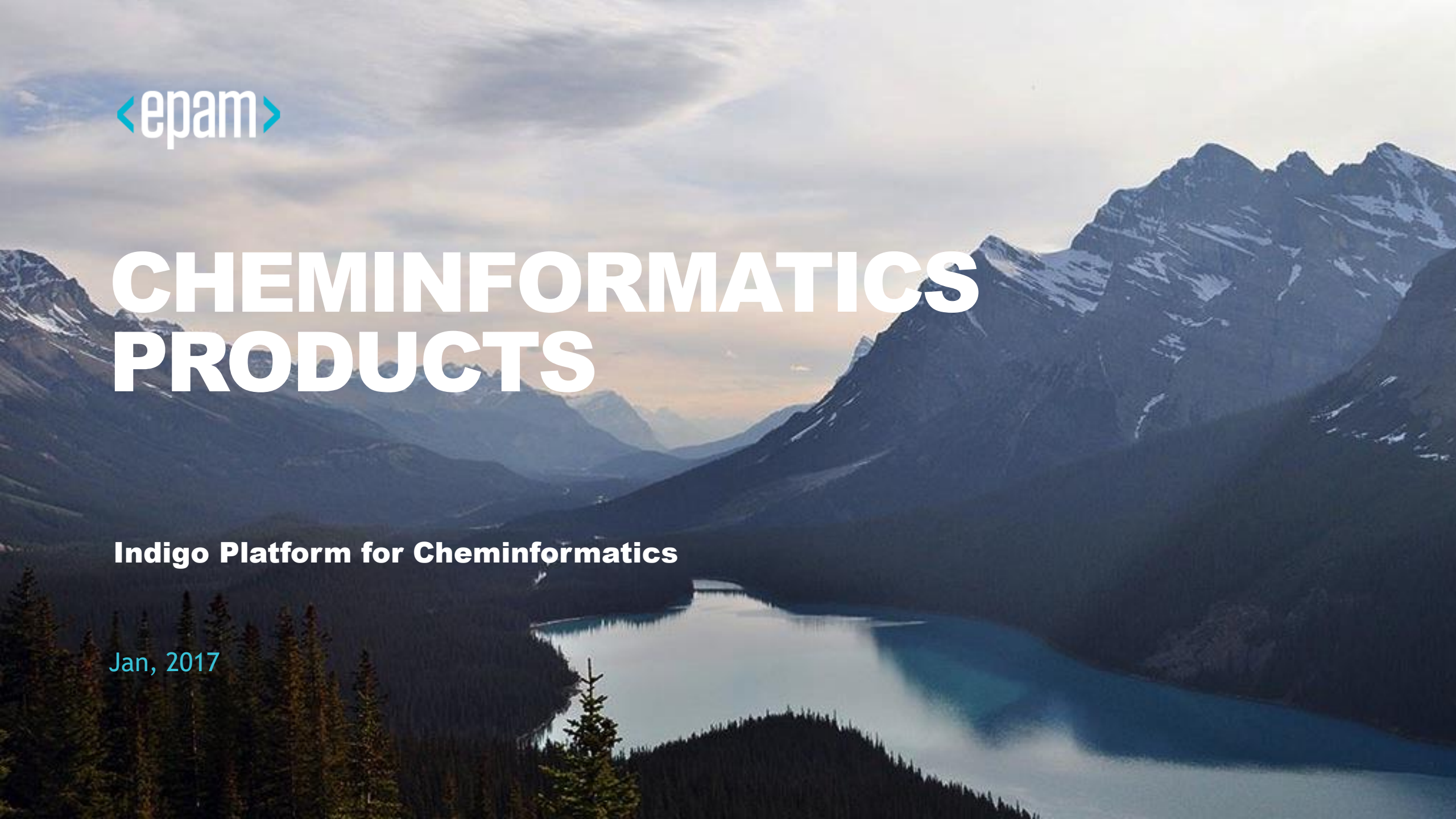


CHEMINFORMATICS PRODUCTS

Indigo Platform for Cheminformatics

Jan, 2017

INDIGO PLATFORM

Indigo Platform

Web (view)

KETCHER

Import chemistry	Export chemistry	Editing of chemical structures	Lightweight portable JavaScript

Middleware (controller)

INDIGO

• Import, export chemistry • Languages: Java, Python, C#	Transform chemistry: • API • Reactions	Calculate: • Properties • Layout • AAM • Scaffold	Render: • SVG • PNG • JPG • Grid

