Responding to the Challenge Posed by the Generic Control of Substances

Nikolett Mihala
Ádám Villei
Balázs Zaicsek
István Őri
Péter Kovács
Róbert Wagner
Henrik Miszoglád
Ákos Papp[1]

[1] Ákos Papp has been the Product Manager of Compliance Checker since its launch in 2016.
If you have any questions or comments please contact apapp@chemaxon.com

Chemaxon

Abstract

Drug monitoring organizations report that new psychoactive substances continue to emerge, posing health and social threats and creating an unprecedented challenge for national and international law systems. By the time a new regulation comes into force, there are already novel drugs ready to be launched into the market. The number and the chemical diversity of new psychoactive substances have increased at a speed that drug control laws often struggle to match. To keep pace with this ever-evolving situation, countries have introduced new types of legal responses such as generic control of substances. Screening large compound collections against these new generic definitions is a challenge for traditional search engines. New approaches like the introduction of novel homology groups and state-of-the-art Markush search technologies are emerging to address these challenges, and this paper gives examples of the successful application of such new technology.

Drug Control Legislation – a Brief History

The 1912 Hague **Opium Convention** was the very first international drug control treaty. The agreement signed by 12 countries aimed to control the excesses of an unregulated free trade regime, primarily regarding opium. A later series of conventions added cocaine and cannabis while remaining more regulatory than prohibitive.

The 1961 United Nation **Single Convention on Narcotic Drugs**[2] **(Yellow List)**[3], which replaced the previous international treaties, was designed to establish internationally applicable control measures with the aim of ensuring that psychoactive substances are available for medical and scientific purposes, while preventing them from being diverted into illegal channels.

Controlled substances are assigned to one of four schedules depending on their medical applicability, addictive potential, and being subjected to abuse. Originally, it was signed by 64 countries, and as a result of synthetic stimulant abuse, it was amended by the 1971 **Convention on Psychotropic Substances (Green List)**[4]; and the 1988 **United Nations Convention against Illicit Traffic in Narcotic Drugs and Psychotropic Substances (Red List)**[5] which classifies drug precursors – forming together the international legal framework of the global drug control regime. Today it has 186 signatories. In addition to psychotropic substances, chemical weapons and hazardous chemicals are also controlled in most countries too.

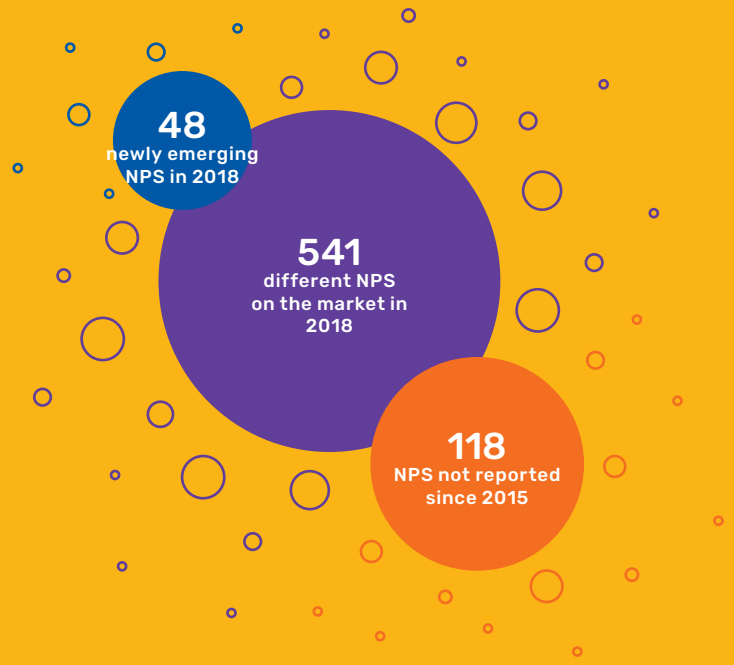
- Montreal Protocol on Substances that Deplete the Ozone Layer[6] (1987)
- Chemical Weapons Convention[7] (1992)
- Rotterdam Convention on the Prior Informed Consent Procedure for Certain Hazardous Chemicals and Pesticides in International Trade[8] (1998)

These lists collect specific named compounds, but even the earliest Yellow List contains generalization clauses to include isomers, esters, and ethers, as well as salts of the drugs including salts of esters, ethers, and isomers.

