



# A new system for searching historic and ELN reaction data

Gregory Landrum Ph.D., Anna Pelliccioli Ph.D.










NIBR IT

Novartis Institutes for BioMedical Research

Basel

ChemAxon UGM 2012 Budapest

# Reminder... moving away from ISIS and ISIS/Host

- Screening sample database   Instant JChem
- ACD   Instant JChem
- MDDR, CMC
- Metabolite 
- Chirbase   Instant JChem
- Current Chemical Reactions 
- Historical collection of Novartis reactions  **New Web application**

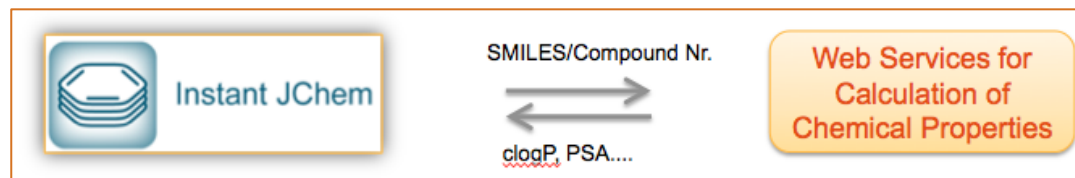
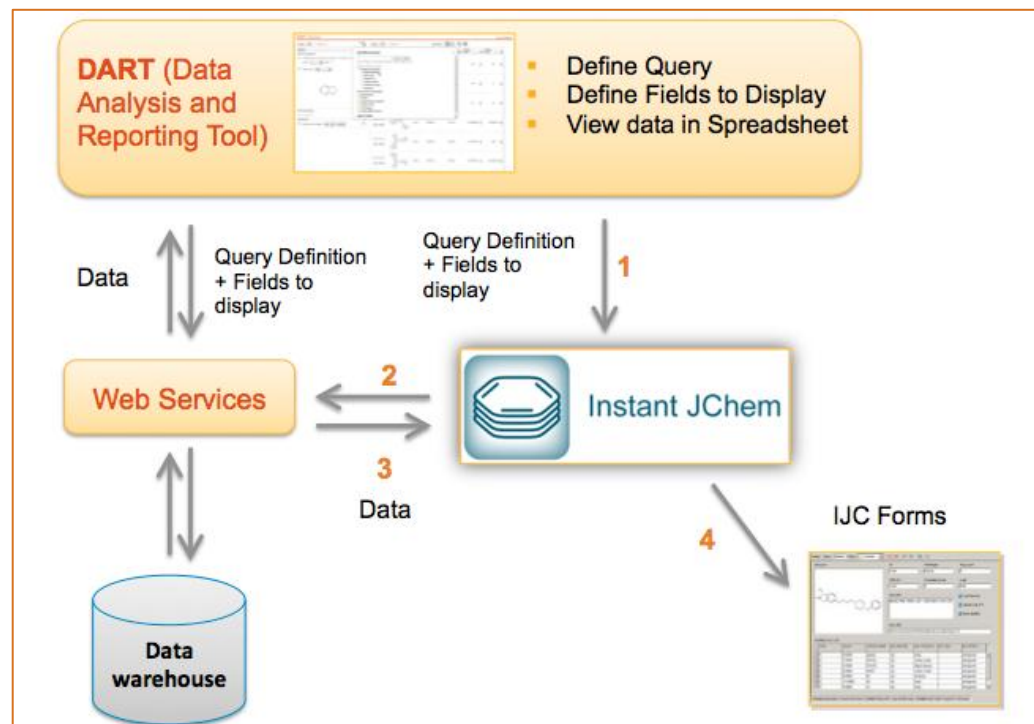
# Why build a new reaction-database application?

---

- Current situation :
  - Literature and historic (pre-eLN) reactions: stored in an ISIS/host database; searchable from ISIS/base.
  - eLN reactions: stored in the eLN (CambridgeSoft) schema; “searchable” from eLN
- Objective:
  - allow NIBR scientists to quickly and efficiently search both sets of reactions in one place
- Requirements:
  - web-based UI for searching and browsing
  - web services for searching and retrieving the data

# An Aside: services

- Serious effort underway to expose every computation possible via web service.
- Some examples:
  - Querying our new data warehouse
  - Calculating descriptors, using predictive models, etc.
  - Registering compounds/biologics, querying registration system
- Advantages:
  - portability
  - instantly global
  - easy updates/deployments
  - write once, use everywhere
  - mashups



# Reduction to practice

---

- Collaboration with ChemAxon
- Collection of user stories within NIBR
  - global team of medicinal chemists driving requirements
  - informed by Pistoia Alliance questions together with experience with CERES + various commercial tools
- Concrete requirements and design done collaboratively
- Implementation by ChemAxon