IMAGO

Convert images to chemistry:
- PNG
- JPG
- PDF

Storage (model)

BINGO

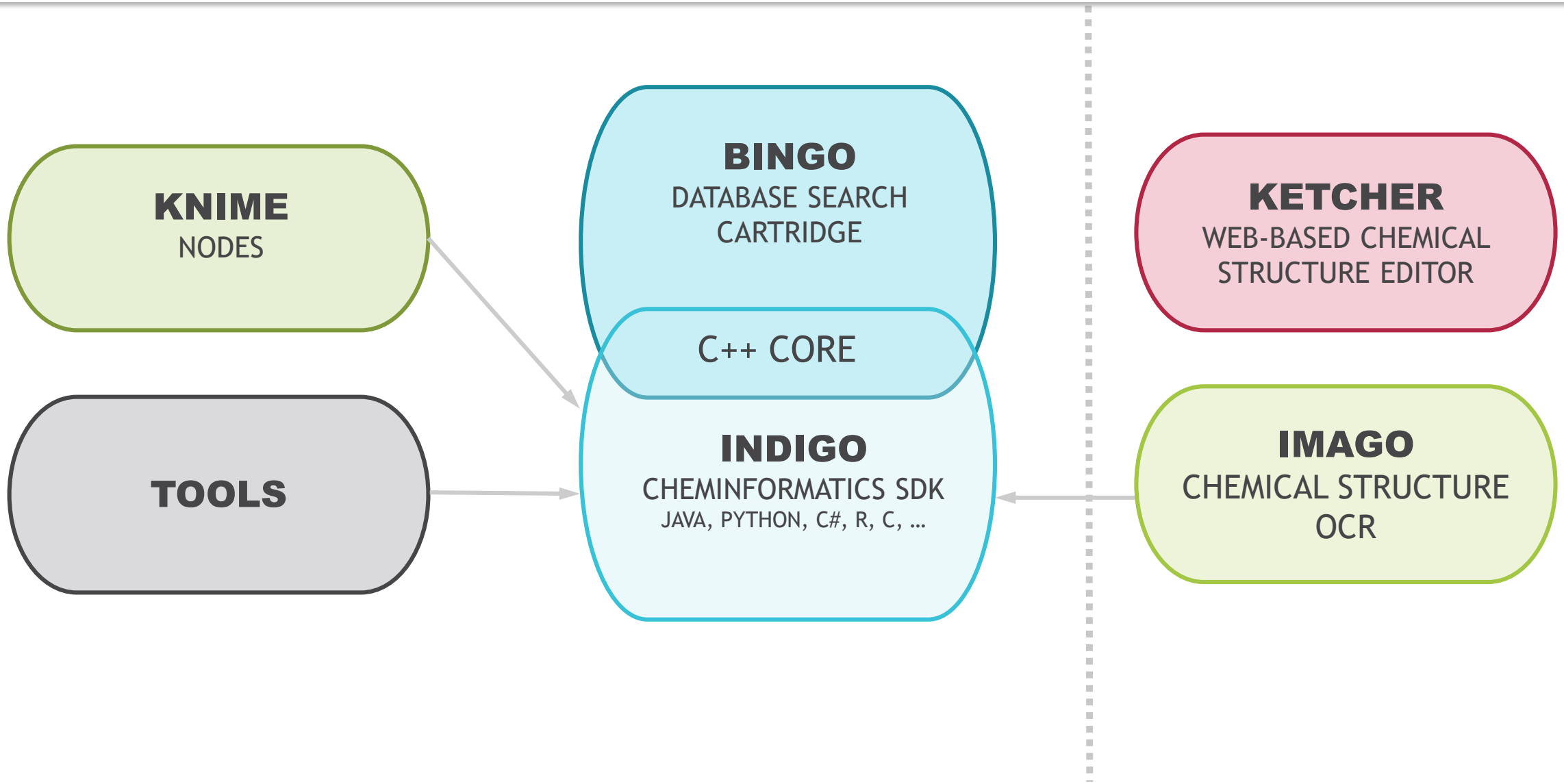
Import, export chemistry	RDBMS: • Oracle • PostgreSQL • Microsoft SQL Server	Search: • Sub • Exact • Similarity	Utility Properties: • SMILES • Aromatic • AAM