[2] https://www.unodc.org/unodc/en/treaties/single-convention.html
[3] https://www.incb.org/documents/Narcotic-Drugs/Yellow_List/58th_Edition/Yellow_List_-ENG.pdf
[4] https://www.unodc.org/pdf/convention_1971_en.pdf
[5] https://www.unodc.org/unodc/en/treaties/illicit-trafficking.html
[6] https://ozone.unep.org/treaties/montreal-protocol
[7] https://www.opcw.org/chemical-weapons-convention
[8] http://www.pic.int/

Controlled Substances in R&D

In recent decades drug markets have become increasingly complex. Well known plant-based substances such as cannabis, cocaine, and heroin have been joined by hundreds of synthetic drugs, many of which are not under international control. Narcotics, hallucinogens, stimulants, depressants, and anabolic steroids are the most typical compound families. Currently, most of those are stimulants, followed by synthetic cannabinoid receptor agonists, and a smaller number of opioids. We have also witnessed a rapid rise in the non-medical use of pharmaceutical drugs. According to the UNO[9] data, roughly 500 new psychoactive substances (NPS)[10] are found on the national markets of Member States each year. Illicit drug manufacturers need to seek alternatives, i.e. compounds specifically designed to circumvent controls.



[9] UN World Drug Report 2020 https://wdr.unodc.org/wdr2020/index.html

[10] New psychoactive substances (*NPS is a new narcotic or psychotropic drug, in pure form or in preparation, that is not controlled by the 1961 United Nations Single Convention on Narcotic Drugs or the 1971 United Nations Convention on Psychotropic Substances, but which may pose a public health threat comparable to that posed by substances listed in these conventions (Council Decision 2005/387/JHA))

To mitigate the impact of this explosion in the number of eventually-controlled substances (mostly NPS), some countries have extended the coverage of existing drug laws by defining groups of substances, rather than listing individual drugs as had been done previously.

This method – adopting a group definitions approach – was initiated by Germany, and in 2014 Belgium established the legal basis to implement it. However, the Netherlands rejected it in 2012 because of the complexity of targeting some substances that may have valid use.[11]

Policymakers need to formulate a generic definition which effectively captures harmful substances to prevent their illegal use, but at the same time doesn't interfere with compounds which might have a research interest. Generic group definitions have been used for years in Ireland and the United Kingdom, and they were introduced into the drug laws of other countries, including Denmark, France, and Italy. Many countries published a list of exemptions to the generic definition to ensure that molecules with pharmaceutical or other use would not have become controlled. General exemptions that may help research (e.g. maximum 1 mg total in a preparation) also exist at several regulations, but these limits are usually accompanied by several restrictions that are difficult to fulfill or prove (e.g. the compound "cannot be recovered by readily applicable means"), so identification and special care of these exempt compounds are still required.

Organizations handling large collections of molecules (i.e. either real or virtual compound libraries) need robust cheminformatics tools to identify controlled substances. It is especially important to have a high performance solution when generic definitions are translated into a digital representation, typically in the form of Markush structures.

Markush structures are generic descriptions created for specifying collections of chemically related compounds. Their very first application was by Eugene A. Markush, who claimed generic chemical structures in applying for a patent for pyrazolone dyes in 1924 to the U.S. Patent Office.[12] Since then, they have gained widespread use in chemical patents and other fields. The invariable part of the structure – the scaffold – includes the common structural features of the collection.

