



Instant JChem

Chemistry on your desktop

Tim Dudgeon <tdudgeon@chemaxon.com>



IJC team

IJC Core

- Petr Hamernik - everything!
- Petr Zajac - schema editor, JWS
- Martin Adamek – IJC server
- Max Sauer - query, DB
- Daniel Butler - docs + QA
- Martin Krauskopf - visualisation
- Radim Kubacki – just started!

Also

- James Illston
- Tim Miller, Steve Hajkowski

Marvin/JChem

- Csizi
- Szilard Dorant
- Gyorgy Pirok
- Peter Kovacs
- Szabolcs Csepregi
- Miklos Vargyas
- Tamas Vertse
- Istvan Rabel
- ... and many many many many many many many many more

- What is IJC
- What's new since last year
- What's coming in 5.4
- IJC databases - case studies

What is IJC?

*A desktop tool for managing,
searching and analysing chemical
structures and associated information*

Who is IJC for?

*Any biologist or chemist.
Guru status not required*



Key features

- Powerful report designer
- Database management
- Comprehensive search capabilities
- Structure based predictions
- Sharing and collaboration

New since last year

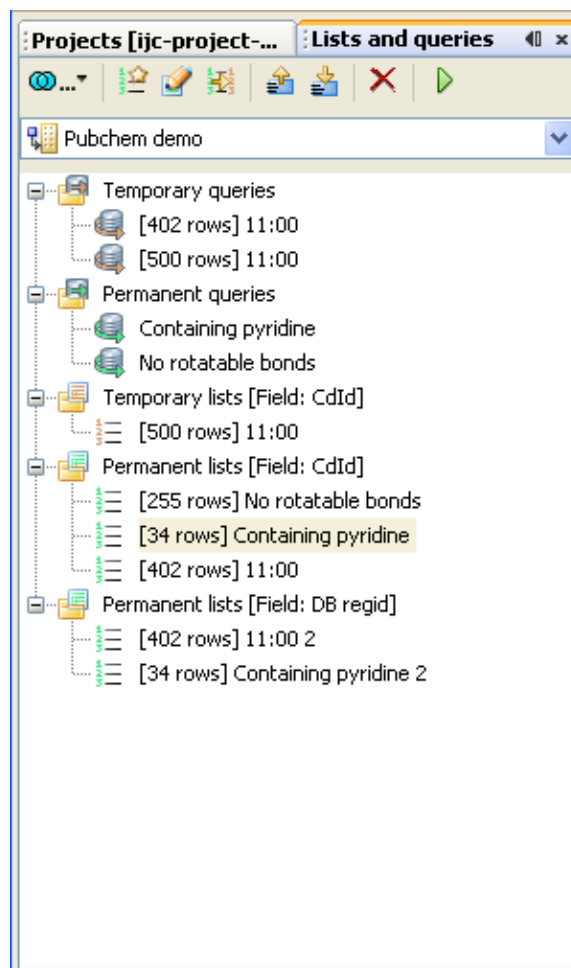
Several important features added since last year's UGM

Database features

- Support for multiple Oracle schemas
- Support for text based primary keys
- Support for database views
- Improved support for JChem cartridge
- Improved schema editor
- Additional options for security
- Performance improvements

List management

- Lists for any integer or text field
- Convert lists between different fields
- Copy/paste
- Drag and drop



Miscellaneous

- Improved printing
- Reactor integration
- Additional similarity search metrics
- Improved look and feel
- More flexible structure renderer
- Import of Markush DARC files

New for 5.4

- License changes
- Visualization module
- Form builder improvements
- Performance improvements
- Calculations *
- Scripting support *
- LogP/D + pKa training

