Methods for “Trends in the availability and types of drugs sold on the internet via cryptomarkets, June 2022 - May 2023”

Nicola Man, Qingyuan Linghu, Raimondo Bruno, Monica J. Barratt, Rachel Sutherland and Amy Peacock
This report was prepared by the National Drug and Alcohol Research Centre, UNSW Sydney
For further information: drugtrends@unsw.edu.au

1. General Note

Data presented in the public online interactive visualisation and reported in bulletins are obtained from the regular scraping, collation, and analysis of drug listings on cryptomarkets, conducted as part of the Drugs and New Technologies (DNeT) project. The DNeT project has been running since 2012 and forms part of Drug Trends, an illicit drug monitoring system in Australia.

The primary aim of cryptomarket monitoring within Drug Trends is to understand trends in the online availability of psychoactive substances on cryptomarkets, also known as darknet markets. Specifically, we focus on listings advertising the sale of illicit drugs (e.g., heroin), key licit drugs (e.g., alcohol, tobacco, e-cigarettes) and pharmaceutical medicines, as well as drug-related paraphernalia (e.g., needles and syringes, reagent test kits).

In this Methods document, we outline the background to this program of work and the methods underpinning data presented. There are various approaches to collecting, collating, categorising and analysing cryptomarket data, and inherent challenges in these processes. There are also limitations and constraints on the appropriate interpretation of these data (see below for further detail). For this reason, we have attempted to be as transparent as possible about our procedures.

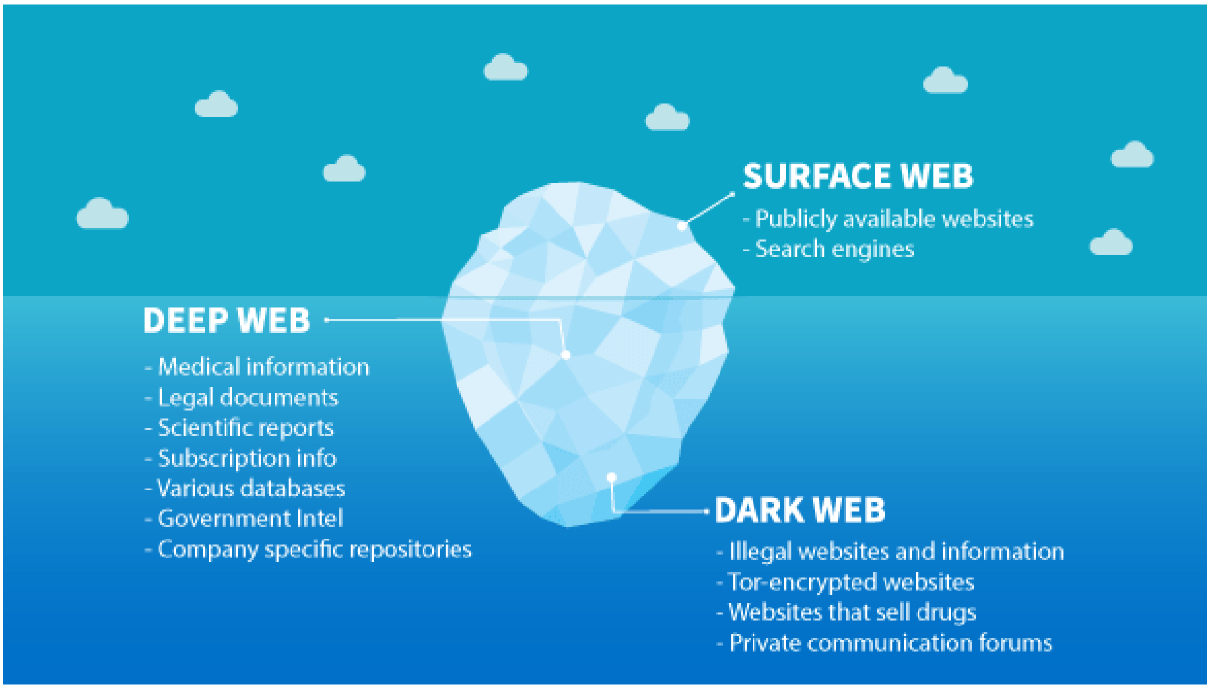
Our analysis of cryptomarket listings is an ongoing process, with ongoing refinements to the categorisation and monitoring process. Data may not be comparable from one output to the next as we implement improvements. We welcome feedback and suggestions so that we can continue to improve utility of these data and our reporting on them (drugtrends@unsw.edu.au).