[11] Legal approaches to controlling new psychoactive substances (Perspectives on drugs) EMCDDA https://www.emcdda.europa.eu/topics/pods/controlling-new-psychoactive-substances_en

[12] https://worldwide.espacenet.com/publicationDetails/biblio?CC=US&NR=1506316&KC=&FT=E&locale=en_EP

The variable parts can be described by:

- Substituent variation – listing a set of different substituents at a position
- Position variation – different attachment points/position for a substituent
- Frequency variation – allowing substituents to occur multiple times in a chain or a part of a ring
- Homology groups – general nomenclatural expressions covering a large or theoretically infinite number of substructures with a common structural feature like “aryl”

Figure 1. depicts some parts of a possible query covering the fentanyl derivatives of the *Italian Decreto del Presidente della Repubblica del 9 ottobre 1990, n. 309 TABELLA I.*[13] and the fentanyl molecule as a hit.

As a result of the structural/chemical diversity, a single Markush structure might cover a huge chemical space possibly representing an unlimited number of molecules.

Some countries – particularly Belgium – draw explicit Markush structures as definitions, thus making digital transformation easier. In some cases the legislators opt to create regulations where digitalization results in Markush structures covering 10^{30} or more molecules. Some examples of these “giant” Markush structures are for the control of synthetic cannabinoids in the UK, and for fentanyl derivatives in Belgium, Italy, and the UK.

Efficient search against Markush structures requires advanced methods that go well beyond simple structure-based search. The complex Markush structures resulting from generic definitions present a big challenge for search engines. With the introduction of generic group definitions (homology groups) searching speed has been significantly increased. In addition, a new query technology has been also invented, which is optimized to process enormous Markush structures in substructure search scenarios at high speed.

[13] https://www.normattiva.it/eli/stato/DECRETO/2020/06/30/0/ORIGINAL

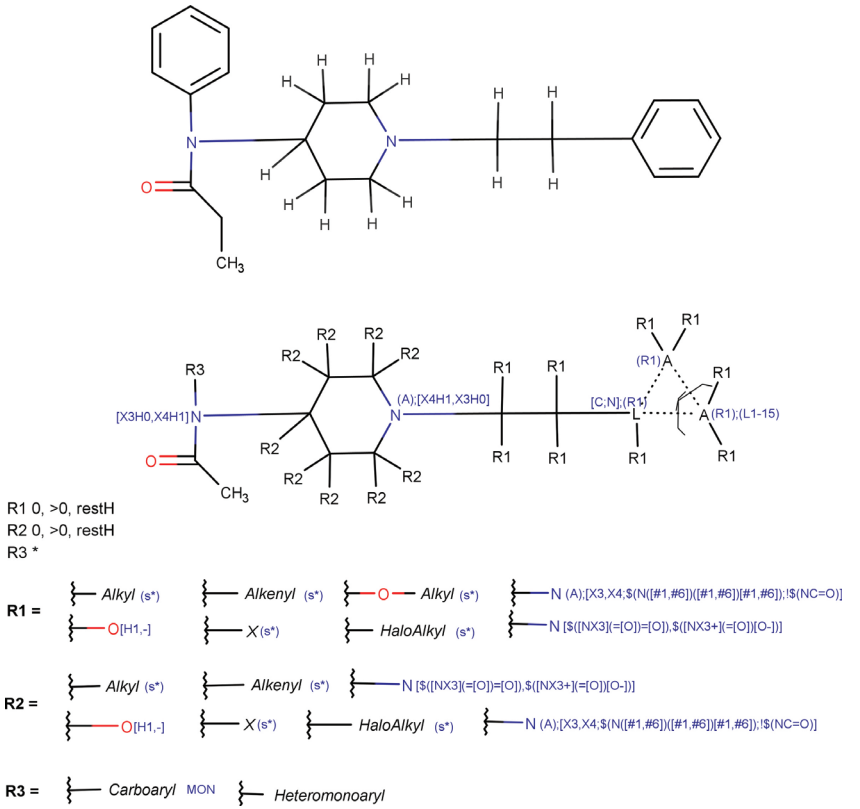


Figure 1. A Markush query for fentanyl derivatives

Stepwise search of all the enumerated structures represented by a Markush structure isn’t a suitable solution when the Markush structure covers a huge chemical space, even if this space is finite. Our algorithm handles Markush structures without fully enumerating the Markush space.

