



RaaS: Registration as a Service Full story after half-year

Rama Bhamidipati
25 Sep 2012

Background

- 2009: GSK looking to simplify IT landscape and reduce IT spend
- One program targeted a Registration-as-Service (RaaS) solution
 - Replace, and if possible enhance, existing registration system
 - Vendor supported, outside GSK firewall
- 2010: GSK Legal raise IP concerns for hosting registration system outside firewall
- 2011: GSK change vendor to Chemaxon, moving to a corporate chemistry solution based on JChem
 - Scope of registration system now replace functionality only
- Solution proposed to host registration system in GSK data centre, but managed by Chemaxon
 - SaaS product, without IP risk

Technology – Breaking New Ground

- Registration arena:
 - Chemaxon limited experience in for large pharma registration systems
 - Proposed technology MySQL, JChem - unproven
- How would multi-million structure database perform?
- Implementation of a registration service
 - Moving away from SDFfiles
- Needs to work with existing processes
- Implementation of V3 molfile format
 - Co-existence with V2 molfile applications

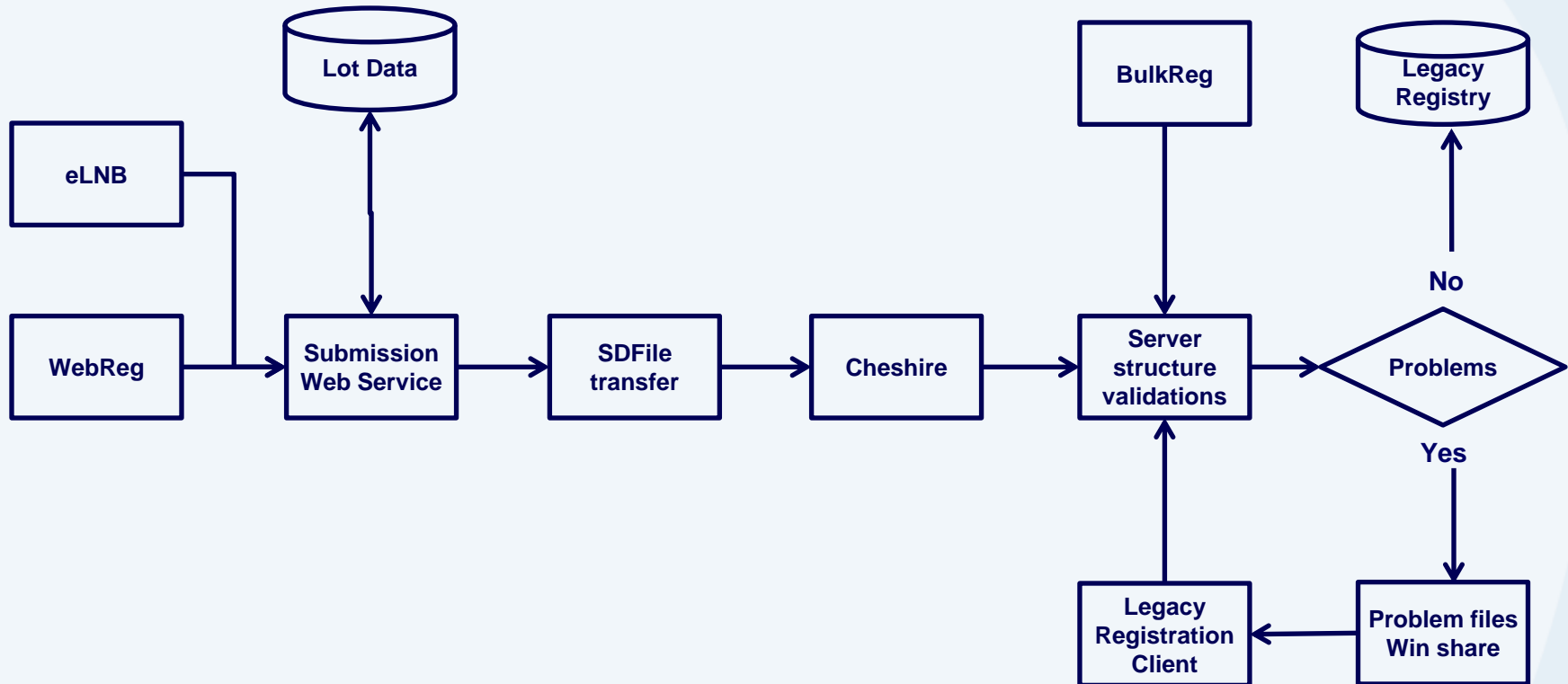
Registration Service

- Maintaining existing data structure
 - Parent – Version – Preparation
- Moving away from preferred/non-preferred compound numbers
 - One compound, (possibly) many compound numbers to one compound, one compound number
- Highly proprietary functionality
 - Stereochemistry handling – atom and bond level markups
 - Peptide handling
- Novelty checking – performance, functionality
 - Handling of no-structure records
- Automation
- Registration throughput

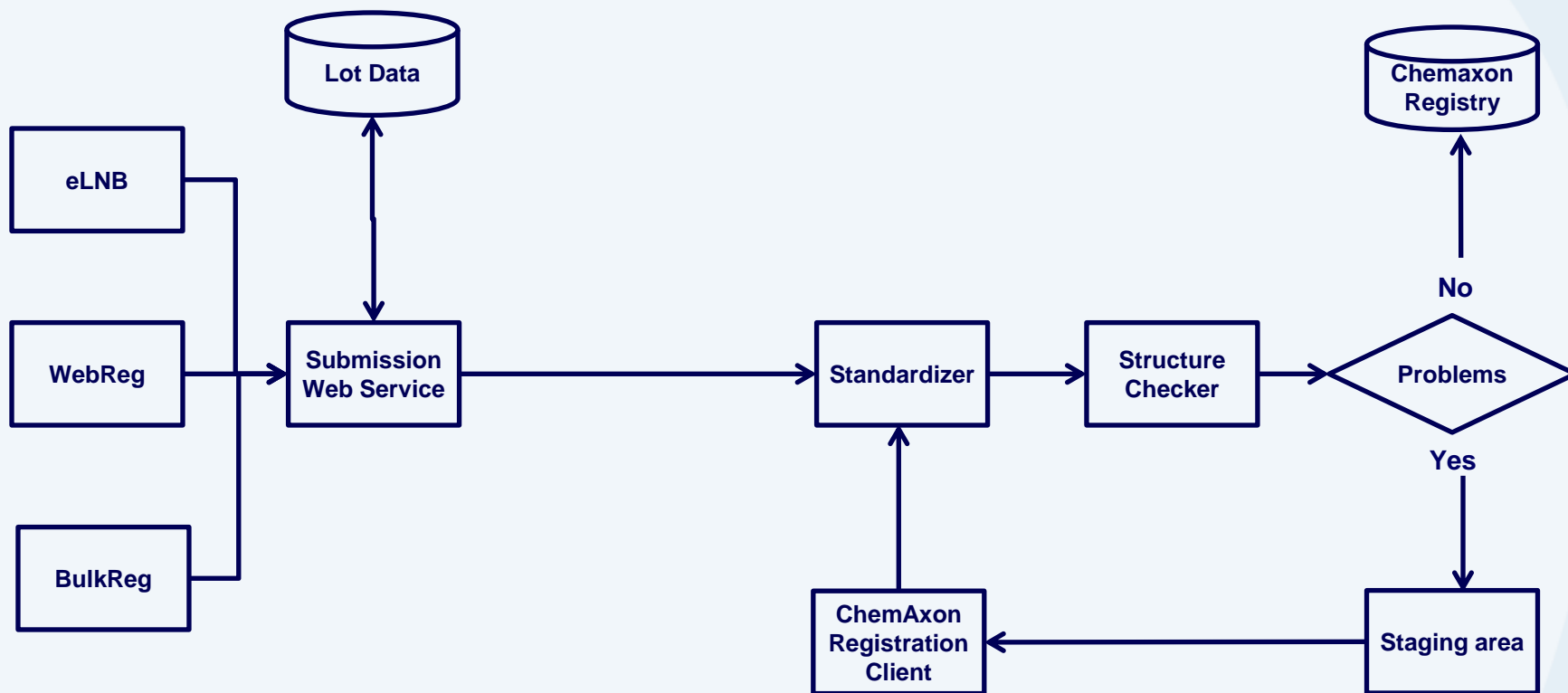
Data handling

- Migration of 5+ million records
 - How long to migrate
 - Confirming successful migration
 - How would this be done
- V2 <-> V3 molfile conversion
- Peptide handling
- Tautomer resolution
- Moving non-preferred data to preferred data tree
- Non-structural data
 - Key for no structure records
 - Many fields