BINGO NO-SQL

Storages:
- File-based
- Lucene

Bingo and Indigo utilities

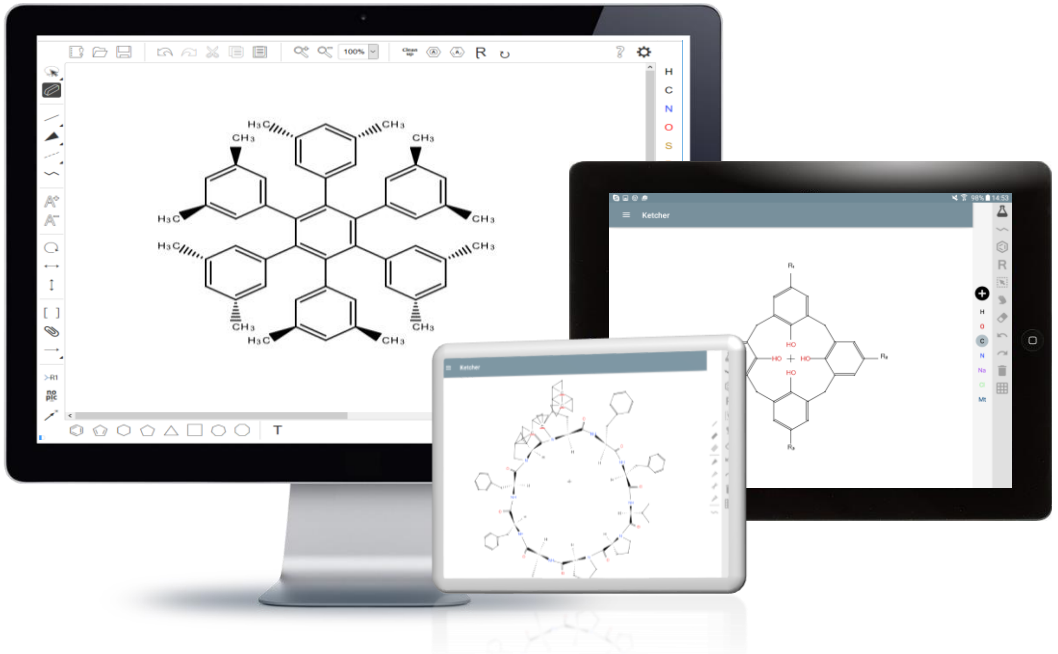
PRODUCTS OVERVIEW



KETCHER
WEB-BASED CHEMICAL STRUCTURE TOOL

SUMMARY

- EPAM developed an open-source web-based chemical structure editor incorporating high performance, good portability, light weight, and ability to easily integrate into a custom web-application.



KEY HIGHLIGHTS

- Use of SVG to achieve best quality in-browser chemical structure rendering
- Ability to load and save structures in MDL/Biovia molecular file format, InChI, ChemAxon Extended SMILES, ChemAxon Extended CML file formats
- Fast 2D structure representation that satisfies common chemical drawing standards
- 3D structure visualization
- OCR - ability to recognize structures at pictures (image files) and reproduce them
- Complete stereochemistry support during editing, loading, and saving chemical structures
- Storing full history of actions, with the ability to rollback to any previous state
- Direct input of atom label and charge
- Languages: JavaScript with third-party libraries

INDIGO CHEMINFORMATICS

SDK

- Indigo SDK is highly configurable and extensible and allows creating C/C++/C#/Java/Python plug-ins for it
- Excellent portability. Supported platforms: Windows/Linux/Mac - 32-bit and 64-bit
- Original Algorithms