Search time is significantly reduced by two optimizations:

- Substituent variation – listing a set of different substituents at a position
- Position variation – different attachment points/position for a substituent

Recently, some generic definitions have come into force that are impossible to represent by traditional Markush structures. To address this, a cheminformatics toolbox with scripting capability plays a critical role in ensuring the required accuracy.

Identify Controlled Substances in Compound Collections

Under the auspices of the Controlled Substance Compliance Expert Community of the Pistoia Alliance, Chemaxon in cooperation with Patcore has created a cutting-edge checking system to screen chemical structures against the controlling legislations. This system – Compliance Checker – is a combined software and content package providing a flexible tool for addition to compound compliance workflows. Chemaxon works with a third party that monitors the drug control legislation of the 18 jurisdictions that are covered by Compliance Checker, and as new or amended legislation is released, the database is updated accordingly to keep the system current.

Case Study

To showcase how Compliance Checker performs in extreme conditions and to gain a deeper understanding of the differences between controlled substances in different countries, we conducted a pilot study on MolPort's[14] 45.8 M molecule database, containing mostly small/drug-like molecules. From the available configurations, we selected the most robust Compliance Checker Premium subscription version. Checks were performed in 500K batches, using csv files containing company IDs and SMILES against all the 18 supported countries (Austria, Belgium, Canada, China, Denmark, France, Germany, India, Ireland, Italy, Japan, Netherlands, Singapore, Spain, Sweden, Switzerland, UK, USA).

Pistoia Alliance[15] is a global, not-for-profit alliance of life science companies, vendors, publishers, and academic groups that work through pre-competitive collaboration to lower barriers to innovation in R&D. Its Controlled Substance Expert Community[16] provides a forum to share best practice in controlled substance compliance and explore and compare the differing legislations worldwide.

Pistoia Alliance

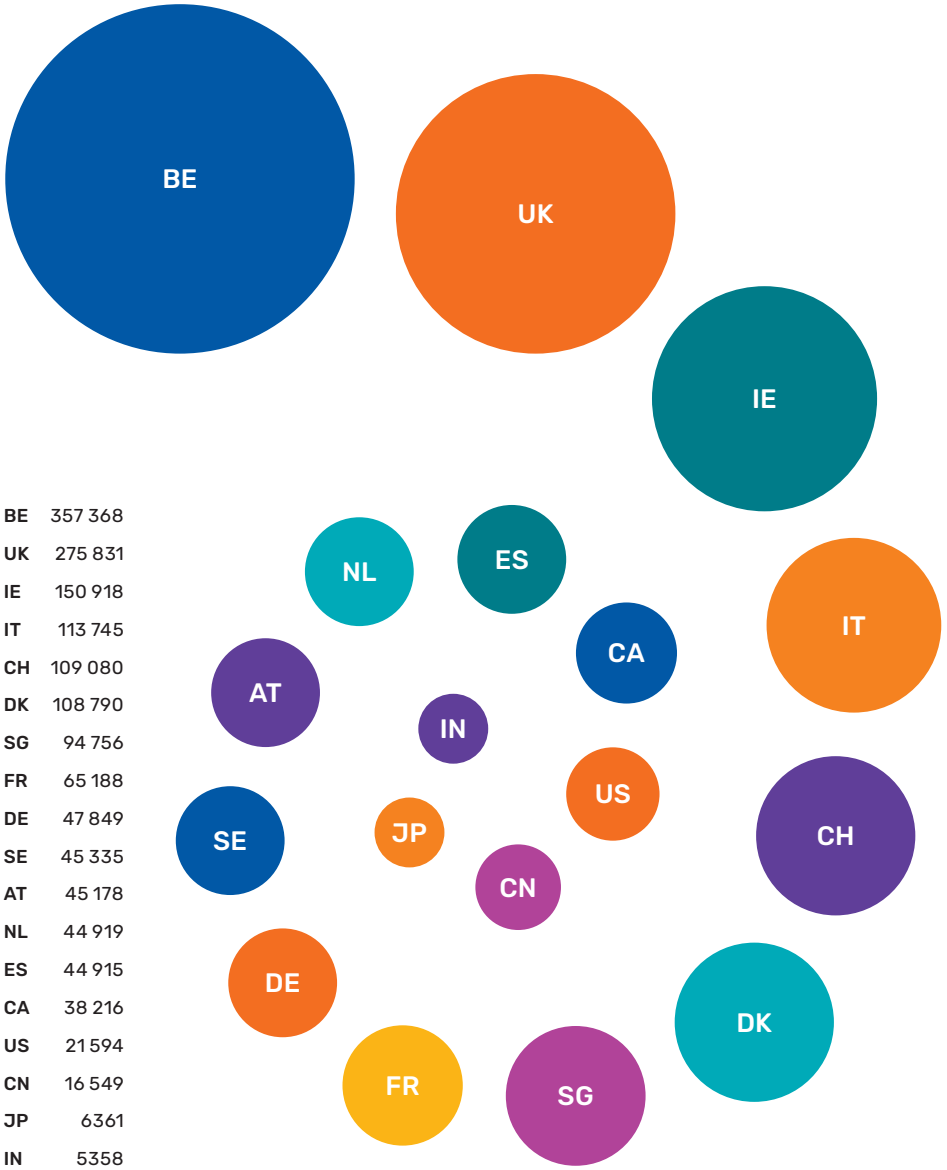
[14] https://www.molport.com
[15] https://www.pistoiaalliance.org/
[16] https://www.pistoiaalliance.org/projects/current-projects/controlled-substance-compliance/