2. Definition of ‘cryptomarkets’

Cryptomarkets (‘darknet markets’) are online marketplaces that facilitate the purchasing of illicit goods and services via multiple sellers, and provide participants with anonymity via its location on the hidden web. Content available on the internet can be divided into the surface web (content accessible via search engines) and deep web (content inaccessible via search engines, including paywalled websites, private websites, company or academic databases).

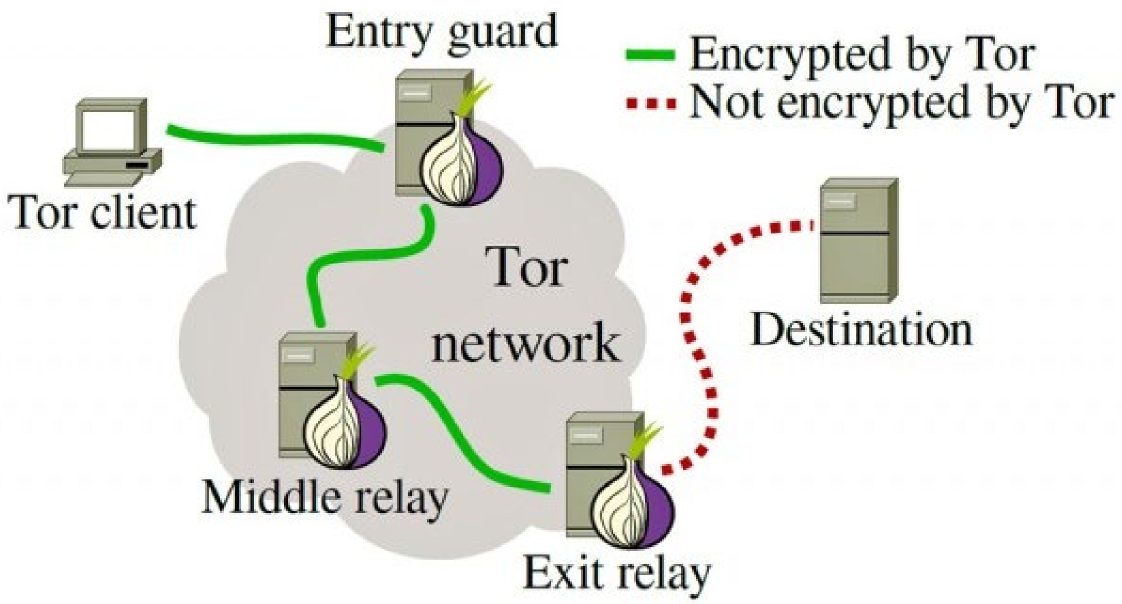
Cryptomarkets are located on a part of the deep web called the 'hidden web' (or 'dark web'): a part of the internet that is accessible only through hidden internet networks such as **Tor Network ('The Onion Router')** (see Figure 1).

Figure 1. Description of the content available on the internet. For further information, see Barratt & Aldridge (2016). Image: vpnoverview.com



Tor is an open source network that re-routes an internet user's IP address through various encrypted nodes (see Figure 2). This process ensures concealment of the location of operating servers and provides a level of anonymity for both buyers and sellers.

