A Sneak Peek at the ChemAxon Compound Registration Service

Jonathan Lee

ChemAxon

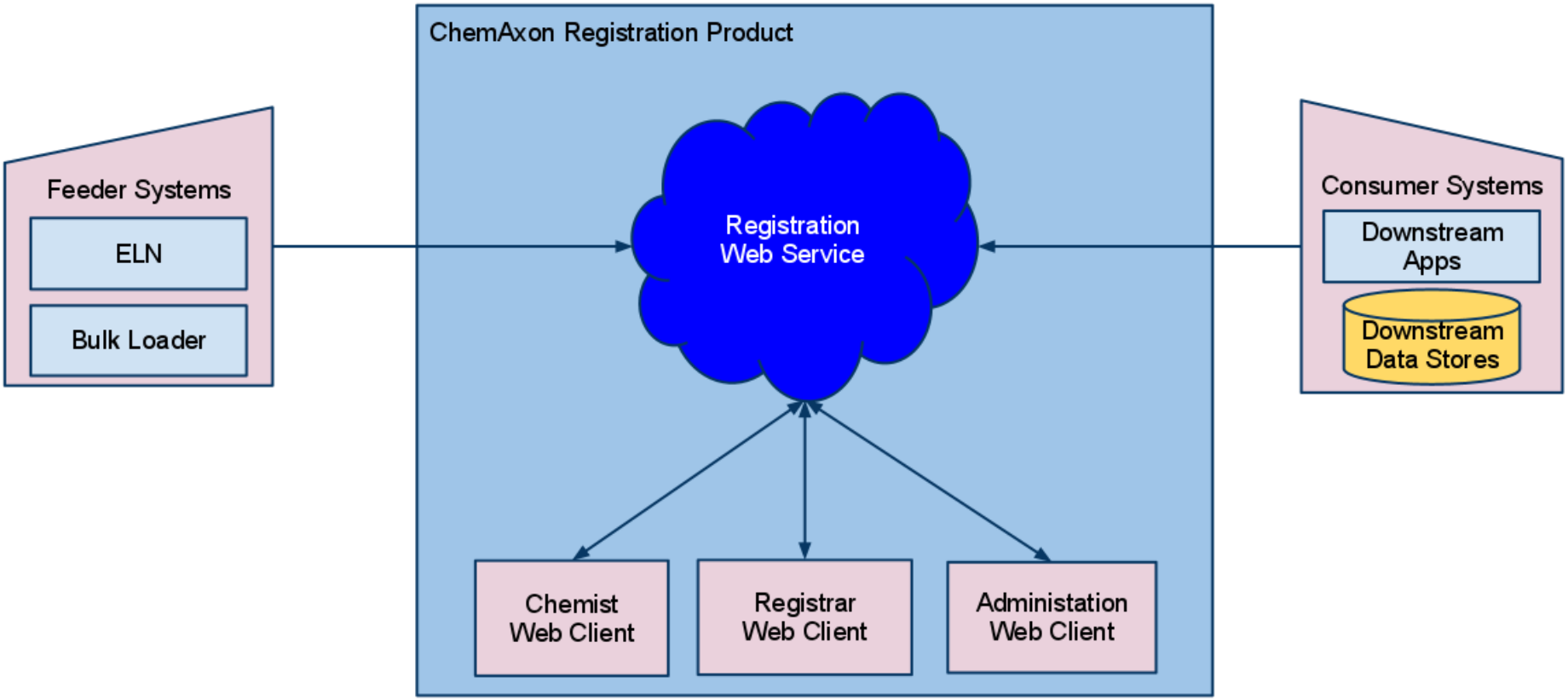
Solutions for Cheminformatics

Introduction

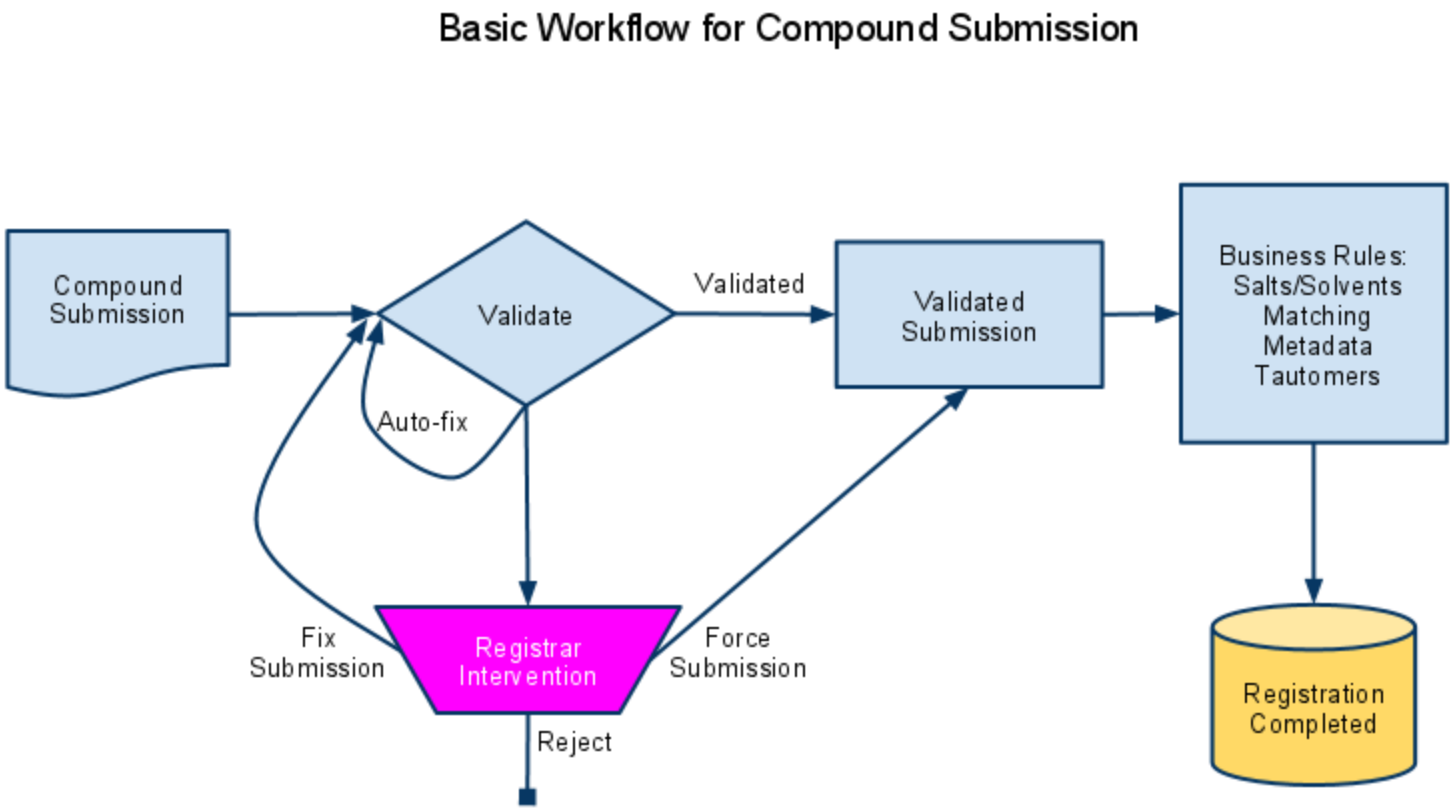
- The Compound Registration Service handles **compound submission and management into a registration data store**.

- ChemAxon and GlaxoSmithKline are **collaborating using an agile development methodology** to replace and improve the current GSK registration service.