# Some query requirements

---

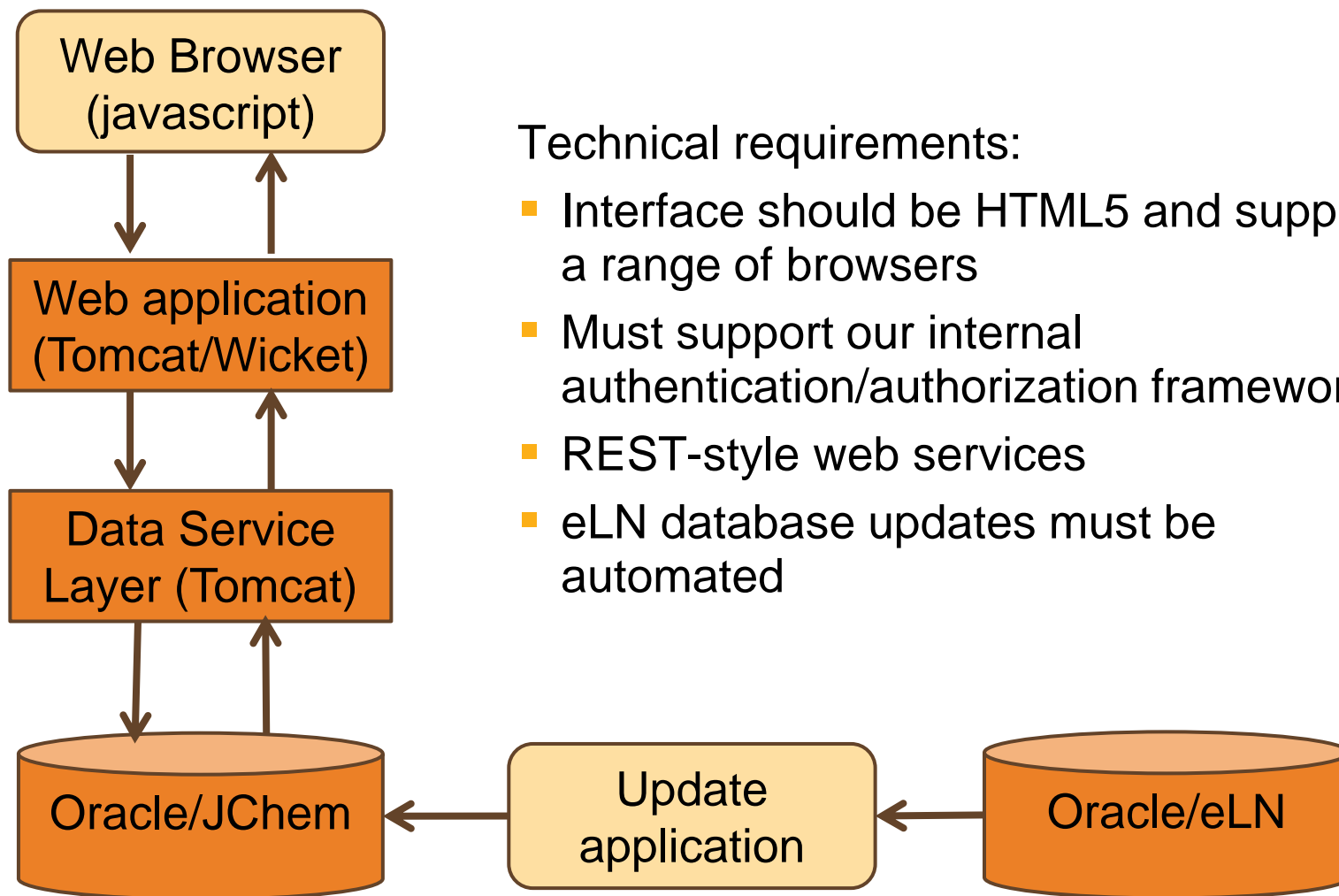
- Retrieve results based on structure searches (exact and substructure) of reactants and/or products
- Retrieve additional information for a particular reaction
- Query across multiple data sources
- Limit results by yield, reagent, solvent, etc.

# Complexity of the historical data

---

- Multiple collections of internal data in separate ISIS/Host databases with different structures
- Constructing a single schema that can hold everything is unrealistic
- Solution: Bring over only the pieces of information that are actually needed for searching/browsing. Present additional data using an alternate “details” page that can be different for each original source
- Standard fields:  
Reaction, date, chemist, source, project, yield, solvent, reagents
- Requires a custom RDF importer for each historical source

# Architecture



## Technical requirements:

- Interface should be HTML5 and support a range of browsers
- Must support our internal authentication/authorization framework
- REST-style web services
- eLN database updates must be automated



# Status

---

- beta version of web interface and services complete
- historic reaction data (~400K reactions) loaded
- starting to test the application with our chemists
- planning eLN -> JChem data migration (~600K additional reactions)

“Demo”

---

# Interface Overview

Structure

Search type  
SUBSTRUCTURE

Open Sketcher

Options

Source

Other fields

Chemist

Project

Yield greater than

Reaction

No Records Found

Source

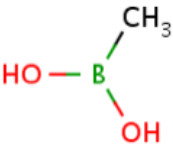
Reagent

Solvent

# Substructure search with results

Structure

Search type  
SUBSTRUCTURE



Options

Source

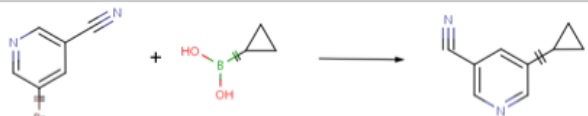
Other fields

Chemist

Project

Product yield greater than

Reaction



yield 100% with K3PO4;SK-CC02-A in Dioxane by Machauer Rainer;

yield 100% with K3PO4;SK-CC02-A in Dioxane by Machauer Rainer;

yield 21% with

yield 21% with

yield 48.7% by Koch Guide; Koch Guide;

