Chem^3 to search SAR space… and much more

Jean-Marc Neefs

Computational Chemistry, Janssen Beerse

We need Chem^3

- 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

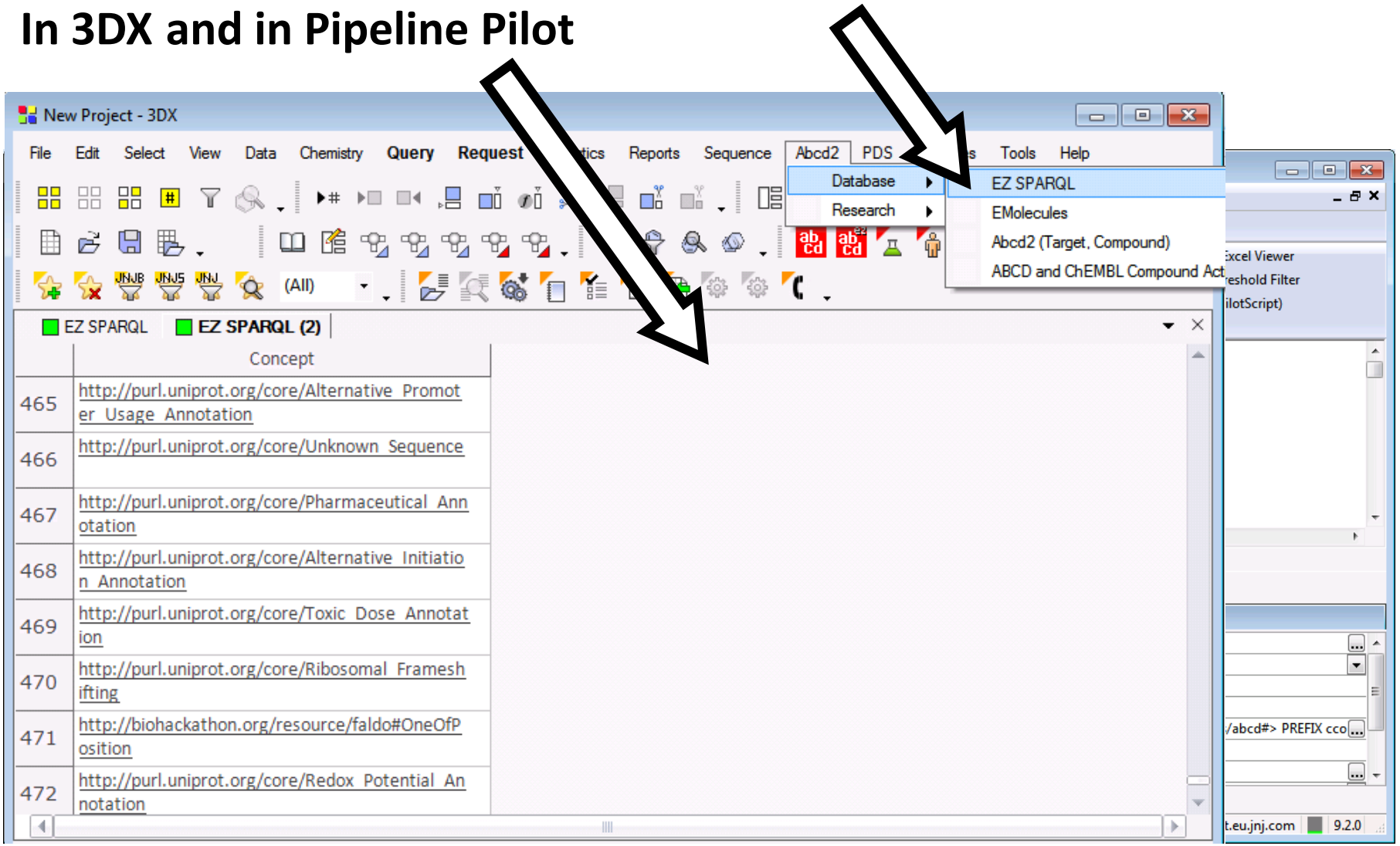
- Searching data beyond proprietary information requires some level of integration
- Integration must happen in hardware, software, processes, mindsets
- One option is to build something entirely new
- Another option is to integrate into existing software familiar to end user chemists and biologists

