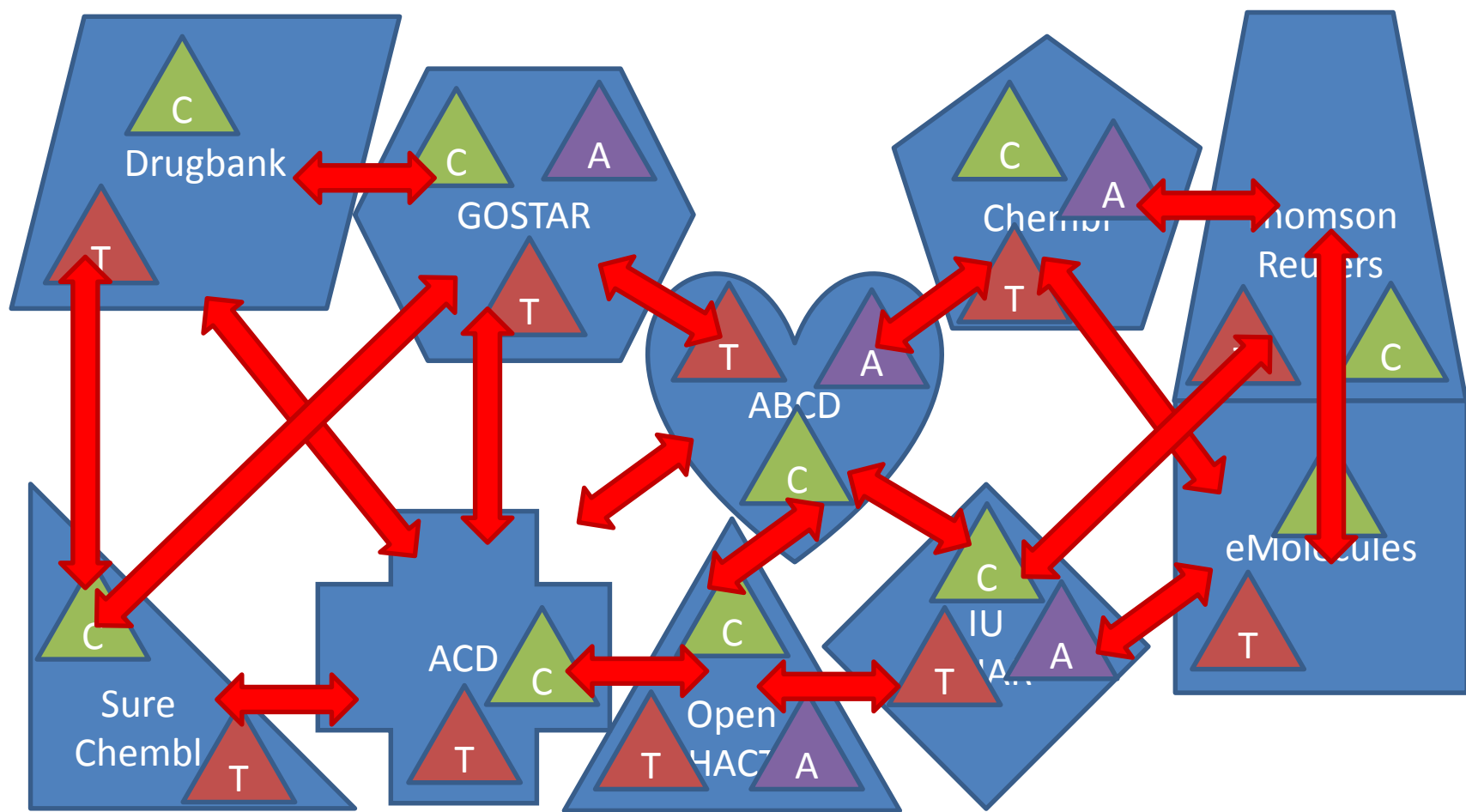
	Concept
465	http://purl.uniprot.org/core/Alternative Promoter Usage Annotation
466	http://purl.uniprot.org/core/Unknown Sequence
467	http://purl.uniprot.org/core/Pharmaceutical Annotation
468	http://purl.uniprot.org/core/Alternative Initiation Annotation
469	http://purl.uniprot.org/core/Toxic Dose Annotation
470	http://purl.uniprot.org/core/Ribosomal Frame shifting
471	http://biohackathon.org/resource/faldo#OneOfPosition
472	http://purl.uniprot.org/core/Redox Potential Annotation

At the bottom right of the interface, the text 't.eu.jnj.com 9.2.0' is visible.