Figure 2. Tor network structure. Image: The Conversation

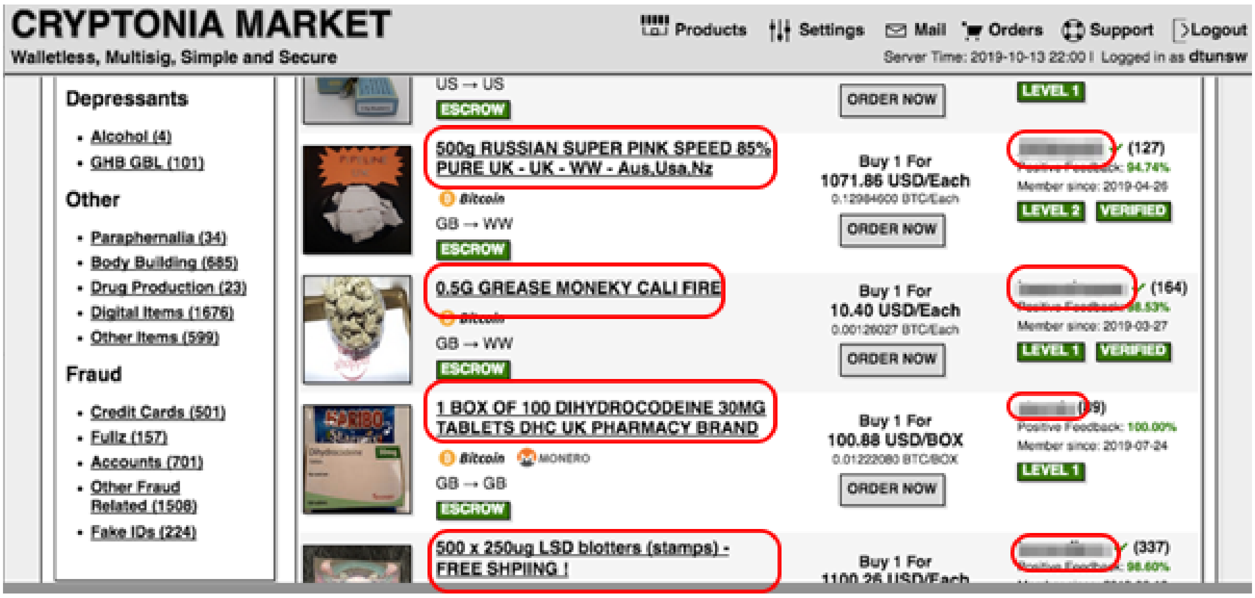


DOI: 10.26190/r9tc-ba06

Anonymity is further maximised through monetary transactions made via the exchange of cryptocurrencies such as Bitcoin and Monero. These transactions are facilitated through escrow systems which are controlled by the cryptomarket on which the purchase is made. Escrow allows a cryptomarket to hold the funds for a transaction until the product is delivered.

Cryptomarket websites are often similar in style to E-commerce websites such as eBay or the Amazon marketplace (see Figure 3 for a picture of the landing page from 'Cryptonia Market' as an example).

Figure 3. The landing page from 'Cryptonia Market' (2019)



Trade in illicit drugs via cryptomarkets is highly dynamic, as cryptomarket lifespans have been constantly interrupted for various reasons. Analysis of 89 marketplaces that were operational from 2010 to June 2017 by the European Monitoring Centre for Drugs and Drug Addiction showed that the main purported reason for market closure was 'exit scamming', where a market site will shut down suddenly, taking the money held in escrows for incomplete orders. Another primary reason was a 'voluntary exit', where a market will close with the mutual consent of those involved, without losses to vendors and buyers. Finally, law enforcement agencies or hackers may decide to target markets and subsequently force their closure.