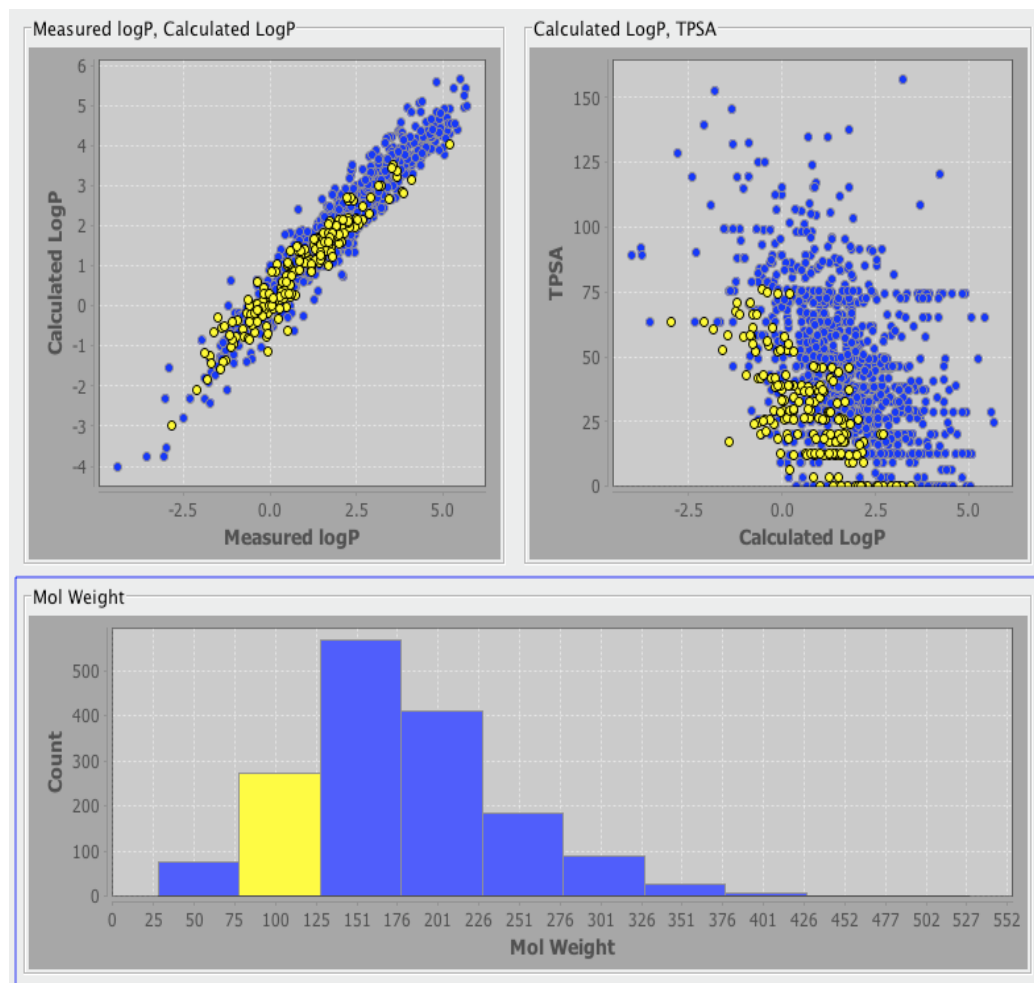
* = may not be in initial 5.4 release

License changes (5.4)

- IJC personal license
 - License now needed for usage with local databases
 - Without license IJC can be used as “free viewer”
- IJC Visualisation
 - Chart widgets
 - Conditional formatting
 - SAR tools
 - Advanced analysis

Visualization module (5.4)

- Chart widgets
 - Histogram
 - Scatter plot
- Conditional formatting *
- Fully integrated with:
 - Selection
 - Query



* = may not be in initial 5.4 release

Visualization module (5.x)

- More chart widgets
 - Line chart
 - X-Y chart
 - Radar (spider, cobweb) chart
 - Pie chart
 - Tukey plot
- Curve fitting
- Advanced SAR analysis
- Advanced structure browsing
- Performance enhancements

More form widgets (5.4)

- Structure matrix
- Multi field sheet
- Tabbed pane
- Allow better forms
 - More compact
 - Less clutter

The screenshot displays the DesignQuery Browse software interface. The main window is titled "DesignQuery Browse" and features a toolbar with various icons. The central area is a "Structure matrix" with columns labeled A, B, and C, and rows numbered 1 through 6. Each cell in the matrix contains a chemical structure. The structure in row 1, column B is highlighted with a blue selection box. To the right of the matrix is a "Screening_Collection" table with the following data:

| Field: | Value: |
|---------------|-------------------------|
| 1 Cdid | 16 |
| 2 Mol Weight | 168.19 |
| 3 Formula | C11H8N2 |
| 4 code | AC 12926 |
| 5 IUPAC Name | 9H-pyrido[3,4-b]indole |
| 6 Smiles | c1ccc2c(c1)[nH]c1cnccc2 |
| 7 Composition | C (78.55%), H (4.79%) |