Results

Number of unique molecules submitted:	45 867 797
Number of controlled substances detected:	1 259 803
Hit rate:	2.7%
Number of unique controlled substances detected:	686 815
Number of molecules controlled by a single regulation:	506 248
Number of molecules controlled in a single country:	478 142
Number of molecules controlled in all the countries screened:	5358
The highest number of regulations controlling a single molecule:	28

Our screening showed that 2.7% of the library is controlled in at least one of the checked countries. The record-holding molecule is codeine, which is controlled by no less than 28 different regulations, followed by other morphine analogues like heroin and the structurally non-related phytocannabinoid, THC. Interestingly, not a single common controlled molecule can be found in the screened countries beyond those 5,358 controlled by the international legislations. The international and the EU regulations control 44,802 molecules overall.

The figure below visualizes the number of hits with the corresponding country.



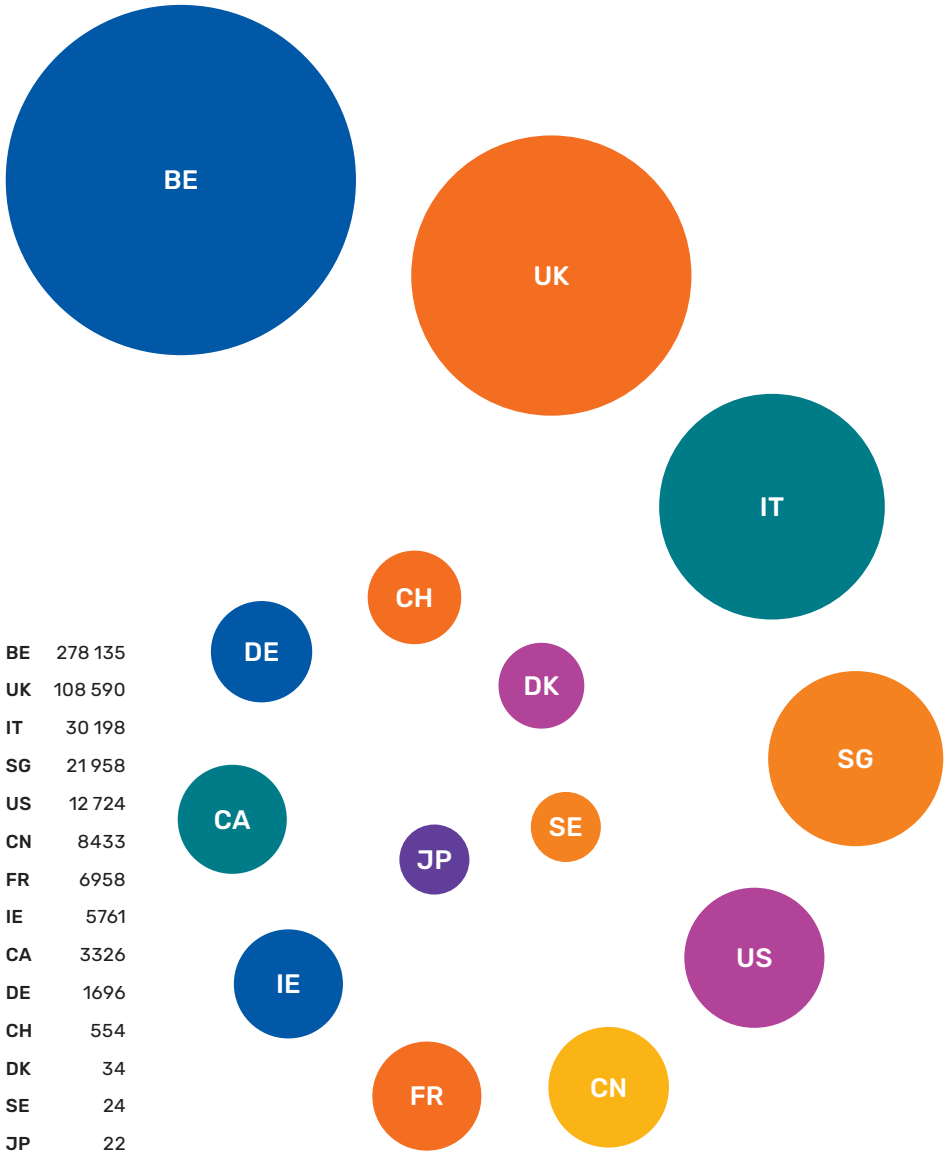
BE	357 368
UK	275 831
IE	150 918
IT	113 745
CH	109 080
DK	108 790
SG	94 756
FR	65 188
DE	47 849
SE	45 335
AT	45 178
NL	44 919
ES	44 915
CA	38 216
US	21 594
CN	16 549
JP	6361
IN	5358

Table 1. Number of controlled substances by country

A very similar list summarizes the regulations which control the highest number of substances.

Regulation	Hit