3. Overview of approaches

The DNeT project has been examining cryptomarket drug listings since 1st February 2014[1]. Cryptomarket websites often exhibit large volumes of data. In addition, security on these websites has become increasingly complex in response to law enforcement seizures. This has made it increasingly difficult to traverse, extract and collate this information on a timely and accurate basis. To mitigate these issues, we have implemented a range of programmed automated processes that operate with minimal manual input. These automated processes, which are described in detail below, consist of (1) traversing the contents of currently active cryptomarkets; (2) parsing the contents of individual webpages and extracting relevant and usable data in

[1] Prior to 2014, fortnightly monitoring of cryptomarkets, Silk Road, Black Market Reloaded (BMR) and the Sheep Marketplace were conducted. However, due to incomplete data, current methodology could not be applied to this data, and hence these data were excluded from reporting.

DOI: 10.26190/r9tc-ba06

relation to the drugs that are listed for sale; (3) classifying these listings into specific drug categories; and (4) data cleaning including deduplication of listings.

It is important to caveat that previous reporting on drug listings on cryptomarkets from DNeT in earlier bulletins were based on different approaches to collection, extraction, drug categorisation, data cleaning and reporting. For this reason, reporting may not be comparable over time. These differences will be described in the relevant sections below. As much as is possible, we have applied the automated approaches described in this methods document to the historical data and presented them in our online visualisation.

This methods document is for our bulletin which focuses on cryptomarket data in the period from 1st June 2022 to 31st May 2023. This bulletin and the accompanying interactive data visualisation are based on twice monthly snapshots of cryptomarket listings, undertaken in the two weeks starting on the 1st and the 15th of each month (see section 4.2). **For a historical record of the weekly cryptomarkets monitored by DNeT, we refer the reader to our interactive timeline and this bulletin for the weekly data up until 31st May 2022.** We have performed deduplication of listings for quantity variants in this bulletin (see section 5.2.1). **For the historical record of the weekly cryptomarket data with quantity variants, please refer to this interactive timeline and the summary bulletin which report on key findings in the period from 1st February 2014 to 31st January 2020.** In addition, we have improved on the drug categorisation and added new drug terms as part of our process of continuous improvement for data from 1st February 2020 onwards (see section 5.1.2).

This project has institutional approval and ethical approval from the University of New South Wales Human Research Ethics Committee (HC220754). Due to anonymised online details, no individual results were gathered and data are published in tabular, aggregate form only.

4. Method of collection and extraction of cryptomarket data

4.1. Identification of cryptomarkets

Active cryptomarkets and their Tor links are identified through surface-web and/or hidden-web sites such as darkfail.net and tor.taxi. These sites provide regularly updated information regarding the status of sites hosted on the Tor network. Certain internet forums accessible from both the hidden and surface web are also used as sources for cryptomarket developments and active Tor links. In particular, Dread (which started on 16th February 2018) is often used as a source of information because it is an active and popular hidden-web forum with a dedicated subDread or forum group for many cryptomarkets. However, Dread had been mostly offline due to persistent DDOS since around August 2022.

Marketplaces are included in monitoring if they:

- Facilitate trade of illicit goods and services via multiple sellers (i.e. >1 vendor on the marketplace);
- Display more than 100 drug listings;
- Are displayed in the English language;
- Vendors ship to and/or from Australia, or ship to and/or from multiple countries; and,
- Have an accessible Tor link for scraping.

Ongoing monitoring of smaller markets is undertaken to determine whether they meet the number of listings threshold with market growth over time. In addition, some markets may meet the >100 drug listing threshold criterion during at least one scrape and thus are included, but experience period(s) of market decline where the number of drug listings is ≤100. These markets and their data are still included for regular monitoring. There may also be markets that we cease scraping because they required too much effort, e.g. difficult CAPTCHA that pop up too frequently (which we note in the bulletin when describing these markets).

4.2. Crawling of cryptomarkets

In order to regularly access a cryptomarket, a dedicated account is created upon its initial monitoring. Since August 2018, custom web crawling (or scraping) programs have been created independently for each market selected for monitoring (semi-automated web crawling programs were undertaken using different language programs prior to this; see section 3). These crawling programs are written in Python and implemented with Selenium, an open-source web-based automation tool. These crawling scripts connect to the Tor network and systematically traverse a specified cryptomarket while concurrently storing the raw web content of each page. Upon initial inspection of a new cryptomarket, a human will code the location of the drug categories that need to be parsed. This code then informs the crawlers of the location of the specified drug categories, allowing the crawler to then iterate through the indexed pages of each of those categories. Note, the crawler does not parse individual vendor or listing pages, but only the indexed pages containing advertised listings.