Below the table is a "Partitioning" section with a "Molecular props" tab. It contains several input fields for molecular properties:

- TPSA: 28.68
- H bond acceptors: 1
- Rotatable bonds: 0
- H bond donors: 1
- Ring count: 3
- Chiral atoms: 0

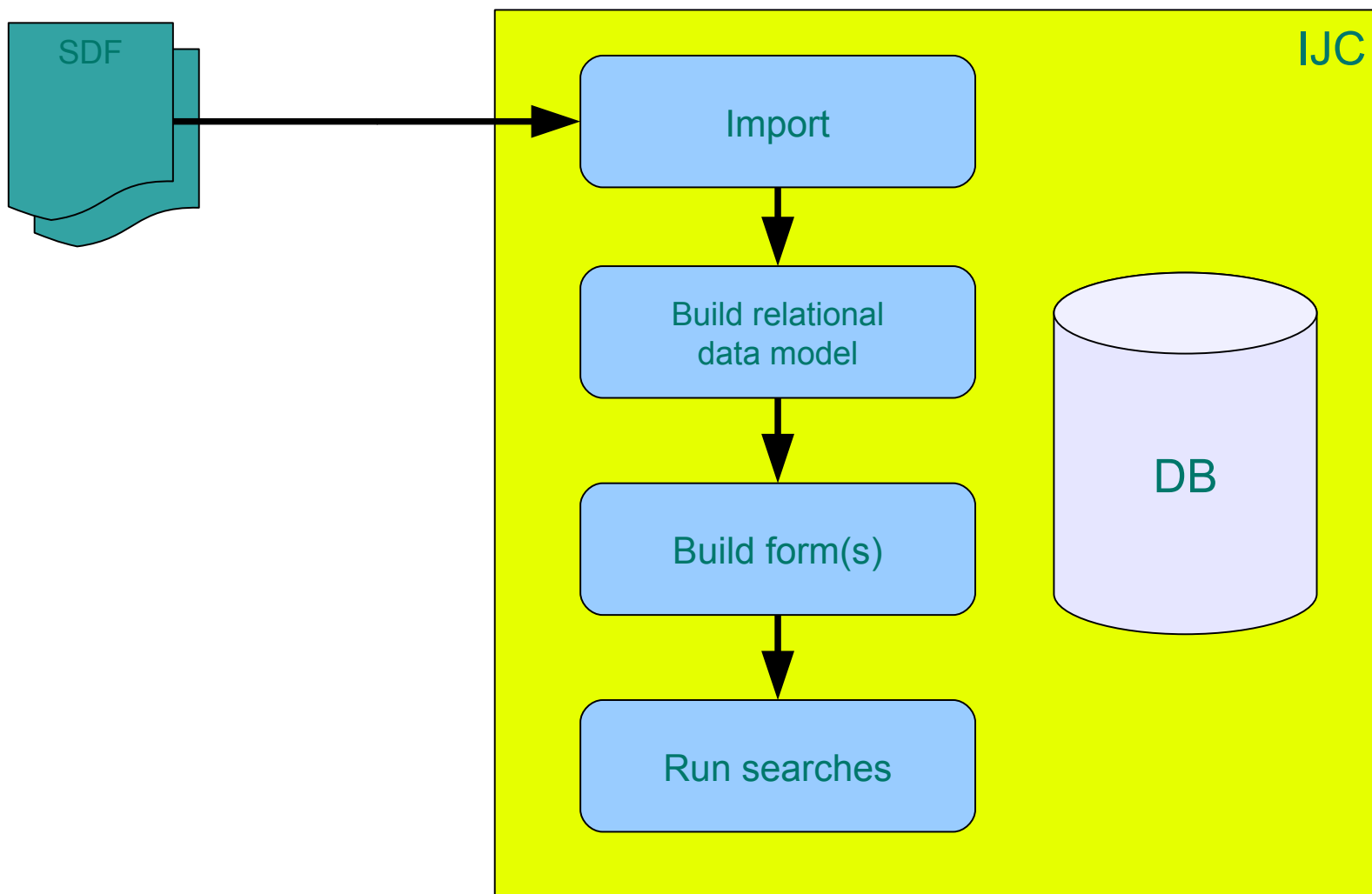
Beyond 5.4

- More database features
- ... and more still
- IJC server
- More visualisation tools
- Improved cherry picking capabilities
- Structure browser (MCS, scaffolds, similarity)
- Additional molecular descriptors (ECFP, BCUT, Pharmacophore...)