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

Chem^3 is here

In 3DX and in Pipeline Pilot



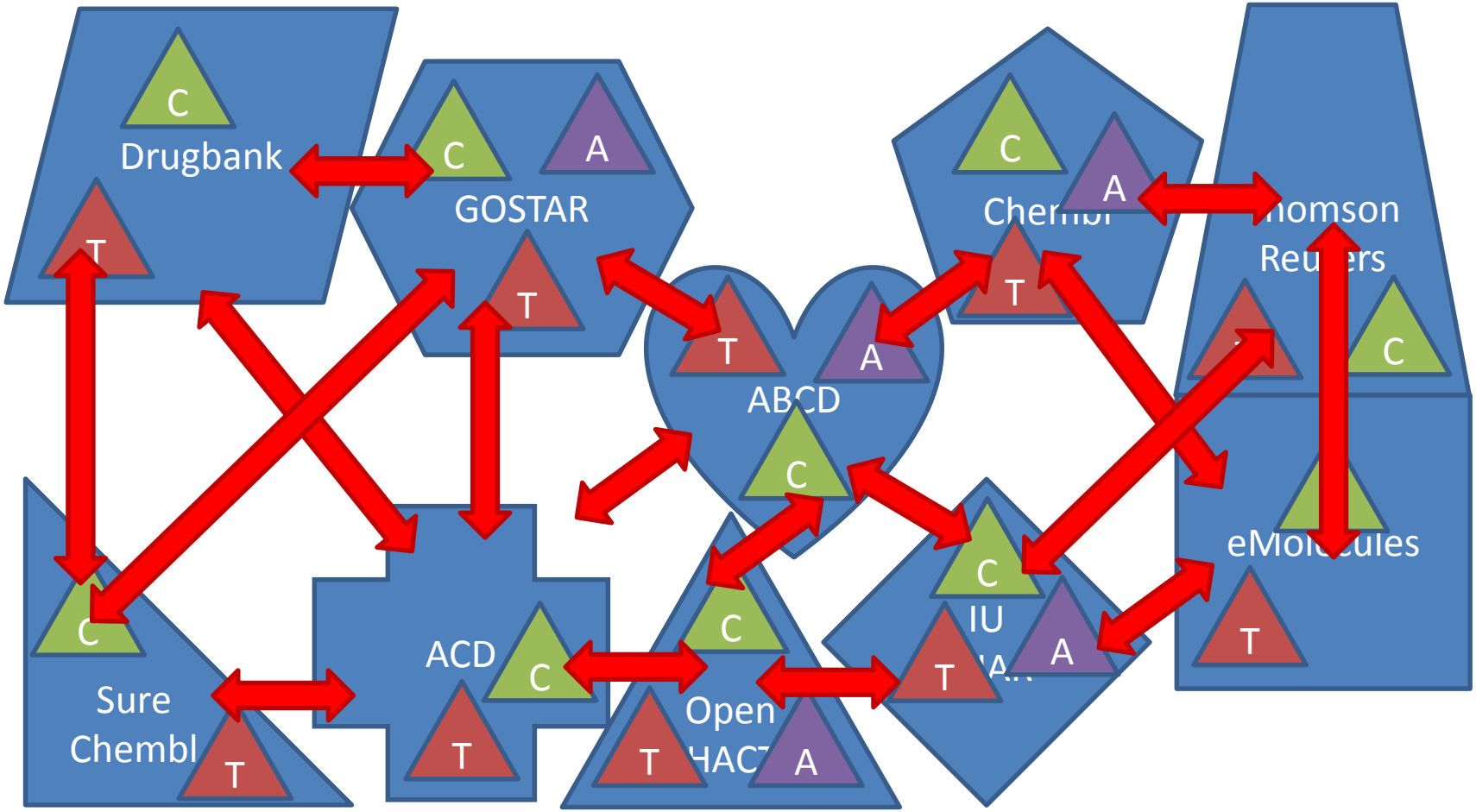
Chem^3 *searches* for *you*

- **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