Registration Process – Legacy



Registration Process – New



Project Delivery

- High level requirements were provided
- Distilled into use cases
 - Learnings on both sides as to level of detail required
- Started with simple cases, built up complexity
- Registration rules implementation
 - Use cases crucial to understanding
- Registration service built to support both registration of data and migration
 - Tested with simple registrations and migrations
 - Iterative process
 - Build up complexity and rules

Project working

- Initially slow delivery and feedback on both sides
- Introduced biweekly sprint planning
 - Took time to stabilize
 - Helped with understanding on both sides, improved scheduling
 - Task prioritisation
- Daily scrums
 - Highlighted problems earlier, quicker solutions
- No 'blame' games
 - Focus on issue, arrive at pragmatic solutions
- Use of technology
 - Live Meeting, Skype, helped
- Helped meet aggressive timelines

Production Launch

- Scheduled for mid-March
 - Not all functionality would be provided
 - Further releases scheduled for May, June to deliver remaining functionality
- Data migration began start Feb
- Registration shut down one week
 - Migration top up
 - Comparison of data, number of compounds etc.
 - All systems switched to use new Registry system
 - Submissions continued, held in queue
- Controlled release of waiting submissions
 - Chemaxon and GSK project team all present

Production Registration

- First registration was an auto-registration!
- Second submission was a huge hybrid peptide
 - Caused some trouble but registered successfully
 - Almost rolled-back to a previous version of registration client
- Trickled through more submissions to ensure registration working as expected
- Two days earlier than targeted, registration restarted
- All GSK registrars worked together via Live Meeting, registering compounds from staging area, documenting any issues and comparing notes

Production Working

- Functionality releases delivered as planned
 - Additional releases made in July, August
- Issues raised in production use also fixed
 - Peptide handling improved
 - No structure registration performance 10x better
 - Monitoring of registration processes, e.g. long running SQL queries
- Client enhancements introduced
- Bulk registration tool released
- Maintaining agreed registration SLAs

Accomplishments

- New registration system released
 - Approx. 6M records migrated with novelty check working correctly
 - Registration continuing with very few interruptions to service
 - Legacy systems retired
- Moved away from SDFiles, shared areas, etc.
 - Simplified processes, environment
- BulkReg application released
 - Performance and throughput being improved gradually

What did not go to plan

- Goal of having Registration as a Service not met
- Have reverted to standard vendor relationship
- System not being supported by Chemaxon
 - Chemaxon will fix bugs and provide enhancements to the registration software
 - GSK responsible for installing updates, maintaining the registration software environment

Working Relationship

- Worked as one team toward a common goal – to deliver the best possible system within the constraints put in place.
- A committed and competent team that was on our side every step of the way
 - Developers found as many bugs as GSK testers
 - Equally concerned to fix bugs
- Releases made on Sundays
 - Support team available to ensure system ready to function correctly on Monday

And Finally...