Using IJC to deploy or access chemical databases

1. Reporting from Activity Base
2. Supplying DBs to partners/customers -
Markush/Patent DBs

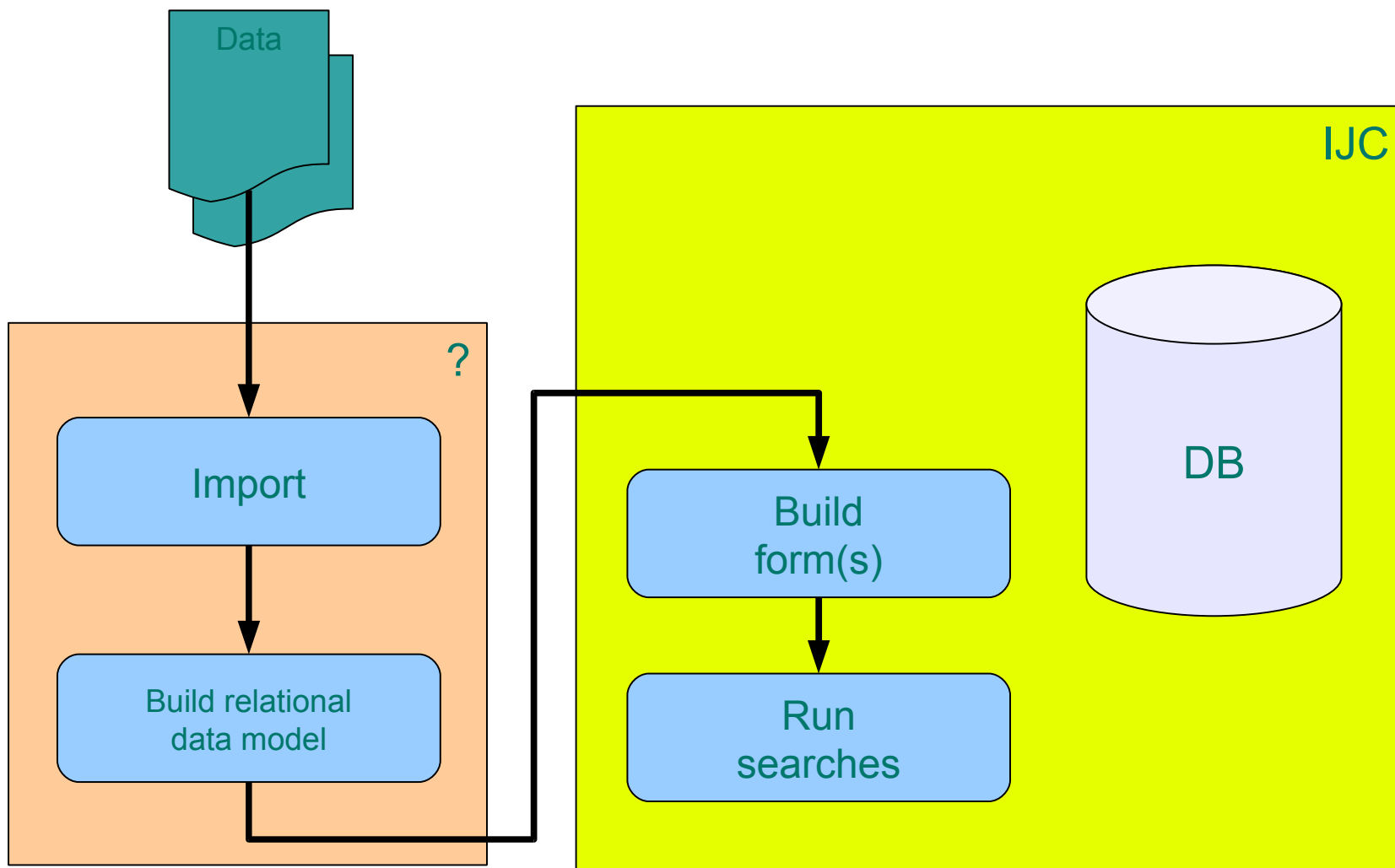
Standard usage



Limitations

- Data is infinitely variable, but
 - Import is not infinitely flexible
 - IJC data model is not infinitely flexible
- Does not allow custom data processing
- Cannot be automated
-
 - We are working to remove these limitations
 - But what to do now when you run into the buffers?