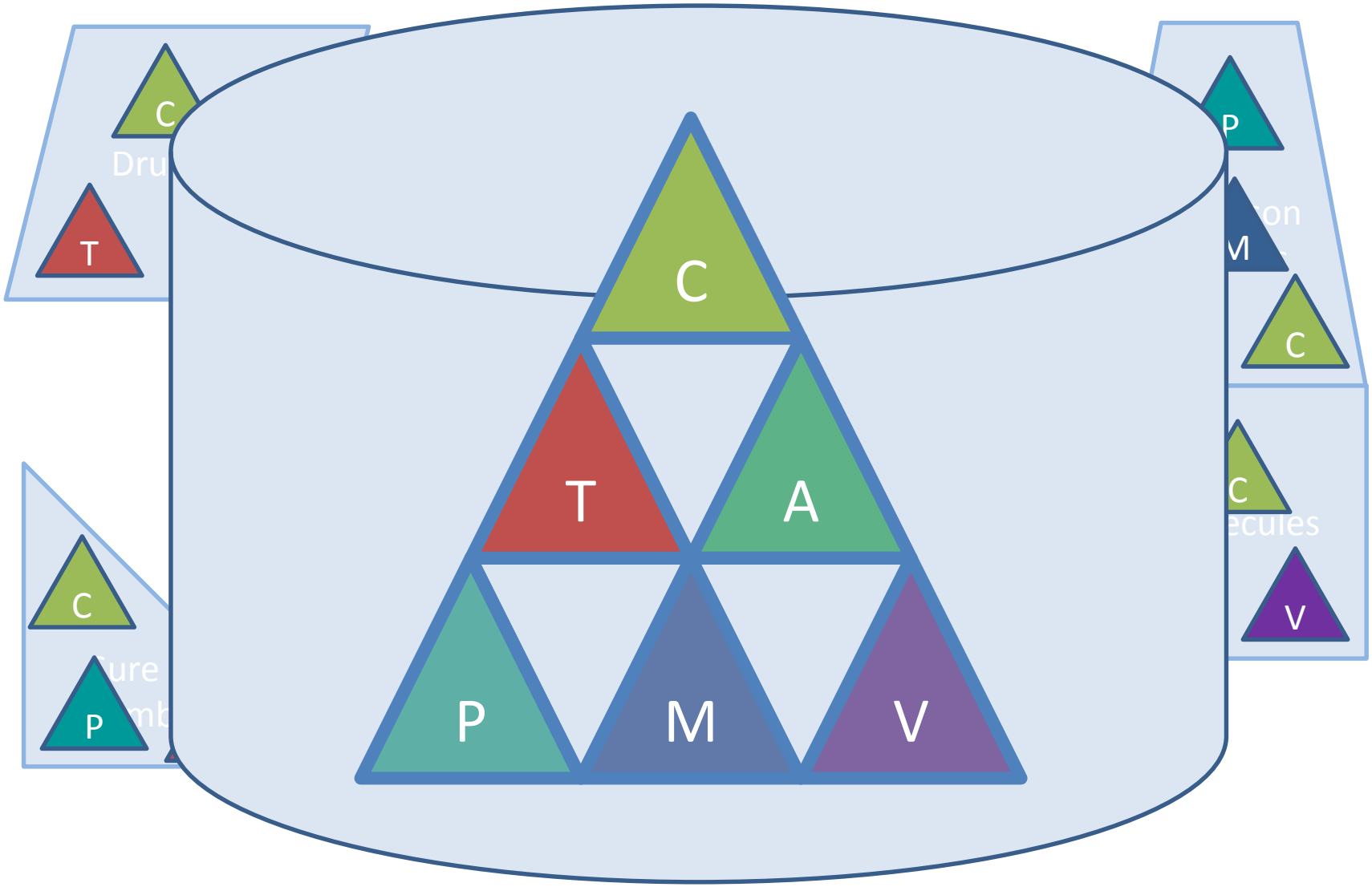
- in **3DX** and **Pipeline Pilot**

- 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:
  - G*enes* 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