ORIGINAL ALGORITHMS

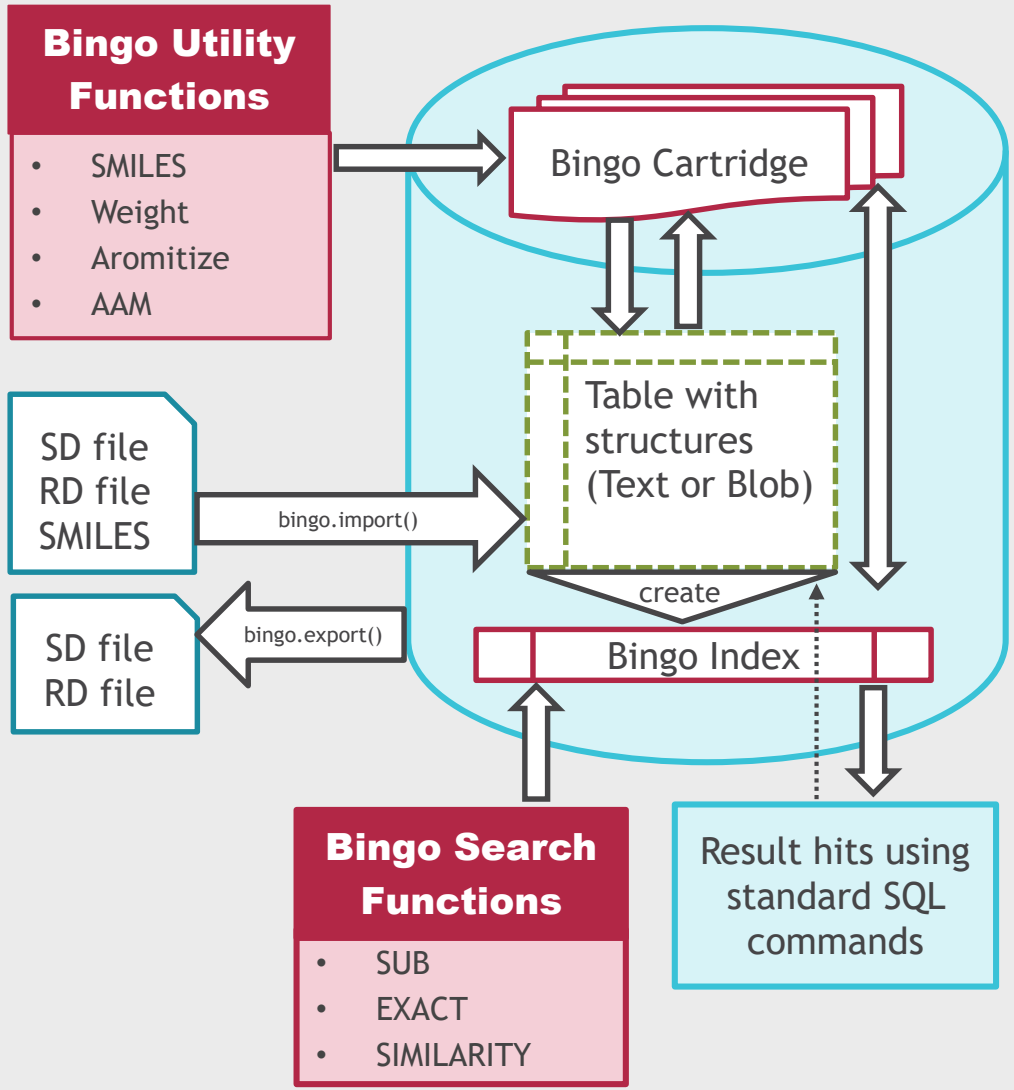
- Substructure matching (subgraph isomorphism)
- Aromaticity matching and de-aromatization
- Tautomer (sub)structure matching
- Chemical resonance (sub)structure matching
- Canonical SMILES (with stereochemistry)
- 2D molecule and reaction layout
- Reaction atom-to-atom mapping (AAM), based on maximum common substructure (MCS)
- Molecule and reaction fingerprints, based on sub-tree enumeration tree canonical codes

FEATURES

- Input formats support: Molfiles/Rxnfiles v2000 and v3000, SDF, RDF, SMILES, SMARTS
- Molecule and reaction rendering. Best picture quality among all available products. Easy SVG support
- Automatic layout for SMILES-represented molecules and reactions
- Canonical (isomeric) SMILES computation
- Exact matching, substructure matching, SMARTS matching. Highlighting support
- Matching of tautomers and resonance structures
- Molecule fingerprinting, molecule similarity computation
- Molecular weight, molecular formula computation
- R-Group deconvolution and scaffold detection