Paradigm for non-standard usage



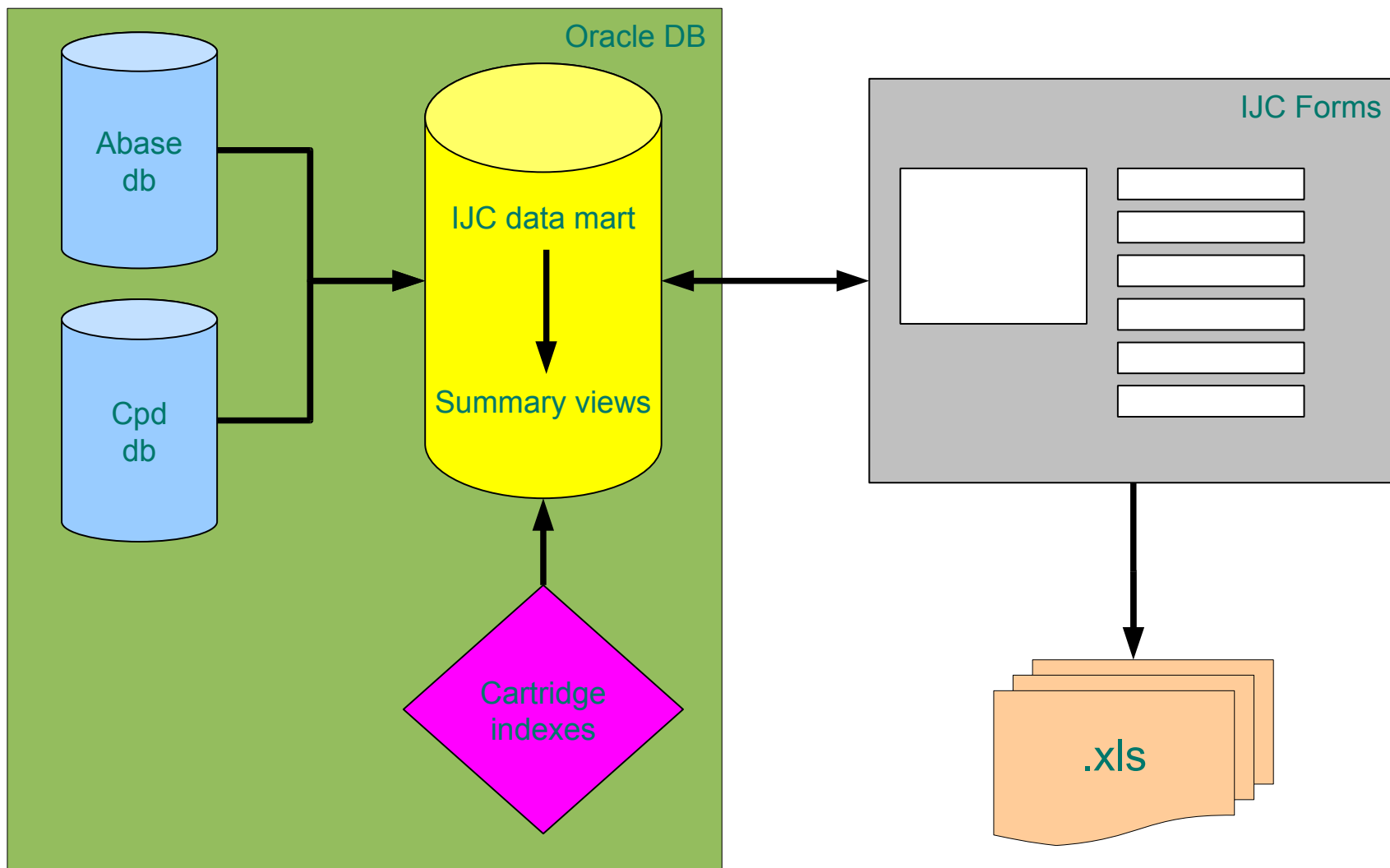
Data processing external to IJC

- Database is a standard SQL database
 - Nothing hidden, even for local databases
 - Can be manipulated by anything that can use JDBC and JChem API
- Allows much more flexibility and automation
- Many approaches possible
 - SQL
 - Pipeline Pilot, Knime...
 - Scripts/programs
- Once processing is complete data can be picked up and used by IJC