Architecture



Submission Workflow



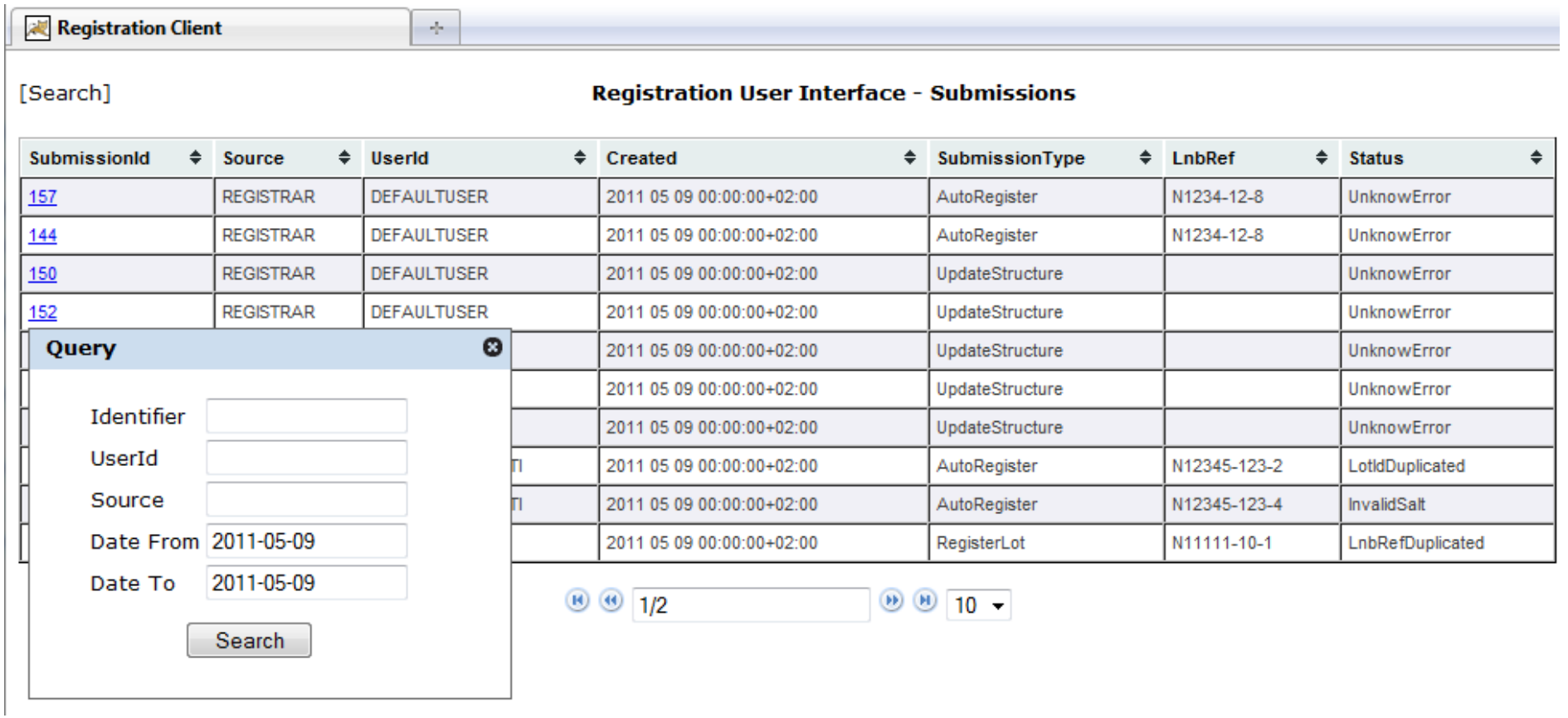
Major Features

- Automatic validation and automatic fix
  - Standardizer
  - Structure Checker
  - Configuration file adjustment
- Single and multicomponent compounds
  - Alternates, formulations, mixtures
- Salt/Solvate handling
  - Registering with parent to salt/solvate ratio
  - Automatic salt/solvate splitting
  - Administration of salt/solvate dictionary (search, add, remove)

Major Features

- Staging area and correction page
  - Possible hits (shows matches of current struct)
  - Match list (unique, replace, accept)
  - Version correction according to business rules
  - Single and batch correction
- Amendment of registered compound (Structure, salt/solvate, metadata, hierarchy)
- Audit history of amendments
- Registry search client (duplicate, taut., 2D)
- Bulk load (SDF to a service)

Demo


Registration Client
[Search]
Registration User Interface - Submissions
SubmissionId
Source
UserId
Created
SubmissionType
LnbRef
Status
157 REGISTRAR DEFAULTUSER 2011 05 09 00:00:00+02:00 AutoRegister N1234-12-8 UnknowError
144 REGISTRAR DEFAULTUSER 2011 05 09 00:00:00+02:00 AutoRegister N1234-12-8 UnknowError
150 REGISTRAR DEFAULTUSER 2011 05 09 00:00:00+02:00 UpdateStructure UnknowError
152 REGISTRAR DEFAULTUSER 2011 05 09 00:00:00+02:00 UpdateStructure UnknowError
2011 05 09 00:00:00+02:00 UpdateStructure UnknowError
2011 05 09 00:00:00+02:00 UpdateStructure UnknowError
2011 05 09 00:00:00+02:00 UpdateStructure UnknowError
2011 05 09 00:00:00+02:00 AutoRegister N12345-123-2 LotIdDuplicated
2011 05 09 00:00:00+02:00 AutoRegister N12345-123-4 InvalidSalt
2011 05 09 00:00:00+02:00 RegisterLot N11111-10-1 LnbRefDuplicated
Query
Identifier
UserId
Source
Date From 2011-05-09
Date To 2011-05-09
Search
1/2
10

Technology Details

- Compound input format
  - Structural information in MOL V3000
  - Chemically significant metadata (e.g. stereo isomer ranges, textual description)
- JChem Base
- **Marvin** and ChemDraw compatible
- **MySQL**, Oracle database
- Tomcat web container
- AJAX web client
- Service Oriented Architecture

Future Plans

- Marvin input format
- Robust migration and bulk loading tools
- Pluggable compound ID generation
- Configurable hierarchy levels
- Role-Customizable amendment and access

Working with ChemAxon

- Agile software methodology
- See progress in short iterations (e.g. 2-3 weeks)
- You easily refine requirements; Developers react to changing requirements
- Measurable velocities provide a better prediction of completion dates

Registration Service Team

ChemAxon

- Edvard Buki, Jonathan Lee, Roland Molnar, Diana Nagy, Akos Papp, Sarolta Pilbak, Attila Tajti

Special Thanks to the team from GlaxoSmithKline

- Rama Bhamidipati, Matt Biggs, Matt Jones, Ian Mawer, Paul Wallace, Shane Weaver, Charlie Wilkins

Thank you for your attention

- We’re looking for Pilot customers

- Questions?

Chemistry Features

- Tautomer Recognition
- Stereochemistry Intelligence
  - Double Bonds
  - Tetrahedral Stereo
  - Enhanced Stereo
  - Basic Allenes
  - Syn/Anti, Endo/Exo
- Future: More stereochemistry features
  - All allenes and cumulenes
  - Hindered rotamers (atrop)
  - Fischer/Haworth projections
  - Helicenes