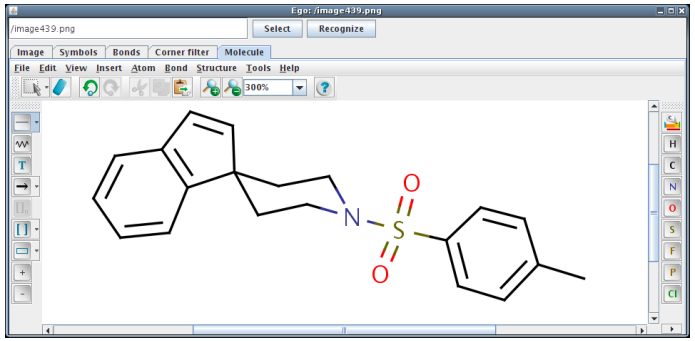
BINGO
DATABASE SEARCH CARTRIDGE

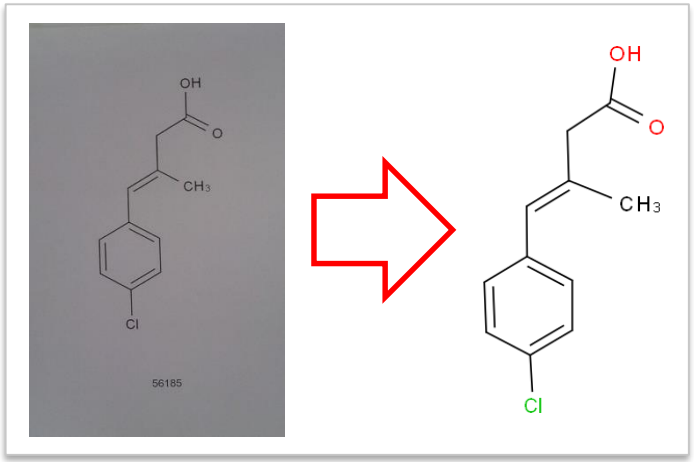
- Open-source chemistry cartridge for Oracle, SQL Server, PostgreSQL RDBMS.
- Store, index, and search molecules and reactions with ordinary SQL commands.
- Powerful cheminformatics functionality combined with all the comfort of relational database.
- Supports the most requested features for molecule and reaction databases: exact/substructure/similarity search, format conversion, canonical SMILES, reaction AAM.
- Performance, scalability, portability, and unique chemical features
- Supported platforms: Windows/Linux/Mac; 32-bit and 64-bit
- Supported database systems:
  - Oracle
  - Microsoft SQL Server
  - PostgreSQL
  - NoSql (Lucene SOLR, Sinequa)
- No problems with large databases like PubChem (> 40M entries)



IMAGO
CHEMICAL STRUCTURE OCR LIBRARY

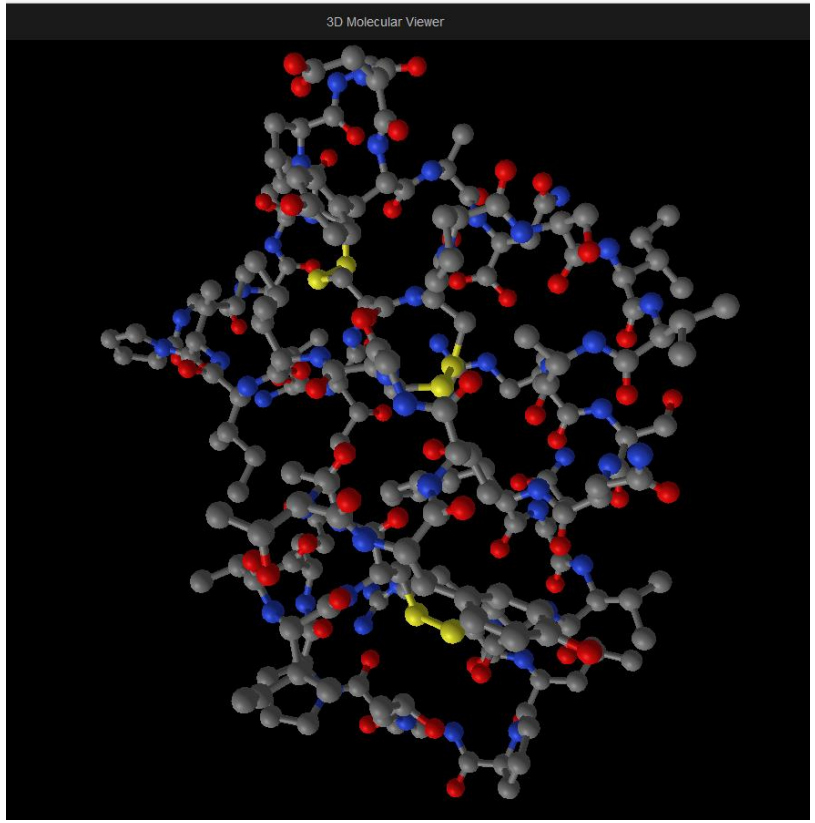
- Support for PNG files for input, each file containing one molecule
- Support for MDL/Symyx Molfile for output
- Recognized molecular elements: Single, double, triple bonds, Atom labels, subscripts, isotopes, charges, Superatoms, Aromatic rings, Touching characters
- Stereochemistry (up- and down-bonds)
- Java GUI and command-line utility
- OpenCV, C/C++
- Patent processing
- Mobile input via camera photos
- Papers indexing for molecule search
- Exposing C interface to applications
- Java wrapper available for all supported platforms