1. Reporting chemical and assay data

- Biotech company with assay data in Activity Base
- Lots of data, but no good way of viewing, searching, or distributing it
- Needed simple, easy to use, but inexpensive solution

Solution



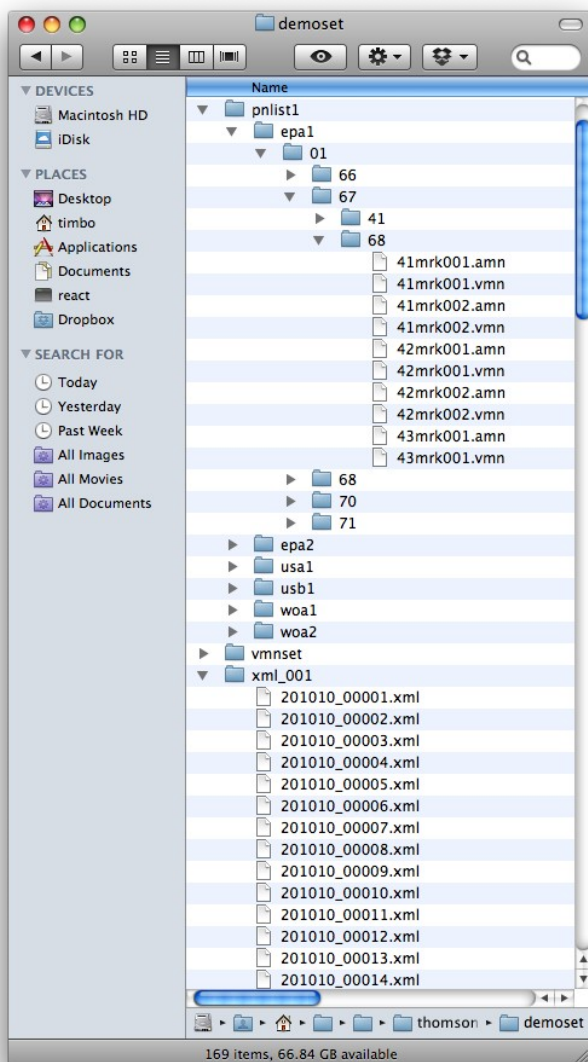
Use of materialised views

- Allows data from highly normalised data model to be converted to a form better suited to reporting
- Provides fast access to data
 - Data is already pivoted
 - Indexes added as needed
 - Summarized (averaged) data available
- JChem cartridge indexes can be added to materialised views (e.g. smiles column)
- Materialised views refreshed as needed (e.g. every night)

2. Markush demoeset

- Markush demo data from Thomson Reuters
 - VMN files containing Markush structures
 - XML files containing patent data
 - Other files (images, PDFs...)
- IJC would be ideal reporting tool for this data
 - But can't import the data as its too complex
- Quick solution needed to allow customers to evaluate Markush technology + data