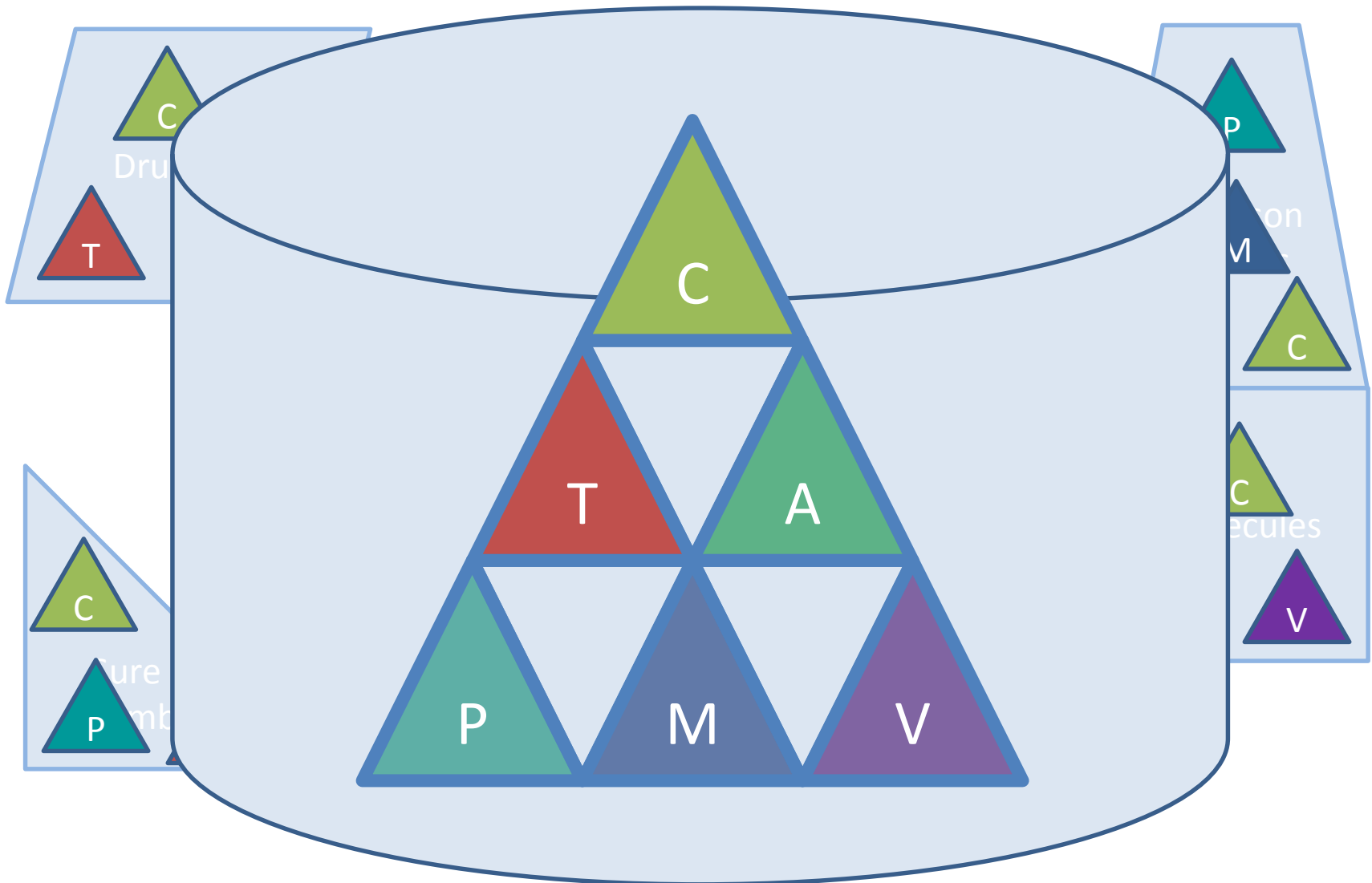
Chem³ *searches for you*

- **Current Queries** in the prototype:
 - With a target (list), *search* active small molecules
 - With a molecule (list), *search* structures and activities
 - *Searches* several SAR dbs *together* : ABCD + ChEMBL
 - *New this week*: free text search
- in **3DX** and **Pipeline Pilot**
- It is **programmable** – it will be able to *search* extra information not in our own database

Relational Searching is Hard



Semantics makes Integration and Searching a lot easier



Some Semantic 'Rules' Examples

- Examples for SAR:
 - *Assays* measure *Activity* of *Compounds* on *Targets*
 - *Activities* have *Values* and *Units* (optional)
 - *Compounds* have *SMILES* codes
- Examples for Much More related Data:
 - *Genes* code for *Targets*, *Genes* have *Expression Levels*
 - *Targets* have *Mutations*, *Mutations* cause *Diseases*
 - *Drugs* affect *Targets*, may cause *Toxic Effects*
 - *Target* levels affect *Cells*, *Cells* have *imaging data*
- ***Possibility to Identify indirect relations***

Relational vs Semantic Data Integration

Relational – ‘ORACLE’

- Table input
 - Data in rows, Properties in columns
- Structure of the data **is** pre-defined for each data source
- Relations between **tables** using keys + ‘mapping’ tables for all relations between **tables**
- Database Schema - based
- Query: SQL

Semantic – ‘OpenLink Virtuoso’

- ‘Triple’ input in RDF format
 - Subject – Verb – Object, Data AND Properties in triples
- Structure of the data **is not** pre-defined for each data source
- Relations between **entities** using more ‘mapping’ triples for all relations between **entities**
- Ontology - based
- Query: SPARQL

Many Thanks

- Open PHACTS community
- Herman van Vlijmen
- Edgar Jacoby
- Dmitrii Rassokhin
- Doug Martin