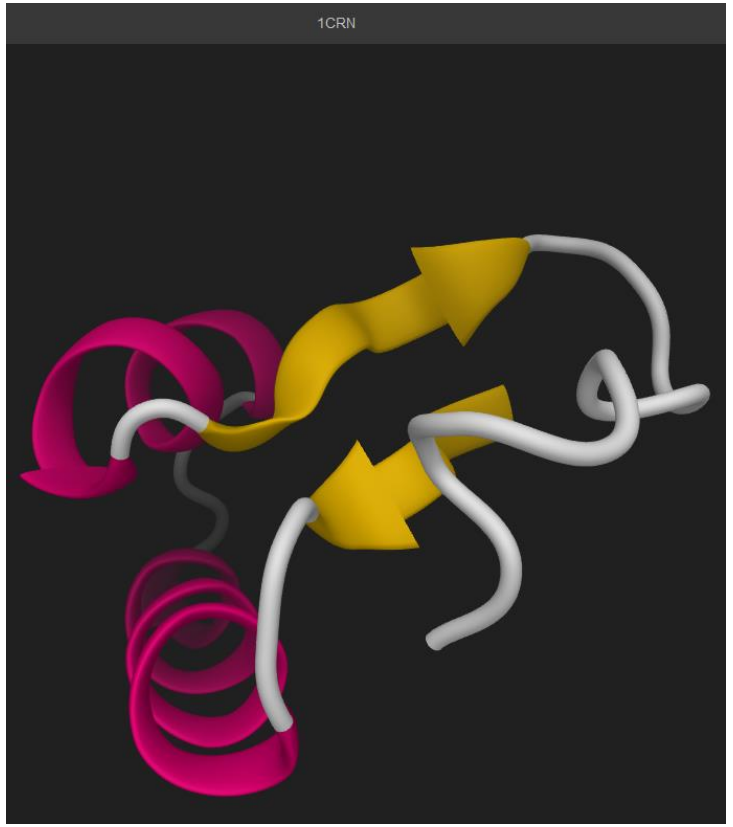




INDIGO PLATFORM PRESENT/FUTURE

- Ketcher 3D viewer integration
- Imago OCR algorithms and improvements
- Bingo and Indigo improvements
- Make it faster
- Make it more functional
- Indigo service