- Thanks to the following for making this a success

GSK

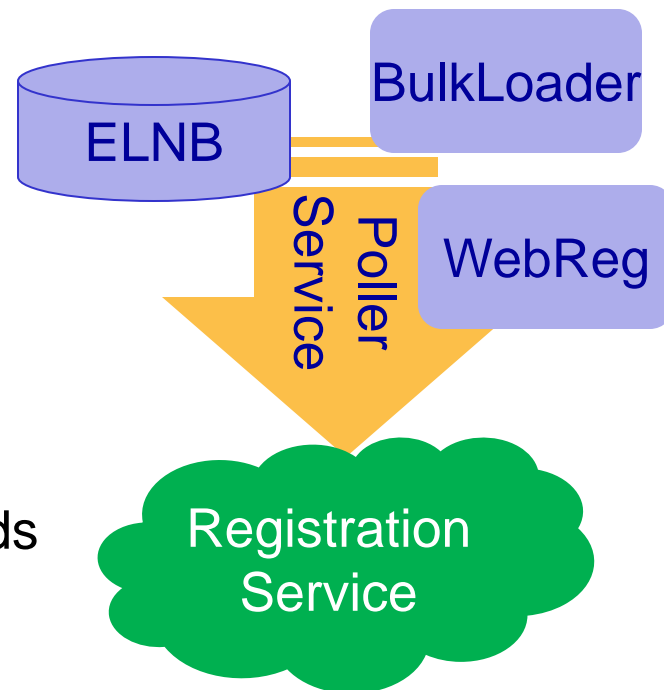
- Charlie Wilkins
- Matt Jones
- Shane Weaver
- Matt Biggs
- Ant Stephenson
- Susan Taylor
- Ian Mawer
- Chris Eubank
- Paul Wallace

Chemaxon

- Akos Papp
- Sarolta Pilbak
- Csaba Peltz
- Srdjan Bejakovic
- Diana Nagy
- Edvard Buki
- Roland Molnar
- Mate Beres

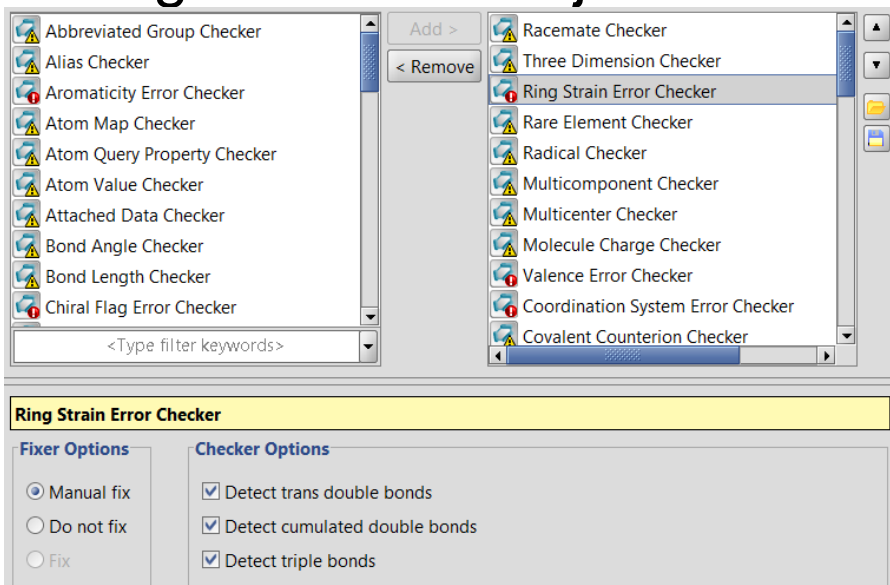
Automatic registration

- Automatic registration of compounds from
 - eLNB
 - Direct submission to registration
 - Web registration client
 - Form based registration
 - Bulk loader
 - Mapping of SDFFile fields to DB fields
- Poller web service to submit any pending data to the registry service



Validation

- Automatic validation and automatic fix
 - Standardizer
 - Structure Checker
 - Quality checks (substructures, unbalanced charge, etc.)
 - Stereochemistry related automatic fixes
 - Configuration file adjustment



Staging Area

- MyStaging vs. AllSubmissions
- Status and detailed message
- Searchable fields
- Items assigned to registrars

Register new compound **Staging area** Amendment Administration Salt search Audit Search local version: 2696 Help Logout