BE Arrête royal reglementant les substances stupefiantes, psychotropes et soporifiques	312829
UK The Misuse of Drugs Regulations 2001 Schedule I	192032
UK Third Generation Synthetic Cannabinoids	110450
IE MISUSE OF DRUGS REGULATIONS 1988 Schedule I	105489
CH Swiss Controlled Substances Act (BetmVV-EDI) Narcotics List E	102073
SG Misuse of Drugs Act FIRST SCHEDULE Controlled Drugs	90131
IT Decreto del Presidente della Repubblica del 9 ottobre 1990, n. 309 TABELLA I	69470
UK Misuse of Drugs Act Part I Class A Drugs	69143
DK Bekendtgorelse om euforiserende stoffer Liste B	64696

Correlated figure representing the countries vs. number of exclusively controlled molecules



BE	278 135
UK	108 590
IT	30 198
SG	21 958
US	12 724
CN	8433
FR	6958
IE	5761
CA	3326
DE	1696
CH	554
DK	34
SE	24
JP	22

Table 2. Countries vs. number of exclusively controlled substances

And what is behind the numbers?

Generic definitions, indeed. It is beyond question that this plethora of molecules cannot be described without using generic structures.

The regulation covering the highest number of molecules is the Belgian Arrête royal reglementant les substances stupefiantes, psychotropes et soporifiques[17]. It contains six generic definitions, for the derivatives of amphetamine, cathinone, fentanyl, cannabinoid, tryptamine, and piperazines. (164,503 cannabinoid; 127,029 piperazine; 18,382 amfetamine; 2,267 cathione; 305 fentanyl; 64 tryptamine were identified.)