To ensure an automated crawl is complete and exhaustive, numerous checks take place during and after the completion of a crawl. Firstly, the crawler is designed to pause for a pre-defined amount of time until specific elements of a webpage are visible before saving its contents and continuing onto the next webpage. This process ensures that the crawler does not capture an incompletely loaded webpage with missing content.

These time lags which are often randomized also allow the crawler to mimic a human user, and therefore avoid alerting anti-DDOS (Distributed Denial-of-service attacks) marketplace mechanisms, which may result in the blocking of our accounts' ability to access and view marketplace content. Secondly, as the webpages of a cryptomarket usually share a formulated layout, the saved contents can be automatically parsed and verified to satisfy certain conditions (e.g., checking every stored webpage contains the expected listed items for sale). Finally, as the crawler saves the web content in a directory that resembles the organisation of the marketplace, the correct number of pages can be easily verified against what is displayed on the marketplace.

For the large majority of markets, crawling is often interrupted by CAPTCHA challenge-response tests that were intended to determine if the cryptomarket users were human. To counteract this, automatic detection mechanisms have been implemented, allowing a human user to be notified of an interruption, and then given sufficient time to solve the CAPTCHA challenge. After successful completion of the CAPTCHA challenge, the crawler resumes the traversal of cryptomarkets.

When a marketplace is inaccessible or only partially accessible, the data are treated as missing. If a cryptomarket can only be partially crawled at a given time point, the collected web content is not entered into the dataset. A marketplace may be down for multiple reasons, including server outages, distributed denial of service attacks (DDoS; in which multiple sources are used to generate a large amount of traffic to an online service, thereby overwhelming its servers), law enforcement seizures, exit scams and hacking attacks. If a marketplace is down at one time point, unless there was reason to believe it would not return (in the case of seizures or exit scams), attempts are made to access it at the next time point. If consecutive attempts to access a cryptomarket fail for several scrapes, we search for evidence of the market's closure (e.g. on Dread, from other hidden web sources, and/or from surface web sites) to confirm it is no longer operational.

Automated crawling is conducted on a regular basis (monitoring is ongoing) on a stand-alone computer, and can take minutes to hours to complete for each market, depending on the size and the operating status of the markets. As instantaneous snapshots of the entire marketplace are not possible with our current methods, an automated crawl is intended to be completed in the shortest possible time, and therefore, act as the best approximation of an observed marketplace at a given time.

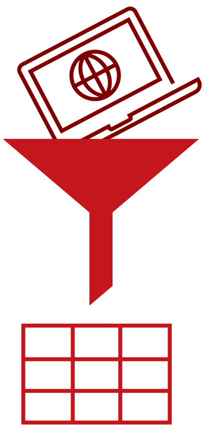
From 1st August 2022, data are scraped twice monthly in each of the two weeks starting on the 1st and the 15th of each month. The snapshot is taken on the first Thursday of the two-week period. If data from a cryptomarket are missing or cannot be scraped completely, scraping is attempted again for the market in the following week. As such, twice monthly snapshots of cryptomarket listings are presented in this bulletin as data in the two weeks starting on the 1st and the 15th of each month. Please note that historical data before 1st August 2022 were captured on a weekly basis. The twice monthly snapshots in the current interactive data visualisation were created from these historical data before 1st August 2022 (see section 5.3).

Prior to the implementation of fully automated crawling on 9th August 2018, markets were traversed, with raw URLs saved, and data extracted using semi-automated VBA programming processes.