My staging Unassigned All Submissions Search

	id	LnbRef	created on	source	submitted...	opened by	current status
<input type="checkbox"/>		<input type="text"/>	---	<input type="text"/>		<input type="text"/>	
<input type="checkbox"/>	44	N2196-2-4	15-May-2012 08:54:46 GMT	ELNB		admin	StructureChec...
<input type="checkbox"/>	43	N3538-1-1	15-May-2012 08:53:27 GMT	ELNB		admin	StructureChec...
<input type="checkbox"/>	42	N9046-7-5	15-May-2012 08:52:50 GMT	ELNB		admin	UnBalancedC...
<input type="checkbox"/>	40	N3692-5-2	15-May-2012 07:35:52 GMT				ParentTautom...
<input type="checkbox"/>	38	N2425-2-7	15-May-2012 07:32:34 GMT	ELNB		admin	StructureChec...
<input type="checkbox"/>	35	N2401-4-2	15-May-2012 05:17:05 GMT	REGISTRAR		admin	UnBalancedC...
<input type="checkbox"/>	33	N1255-3-7	15-May-2012 05:15:59 GMT	REGISTRAR		admin	UnknownCST
<input type="checkbox"/>	31	N4227-5-6-	15-May-2012 05:15:04 GMT	REGISTRAR		admin	InvalidLnbRef

An existing tautomer of the parent structure was found in the database.

Submission Correction Page

- Manual correction of structure or data
- System-, quality- and stereo-checkers/fixers
- Batch registration using the fixers

BULK LIST

Register all

N4227-5-6- (31)

N1255-3-7 (33)

N2401-4-2 (35)

N2425-2-7 (38)

N3692-5-2 (40)

N9046-7-5 (42)

N3538-1-1 (43)

N2196-2-4 (44)

SUBMISSION EDITING

LnbRef: **N1255-3-7**

Molweight: **310.349**

CST: **BlueSlime**

SubmissionId: **33**

Formula: **C13H16N3O4S**

Restriction: **0**

Status: **UnknownCST** - A previously unknown CST was entered.

Register

Register with

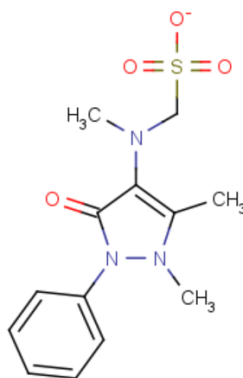
CN

[show possible hits](#)

[unlock submission](#)

[bulk restriction change](#)

Type: **Single**



BlueSlime

SALTS AND SOLVATES

+ add solvate + add salt

Parent Multiplicity 1

Salt components

Sodium 1

SWITCHES

Register New Lots

Register 2D Parent Matches

Register New Versions

Analyze Salt Solvate Fragments

Perform Quality Checks

GSK MDL Either DoubleBond
Checker

Register as Drawn

GSK EZ DoubleBond Checker

Register as Drawn

GSK Multiple Stereocentre Checker

Register as Drawn

GSK Marked Chiral With Chiral Flag
Checker

Register as Drawn

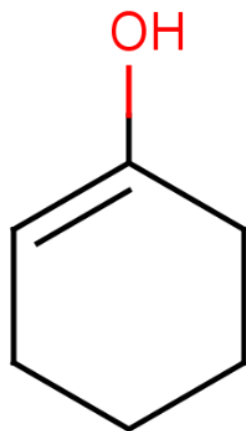
Match Service

- Exact, Metadata, 2D, Tautomer, 2D&Tautomer
- Possible hits (mock registration)
- Match list (unique, replace, accept options)

Matching components in the database

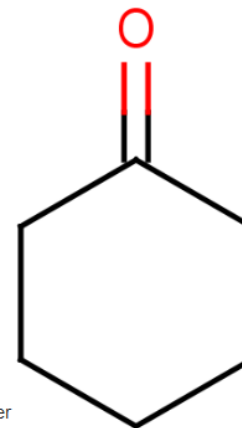
Component 1

Register as unique



1. Tautomer - GSK19

1 / 1



- PCN: GSK19
- Match Type: Tautomer

Accept

Replace



Amendment

- Amendment of registered compounds
 - Amendment on Parent/Version/Lot level
 - Structure, metadata, MW, salt/solvate info
 - Version correction according to business rules
 - Multicomponent compounds (number of components, structure, range/percentage)
 - Restriction level