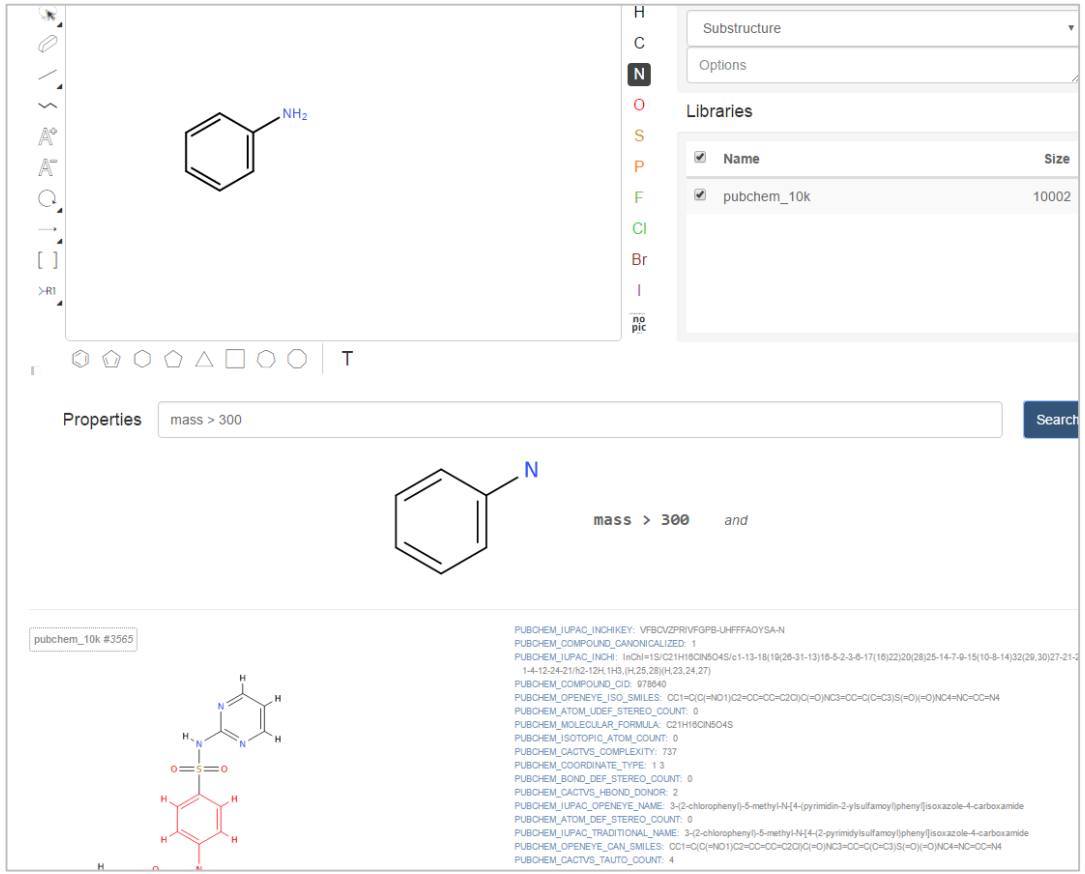
INDIGO PLATFORM PRESENT/FUTURE

INDIGO SERVICE

- Fast insert (about 10k per second)
- Google-like search
- Support users/projects
- Easy deploy REST (Docker)
- New technologies and features…

INNOVATIONS

- RESTful service (Python + Flask+Indigo)
- Easy deployment (Docker)
- Search chemistry (Bingo)
- Search properties (string and numbers)
- Fast upload* (>5000 structures/sec)
- Modern interface
- Etc.



OPEN SOURCE

- About 3 000 downloads in 2016
- 847 downloads by Industry
- Downloads from more than 40 countries
- Downloads from more than 250 companies
  Merck, Novartis, Lilly, BASF SE, Bayer, Royal Society of Chemistry, ChemBridge Corporation, IBM, Intel, other
- Downloads from more than 400 universities

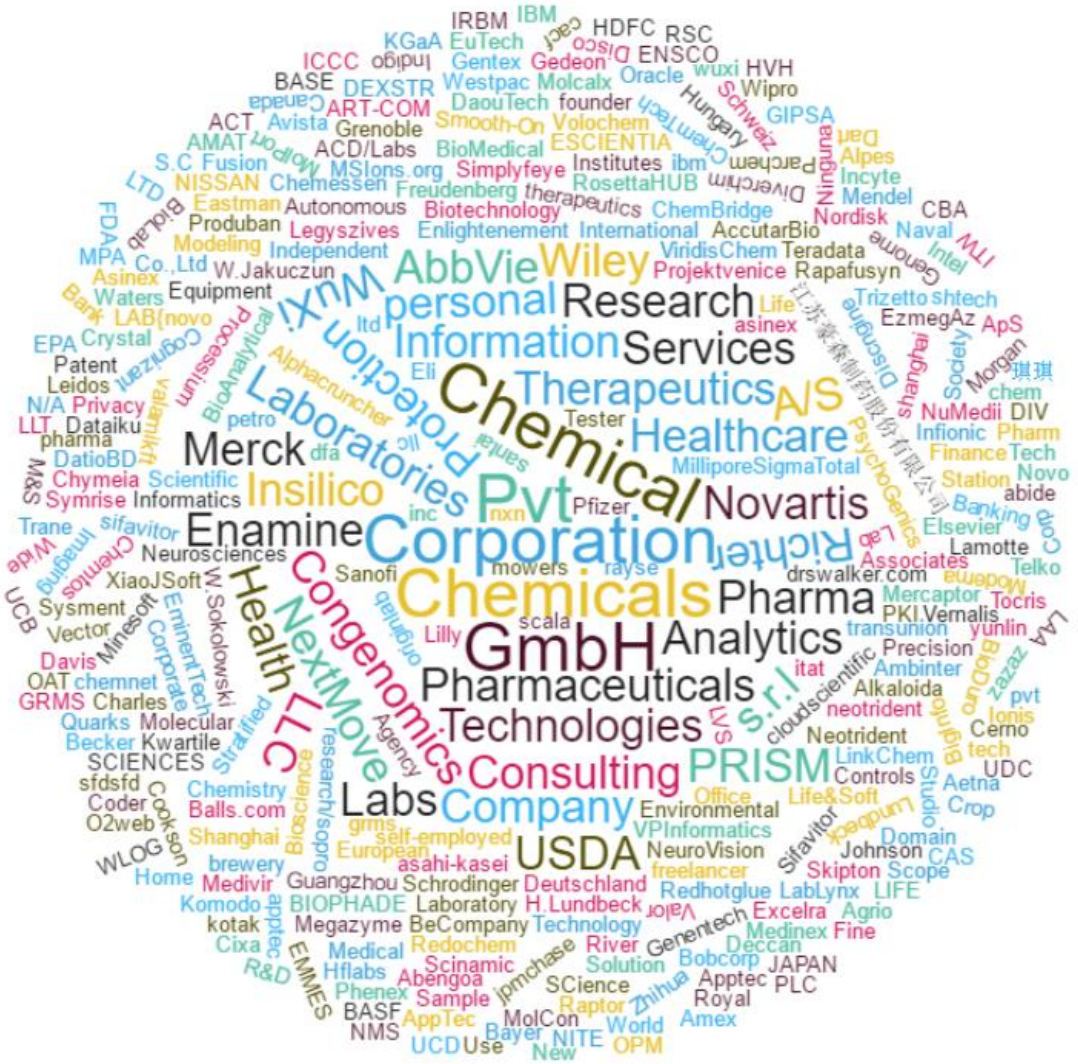
Product \ Position		Business/Industry	C-Level	Manager	Other	Student/Academic	Totals
Bingo		166			1	204	**371**
ChemDiff	4	10				55	**69**
Imago		95				247	**342**
Indigo	7	209		1	1	753	**971**
Indigo ELN	7	177				482	**666**
Ketcher	5	78	1		1	211	**296**
Legio	3	7				28	**38**
Parso	4	105			1	95	**205**
Totals	**30**	**847**	**1**	**1**	**4**	**2,075**	**2,958**