Source

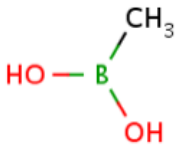
Reagent

Solvent

# Search-results filters

**Structure** ▾

Search type  
SUBSTRUCTURE ▾



**Options** ▾

**Source** ▾

**Other fields** ▾

Chemist

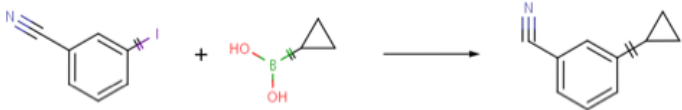
Project

Product yield greater than

**Reaction**

yield 100% with K3PO4;SK-CC01-A in Dioxane by

yield 100% with K3PO4;SK-CC01-A in Dioxane by



yield 100% with K3PO4;SK-CC01-A in Dioxane by Machauer Rainer; Z

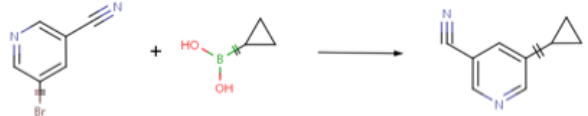
yield 100% with K3PO4;SK-CC02-A in PhMe;H2O by Machauer Rainer;

yield 100% with K3PO4;SK-CC02-A in PhMe;H2O by Machauer Rainer;

yield 100% with K3PO4;SK-CC02-A in PhMe;H2O by Machauer Rainer;

yield 100% with K3PO4;SK-CC01-A in Dioxane by Machauer Rainer; Z

yield 100% with K3PO4;SK-CC02-A in PhMe;H2O by Machauer Rainer;



yield 100% with K3PO4;SK-CC02-A in Dioxane by Machauer Rainer;

yield 100% with K3PO4;SK-CC02-A in Dioxane by Machauer Rainer;

**Source** ▾

Check all | Uncheck all

<input type="checkbox"/> Novartis Backlog	359
<input checked="" type="checkbox"/> SynLab1	128
<input type="checkbox"/> NovRes	109
<input type="checkbox"/> SynLab2	90
<input type="checkbox"/> TanPromoted	80

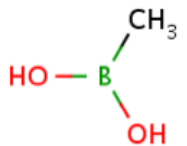
**Reagent** ▾

**Solvent** ▾

# Additional search criteria + filters

**Structure**

Search type  
SUBSTRUCTURE



**Options**

**Source**

**Other fields**

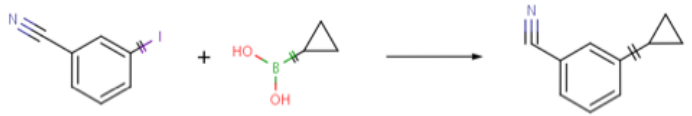
Chemist

Project

Product yield greater than  
80

Search Reset

**Reaction**



yield 100% with K3PO4;SK-CC01-A in Dioxane by Machauer Rainer; [redacted]

yield 100% with K3PO4;SK-CC02-A in PhMe;H2O by Machauer Rainer; [redacted]

yield 100% with K3PO4;SK-CC02-A in PhMe;H2O by Machauer Rainer; [redacted]

yield 100% with K3PO4;SK-CC02-A in PhMe;H2O by Machauer Rainer; [redacted]

yield 100% with K3PO4;SK-CC01-A in Dioxane by Machauer Rainer; [redacted]

yield 100% with K3PO4;SK-CC02-A in PhMe;H2O by Machauer Rainer; [redacted]

yield 86.1% w [redacted]

yield 75% by [redacted]

yield 81.7% b [redacted]

yield 86.1% w [redacted]

**Source**

Check all | Uncheck all

- Novartis Backlog 20
- NovRes 19
- SynLab1 16
- TanPromoted 15
- SynLab2 8

**Reagent**

**Solvent**

Check all | Uncheck all

- CH2Cl2 17
- Dioxane 11
- H2O 10
- THF 10
- PhMe 7
- nPROH 6
- Et2O 4
- EtOAc 4
- Hexane 3
- EtOH 3

# What's next?

---

- Get the application out to our test users and collect feedback
- Get that eLN data loaded
- Start thinking about better ways of organizing the results
- Connecting to other web apps
- ???

# Acknowledgements

---

## NIBR GDC

- Philipp Lehr
- Klemens Hoegenauer
- Berndt Oberhauser
- Wolfgang Miltz
- Gurdip Bhalay
- David Barnes

## NIBR IT

- Thomas Mueller
- Rajshekhar Aurade
- Sadik Keshwani

## ChemAxon

- Tim Dudgeon
- Mihály Medzihradzky
- Gustavo Santucho
- Mario Burdman
- Gabriel Moreno



---

*Thanks!*