Register new compound Staging area **Amendment** Administration Salt search Audit Search local version: 2696 Help Logout

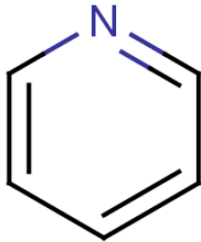
STRUCTURE BROWSER

STRUCTURE EDITING

Amend Delete Undelete Rename LnbRef History Change Restriction Level

MolWeight: 137.17926085903645 CST: -- Created on: 11-May-2012 11:57:0 Modified on: Formula: C5H5N.ClH.1.2(H2O)
Restriction: 0 Submitted by: Registered by: admin Modified by:

Type: Single



HCl

$[H_2O]_{1.2}$

SALTS AND SOLVATES

+ add solvate	+ add salt
Parent Multiplicity	1
Salt components	
Hydrochloride	1
Solvate components	
Water	1.2

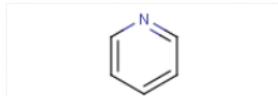
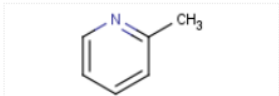
ChemAxon

Audit History

- Audit history of amendments
 - Initial registration
 - Any amendment
 - Deletion/undeletion

Audit

N8957-1-6

Details	Source Tree	Target Tree
<p># 2 Operation: Version Level Amendment Modified by: admin Date: Mon, 14 May 2012 09:28:09 GMT</p>	<p>GSK3 GSK3D N8957-1-6</p>  <p>GSK3D</p>	<p>GSK4 GSK4A N8957-1-6</p>  <p>GSK4A</p>
<p># 1 Operation: Initial Registration Submitted by: Registered by: admin Date: Fri, 11 May 2012 11:56:05 GMT</p>		

Features – Search Client

- Registry search client
 - Exact, 2D, tautomer
 - Metadata, MW
 - Export to file (under development)


Search

SEARCH OPTIONS

2D
 Tautomer
Search type: Substructure

SEARCH QUERY

Structure




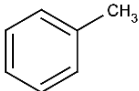
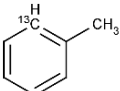

Salt Info

PCN CN LNBREF CST MW MF MatchType Action

Ignore Ignore Ignore Ignore Ignore Ignore

SEARCH RESULTS

displaying results: 1 - 5

		GSK1	GSK1A	N9127-9-1		78.1118	C6H6	Exact	Amend
		GSK2	GSK2A	N6036-4-4		92.1384	C7H8	Exact	Amend
		GSK2	GSK2B	N5954-5-5		93.1311	C6[13C]H8	Exact	Amend
		GSK3	GSK3A	N6470-3-7		79.0999	C5H5N	Exact	Amend

Demo

Register new compound

Staging area

Amendment

Administration

Salt search

Audit

Search

REGISTER STRUCTURE

LnbRef

N9710-3-9



LotId

717229



Source

ELNB

MolWeight

Chem. Sig. Text



Submitter ID

Additional Parameters

Param1:Value1, Param2:Value2

STRUCTURE EDITING

Register

Register with CN

Type Single

preview

File Edit View Insert Atom Bond Structure Calculations Tools Help



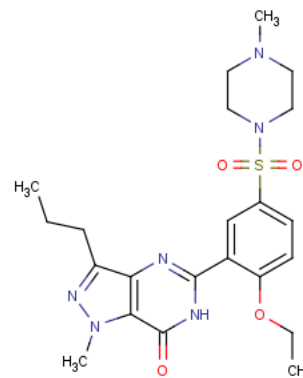
Registration successful!

Registration successful!