The data



```
1 <?xml version="1.0" encoding="UTF-8" standalone="no"?>
2 <!DOCTYPE tsip:dataFeed SYSTEM "dataFeed.dtd">
3 <tsip:dataFeed xmlns="http://schemas.thomson.com/ts/20041221/tsip" xmlns:dwpi="http://schemas.thomson.com/ts/20041221/dwpi" xmlns:
4 <tsip:tddVersion="22" tsip:date="2009-08-12">
5 <invention tsip:action="replace" tsip:src="dwpi" tsip:pan="2006255082">
6
7 <accessions>
8 <updates>
9 <assignees tsip:action="replace">
10 <assignee>
11 <assigneeTotal tsip:form="dwpi">SANOFI AVENTIS DEUT GMBH</assigneeTotal>
12 <name>
13 </name>
14 <assigneeCode tsip:src="dwpi" tsip:codeType="dwpi:CompanyStd">SNFI</assigneeCode>
15 </assignee>
16 <assignee>
17 <assigneeTotal tsip:form="dwpi">SANOFI-AVENTIS DEUT GMBH</assigneeTotal>
18 <name>
19 </name>
20 <assigneeCode tsip:src="dwpi" tsip:codeType="dwpi:CompanyStd">SNFI</assigneeCode>
21 </assignee>
22 </assignees>
23 <inventors>
24 <publications>
25 <applications>
26 <relateds>
27 <priorities>
28 <metaData>
29 <classificationIpcCurrent>
30 <classificationEciaCurrent>
31 <classificationUsCurrent>
32 <classificationJpCurrent>
33 <titleEnhanced tsip:action="replace" tsip:lang="en" tsip:src="dwpi" tsip:cc="DE" tsip:se="102004046492" tsip:ki="A1">
34 <title tsip:lang="en" tsip:input="intellectual">New 4-cyclobutenylaminophenyl-tetrahydroisoquinoline derivatives,
35 </titleEnhanced>
36 <abstractEnhanced tsip:action="replace" tsip:lang="en" tsip:src="dwpi" tsip:cc="DE" tsip:se="102004046492" tsip:ki="A1">
37 </abstractEnhanced>
38 <noveltyAscii>
39 <descriptionAscii> <paraAscii>4-cyclobutenylaminophenyl-tetrahydroisoquinoline derivatives of formula (I) and
40 <paraAscii>R1-R4 = hydrogen, fluoro, chloro, bromo, iodo, cyano, nitro or R11-(CmH2m)-An-;</paraAscii>
41 <paraAscii>m = 0-4;</paraAscii>
42 <paraAscii>n = 0 or 1;</paraAscii>
43 <paraAscii>R11 = hydrogen, methyl, CpF2p+1 or phenyl;</paraAscii>
44 <paraAscii>p = 1-3;</paraAscii>
45 <paraAscii>A = 0, NH, NMe or S(O)q;</paraAscii>
46 <paraAscii>q = 0-2;</paraAscii>
47 <paraAscii>R5 = hydrogen, 1-6C alkyl, optionally (partially) fluorinated, or 3-6C cycloalkyl;</paraAscii>
48 <paraAscii>R6 = hydrogen, hydroxy, trifluoromethyl, Me, Et, isopropyl or cyclopropyl;</paraAscii>
49 <paraAscii>R7 = hydrogen, 1-6C alkyl, 3-6C cycloalkyl, OR12 or NR13R14;</paraAscii>
50 <paraAscii>R12 = hydrogen, 1-6C alkyl, optionally (partially) fluorinated, or 3-6C cycloalkyl;</paraAscii>
51 <paraAscii>R13, R14 = hydrogen, 3-6C cycloalkyl, 1-6C alkyl (optionally (partially) fluorinated), phenyl,
52 <paraAscii>R13+ R14 = a 3-9 membered ring in which one C may be replaced by O or NMe;</paraAscii>
53 <paraAscii>R15, R16 = hydrogen or 1-4C alkyl;</paraAscii>
```

Solution

- Import VMN files using IJC file import
- Load data from the XML files into additional DB tables using custom Groovy script
- Join the two sets together
- [Future] provide links to images and actual patent data
- Provide IJC project to customers as zip file

Groovy script

- Creates tables
- Trawls XML files
- Parses XML
- Assembles data
- Loads into DB
- < 200 lines code

```
...
File root = new File('/Users/timbo/data/structures/thomson/demoset/xml_001')
def files = root.listFiles()
def fileCount = 0
def inventionCount = 0
def markushCount = 0

files.each {
    if (it.name.endsWith(".xml")) {
        println "Processing file $it"
        def dataFeed = new XmlParser().parse(it)
        fileCount++

        dataFeed.tsip.invention.each {
            inventionCount++
            //println "invention"
            String assignee = ""
            it.assignees.each {
                it.assignee.assigneeTotal.each {
                    //println "  assignee = ${it.text()}"
                    assignee += it.text() + "\n"
                }
            }
        }

        def patents = ""
        it.publications.each {
            it.publication.each {
                def status = it.attribute(tsip.status)
                def countryCode = it.documentId.countryCode.text()
                def kindCode = it.documentId.kindCode.text()
                def date = it.documentId.date.text()
                def number = it.documentId.number.text()
                def summary = "${countryCode}${number}-${kindCode}"
                if (status == "dwpi:basic") {
                    summary += " *"
                }
                //println "  patent = ${summary}"
                patents += "${summary}\n"
            }
        }
        //println "  patents = \n${patents}"
    }
}
...
```

Result

Instant JChem 5.3.3

9 / 300

Welcome | Grid view for vmns | Grid view for Inventions | Form view for Inventions

Design Query Browse Entity: Inv...