4.3. Extraction of key features from cryptomarket data

A second set of automated scripts were built and implemented in order to parse and extract relevant information from the collected raw data obtained by the web crawlers. Due to differing webpage layouts between cryptomarkets, independent extraction scripts for each market are created using the HTML scraping tool Beautiful Soup and/or Scrapy.

A conscious decision was made to separate these two steps and allow for the extraction process to be undertaken after the completed crawling stage. This separation was justified due to the short time frame a crawler can remain on a webpage and the continuously changing nature of cryptomarket webpage formats.



These scripts extract the text within the drug listings displayed on each webpage. The scripts parse the stored webpage and extract the relevant features of each observed listing to a readable table. Features within a listing differ by marketplace as each market varies on what they decide to make visible on the listing page. For example, certain markets will include the product's country of origin and available destination countries, while others may decide to only show limited or no shipping information. Other information such as quantity or a vendor's ratings are present on some markets. Certain features remain common between all markets; these are the features we focus on for extraction and reporting. These include the drug listing title and vendor username. Figure 3 shows an example listing page with the features to be extracted and reported on outlined in red.

DOI: 10.26190/r9tc-ba06

5. Method of drug categorisation and data deduplication

In the bulletins published before the year 2020 on the cryptomarket data, drug categorisation was achieved via a dynamic lookup table compiled using previously identified terms and their associated categories. In the current bulletin, drug categorisation on the data from 1st February 2020 onwards were improved with the addition of new drug terms for rule-based classification and in the algorithm for the final assignment of listings to drug categories (see **section 5.1.2.3**).

In bulletins published from the year 2020 onwards, we have deduplicated identical listings from the same vendor listing in the same market and appearing in the same snapshot as described in **section 5.2**. In the bulletins published from the year 2022 onwards, we have also additionally deduplicated listings for quantity variants (see **section 5.2.1**).

5.1. Drug categorisation

5.1.1. Drug categorisation structure

To develop an understanding of what substances are being sold on cryptomarkets, it is necessary to have a consistent framework for labelling drug listings into mutually exclusive categories. Unfortunately, categories imposed by cryptomarkets are often ambiguous or too broad for our purpose (e.g. 'psychedelics'), and categories conflict across marketplaces. Drug listings are also often miscategorised (e.g., cannabis listings categorised as 'stimulants' or non-drug listings appearing in drug-related categories) due to server-side issues (e.g. the server does not process the submitted listing properly and puts it into the wrong category), or vendors listing their products in the wrong category. To ensure consistency across marketplaces, we have assembled a two-tier hierarchy for classifying each recorded drug listing independently. At the first level (Level 1), drug listings are assigned a specific drug name meant to capture the active substance being sold in that listing (noting that there are some exceptions). At the second level (Level 2), drug names at Level 1 are clustered into a smaller set of mutually exclusive broader drug classes (see **Table 1**). The exception comprises key illicit drugs of interest (e.g., MDMA, cocaine, cannabis) which are categorised identically at Level 1 and Level 2 (e.g., 'MDMA' is in both Level 1 and Level 2). Note that the product listings included in our bulletin data include drug-related products such as vaping equipment, drug-testing kits or paraphernalia, i.e. they are not exclusively drugs or substances for consumption.

Table 1. Drug categorisation structure

Level 2 drug class	Description (with ATC code[a] if applicable)	Level 1 drug class
Alcohol	-	Alcohol
Benzodiazepines	Benzodiazepines (N03AE, N05BA, N05CD). Note that certain novel benzodiazepines (e.g., etizolam) have been classified as NPS.	E.g., alprazolam, diazepam
	Drug listings with two or more (non-NPS) benzodiazepines identified	Mixed/uncategorised benzodiazepines
Cannabis	All forms of cannabis, i.e. the plant, oil, seeds, candies, etc.	Cannabis
Cocaine	-	Cocaine
DMT	DMT only, excluding plant sources	DMT