- PCN: GSK21
- CN: GSK21A
-
- LnbRef: N9710-3-9
- LotId: 717229

View this on Amendment

Close



export import

Future Plans

- Make all business logic related part easily configurable
 - Role-Customizable amendment and access
 - Customizable migration and bulk loader tools
 - GUI for the configuration of structure checkers and fixers
- Live feedback to the chemist about quality
 - Report on status of submissions
 - Ability to fix submission without the interaction of the registrar

Acknowledgement

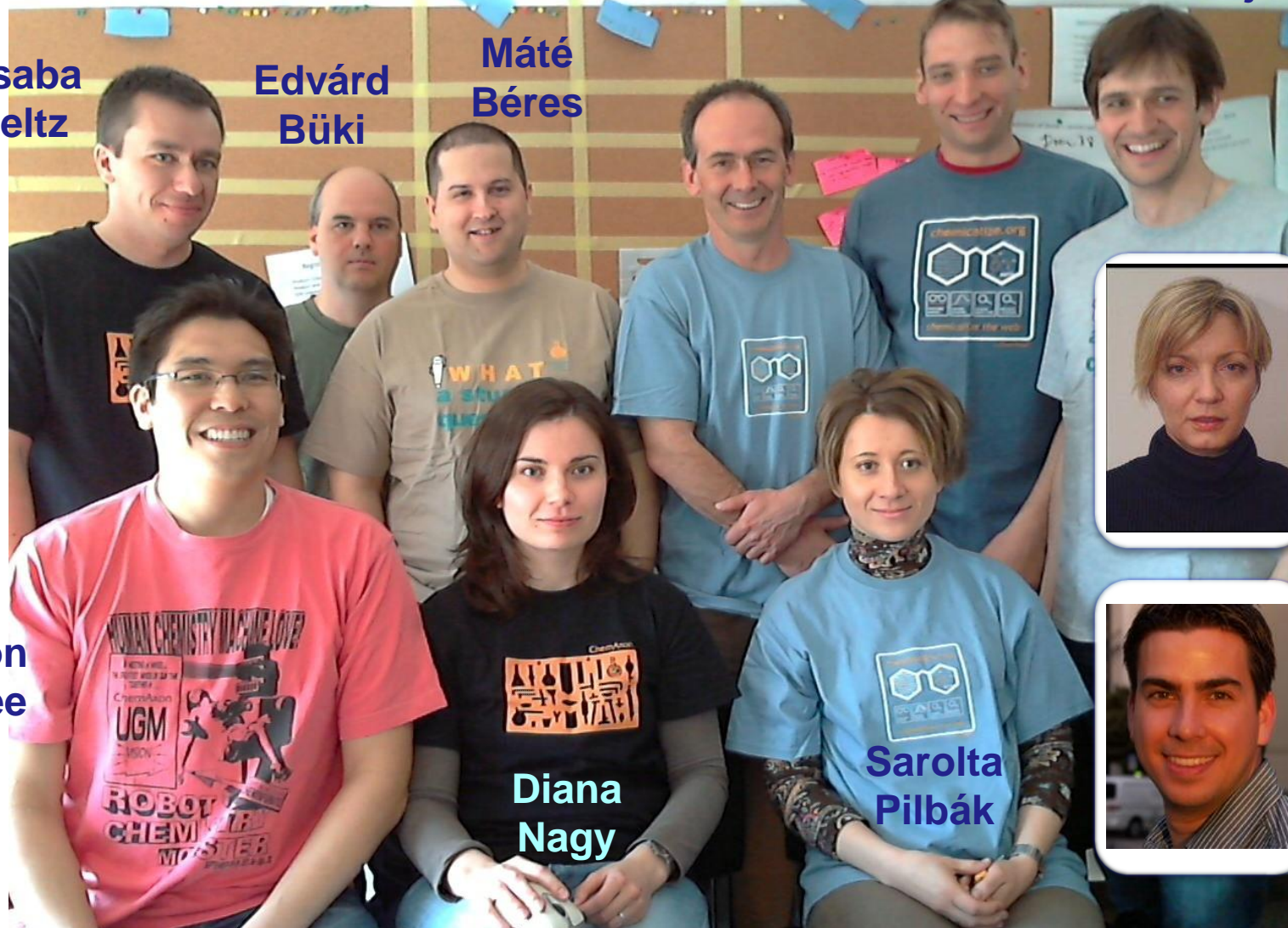
Roland
Molnár

Srdjan
Bejakovic

Csaba
Peltz

Edvárd
Büki

Máté
Béres



Jon
Lee

Diana
Nagy

Sarolta
Pilbák



Nikolett
Mihala



Attila
Tajti