Files
Lists and queries
Services
Projects [jic-project-markush-demoes]

| ID | Title | Description | Assignees |
|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| 9 | New pyridine or pyrimidine derivatives are cannabinoid receptor 1 inhibitors, useful to treat a disease mediated by cannabinoid-1 receptor e.g. eating disorders associated with excessive food intake | 11 Pyridine or pyrimidine derivatives (A) e.g. 4,5-dihydro-1H-pyrazolo(3,4-d)pyrimidine compound of formula (a) or 1H-pyrazolo(3,4-b)pyridine compound of formula (b) and their salts, hydrates, solvates and isomers are new. Y1 = O, NR7 or S; R7 = H, OH or 1-6C alkyl; R1 = 5-10C heteroaryl, 3-12C cyclolalkyl, phenyl or benzyl (all optionally substituted by 1-3 radicals of halo, OH, CN, NO2, 1-6C alkyl, 1-6C alkoxy, halo-substituted 1-6C alkyl, halo-substituted 1-6C alkoxy, NR8R9, S(O)0-2R8, -C(O)OR8 or R10); R2 = 3-8C heterocycloalkyl, 5-10C heteroaryl, phenyl or phenoxy (all optionally substituted by 1-3 radicals of halo, OH, CN, NO2, 1-6C alkyl, 1-6C alkoxy, halo-substituted 1-6C | IRM LLC CHOI H ELLIS D A HE X HE Y LIU H NGUYEN T N WANG Z |

| vmns | Markush structure | compound number |
|------|-------------------|-----------------|
| 1 | | 0326-08801 |
| 2 | | 0326-08802 |
| 3 | | 0326-08803 |
| 4 | | 0326-08804 |

Patents

WO2006047516-A2 *
NO200702352-A
EP1807429-A2
AU2005299421-A1
IN200702514-P1
KR2007057980-A
CN101048408-A
JP2008518016-W
MX2007004936-A1
BR200517015-A
TW200630096-A
US20090247517-A1
KR919524-B1

Use

(A) are useful for treating a disease mediated by cannabinoid-1 receptor (eating disorder associated with excessive food intake (preferred) (obesity (preferred), bulimia nervosa or compulsive eating disorders), metabolic disorders associated with metabolic disorders including obesity, bulimia nervosa, compulsive eating

Mechanism Of Action

Cannabinoid receptor 1 inhibitor. The ability of (I) to inhibit cannabinoid receptor 1 was tested in a mice using a biological assay. The results showed that (I) exhibited competitive inhibitor constant value of less than 100 nM.

Activity

Eating-Disorders-Gen.; Anorectic; Tranquillizer; Antidiabetic; Antiarteriosclerotic; Hypotensive; Gynecological; Cardiovascular-Gen.; Antiarthritic; Osteopathic; Dermatological; Antilipemic; Hypnotic; Neuroleptic; Antiaddictive; Antidepressant; Anticonvulsant; Antimanic; Nootropic; Neuroprotective; Antiparkinsonian; Cerebroprotective; Vulnerary; Hypertensive; Respiratory-Gen.; Cardiant; Immunosuppressive; Antirheumatic; Antimigraine; Antiinflammatory; CNS-Gen.; Vasotropic; Gastrointestinal-Gen.; Antiallergic; Antipsoriatic; Antiasthmatic; Thyromimetic; Cytostatic.

Inventions: 300 out of 300 rows. vmns: 4 out of 579 rows.

Summary

- IJC database + forms + chemistry features provides good solution for serving up database
- Major improvements in the way data can be visualized are coming soon
- Sometimes custom processes will be needed to prepare the data
 - Administrators do the hard work
 - Users reap the benefits