However, if we consider the covered compound families, the absolute winner is the UK Third Generation Synthetic Cannabinoids, which aims to control (as in its name) synthetic cannabinoids exclusively.[18] After the introduction of its previous version in 2016 the pharma industry raised concerns that the generic descriptions were capturing compounds with interest in scientific research which should not be controlled. They noted that due to the broad nature of the generic definition, some compounds were unintentionally captured and that inadvertently subjected them to the strictest level of control. Research organisations in industry and academia were required to obtain a special licence to conduct research on such compounds. To ensure the availability and access to controlled substances exclusively for medical and scientific purposes, while preventing their diversion, a new regulation was formulated with a reduced scope. Before the amendment came into force in November 2019, in this compound library the regulation (Third Generation Cannabinoids[19]) would have controlled 237,968 molecules, while the new amendment halved the number of former hits (110,450).

There is a massive explosion in the number of controlled molecules caused by the introduction of new generic regulations in Italy and Singapore, too[20]. In Singapore, due to the amendment[21] of the Misuse of Drugs Act, identified controlled substances rose from 52,864 to 90,131.

[17] http://www.ejustice.just.fgov.be/mopdf/2017/09/26_1.pdf
[18] https://www.legislation.gov.uk/uksi/2019/1323/made/data.pdf
[19] https://www.legislation.gov.uk/uksi/2016/1124/pdfs/uksi_20161124_en.pdf
[20] https://sso.agc.gov.sg/Act/MDA1973#Sc1-
[21] https://sso.agc.gov.sg/SL-Supp/S751-2020/Published/20200903?DocDate=20200903

Performance

The check of 45.8 million molecules in 500K batches using File Check takes about 2 days using the Compliance Checker's highest performing Software as a Service (SaaS) solution, the Premium Subscription. (This subscription includes four services that execute checks and it autoscales to eight in cases of heavy load. One service instance uses 4vCPU and 8GB RAM)

SaaS solutions

The chemical structures of the research compounds are the biggest asset/treasure of a pharmaceutical company, and therefore production systems were not allowed to cross the firewall in the past. Nowadays, as the trust in cloud based technologies has been increased, the technological and maintenance costs of internal systems tends to overcome the cost related to the risk in using external solutions.

The transition from the on-premise application business model to the cloud-based Software as a Service model is appealing, because hosted subscriptions offer many potential advantages:

- Lower upfront cost – eliminates the upfront cost of installation, as well on-going costs like maintenance
- Quick setup and deployment, ready for integration – the service is already installed and in the cloud. Only needs configuration
- Seamless automatic upgrades

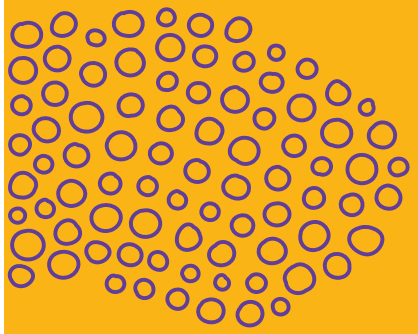
SaaS solutions are attractive for small companies, because they usually do not have the required IT expertise, and for large enterprises alike, because they need a high-performance, cost-effective solution. That is why Chemaxon offers Compliance Checker Software as a Service tailored for every business need. The competitively priced Lite subscriptions were created for those small and medium-sized companies which need a state-of-the-art compliance solution and are ready to make compromises in terms of the number of users, molecules to be checked, and speed.

Summary

It is likely that the foreseeable growth in the market for new psychoactive substances will continue to pose challenges for drug policy bodies and consequently for R&D organizations. A robust and reliable system helping to identify the controlled substances is a very first step in the process to meet the legal requirements.

As generic definitions are being more widely used by regulatory bodies, a new search technology was required, which can quickly and accurately find hits against highly complex Markush structure queries even in large compound collections. Chemaxon's new improved **Markush search engine** makes this possible: it can complete searches significantly faster than traditional Markush search techniques.

The solution also has to be highly automated and integrable, to provide up-to-date compliance information at any time, and anywhere in the compound research-production-logistics workflow. **Compliance Checker's microservices-based architecture** results in an easily scalable system, which is also available via Software as a Service subscriptions. The full check of even such a large collection as MolPort's 45.8 M compound set can be completed in a weekend against all countries currently covered by our **Premium Subscription**.



Contact

info@chemaxon.com

www.chemaxon.com/products/compliance-checker

Chemaxon