BUSINESS DOWNLOADS

All around the world

Country	Totals
USA	199
India	154
China	128
France	53
UK	26
Germany	23
Japan	22
Switzerland	21
Spain	20
Denmark	17
Australia	16
Hungary	16
Italy	14
Korea, South	12
Anguilla	9
Canada	9

From more than 250 companies



INDIGO PLATFORM BUSINESS

Celgene

- Use indigo service as a intermediate storage
- RESTful service tests
- Prepare and provide functional requirements using current existing service

Merck

- Biological registration system
- Implement and perform SCSR transformation using Indigo as a core environment

Elsevier

- Cheminformatics registration system
- Use Ketcher web editor tool

INDIGO PLATFORM COMMERCIAL LICENSES

- Indigo Toolkit – license granted to OriginLab Corporation (USA)
- Indigo Toolkit – license enquiry from U.S. Naval Research Laboratory, but falls under GNU license
- Ketcher – license granted to InSilico Co. Ltd (Korea)
- For small clients current arrangement (2016) is to issue free licence for commercial use/redistribution

KETCHER USAGE

- ChemSpider is a free chemical structure database providing fast access to over 58 million structures, properties, and associated information
-
- WITHDRAWN - database of withdrawn and discontinued drugs that were pulled out of global markets due to safety concerns
- Colorado State University is a public research university located in Fort Collins, in the U.S. state of Colorado. The university is the state's land grant university, and the flagship university of the Colorado State University System.
- Reaxys is a unique web-based chemistry database consisting of deeply excerpted compounds and related factual properties, reaction and synthesis information as well as bibliographic data, navigated and displayed via an actionable interface.
- Fully searchable drug approval documents and extracted data to inform critical drug development decisions.
- Search and Inquire chemicals online

KETCHER PUBLICATIONS

- The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic info
- PubMed Central® (PMC) is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM)
- *Journal of Cheminformatics* is an open access journal publishing original peer-reviewed research in all aspects of cheminformatics and molecular modelling.
- ReadCube was started by a researcher and a computer scientist to address the challenges faced by scientists. What started in a Harvard College dorm room as a tool to help organize and find scientific papers quickly turned into something rather more.
- OMICtools is the first community search platform offering an intuitive search experience to retrieve biological data analysis tools. Connect with users to share and discuss about your tools. Bridge the gap between life science and computational biology.