E-cigarette	This excludes vaping equipment which is classified as paraphenalia.	E-cigarette
GHB/GBL/1,4-BD	GHB or GBL or 1,4-BD	GHB/GBL/1,4-BD
Hallucinogenic mushroom	-	Hallucinogenic mushroom
Heroin	-	Heroin
Ketamine	-	Ketamine
Inhalants	-	Alkyl nitrites, and nitrous oxide
LSD	-	LSD
MDA	-	MDA
MDMA	-	MDMA
Meth/amphetamine (illicit)	Any illicit amphetamine or methamphetamine (includes speed). This excludes substances identified as pharmaceutical stimulants (see below).	Meth/amphetamine (illicit)
New psychoactive substances (NPS)	New psychoactive substances (e.g., acetylfentanyl and 5-MeO-MiPT). Note that this class includes synthetic cannabinoids, novel benzodiazepines, fentanyl analogues and other emerging substances.	New psychoactive substances
Opioids (excluding heroin)	Pharmaceutical opioids (N02A, N01AH), and opium. Note that certain novel synthetic opioids (e.g., acetylfentanyl) have been classified as NPS.	E.g., oxycodone, tramadol, fentanyl, hydrocodone, codeine, morphine, sufentanil, opium
	Opiate drugs used in replacement therapy (N07BC)	E.g., methadone, buprenophine
	Drug listings with two or more opioids (excluding heroin and NPS opioids) identified	Mixed/uncategorised opioids
Other psychostimulants and nootropics	Pharmaceutical psychostimulants (N06B)	E.g., dexamphetamine, methylphenidate, modafinil
Paraphernalia	-	Drug checking equipment, injecting equipment, naloxone, vaping equipment, and other paraphernalia
PCP	-	PCP
PIEDs[b]/weight loss drugs	Performance and image enhancing drugs and weight loss products	E.g., anabolic steroids (A14A), androgens (G03B), skin treatment, antiobesity preparations, excl. diet products
	Drug listings with two or more PIEDs/weight loss products identified	Mixed/uncategorised PIEDs/weight loss
Tobacco	-	Tobacco
Other drugs	Precursors	E.g. ephedrine

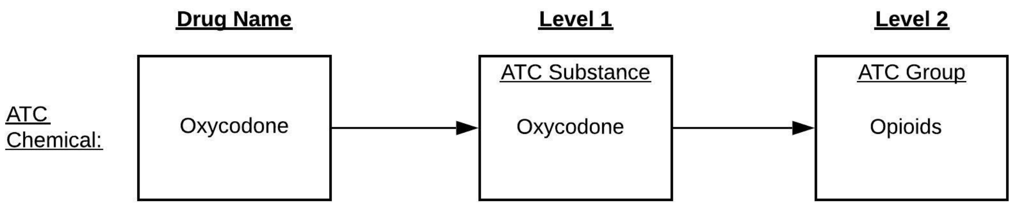
	Drug listings with multiple drugs identified (e.g., "Cocaine+Heroin") spanning across two or more Level 2 drug categories	Mixed/uncategorised drugs
Other psychotropic medicines[c]	Antiepileptics (N03 excluding N03AE which is classified as benzodiazepines)	E.g., pregabalin, gabapentin
	Antipsychotics (N05A)	E.g., quetiapine
	Anxiolytics (N05B excluding N05BA which is classified as benzodiazepines)	E.g. hydroxyzine
	Barbiturates (N05CA)	E.g., pentobarbital, phenobarbital, thiopental
	Hypnotics and sedatives (N05CM)	E.g., methaqualone, scopolamine, clomethiazole
	Antidepressants (N06A)	E.g., sertraline, fluoxetine, mirtazapine, bupropion
	Anti-dementia drugs (N06DX)	E.g. memantine
Other medicines[d]	Cardiovascular system drugs (C)	E.g., propranolol
	Antipruritics, incl. antihistamines, anesthetics, etc. (D04)	E.g., promethazine, benzocaine, lidocaine, procaine
	Drugs used in erectile dysfunction (G04BE)	E.g., vardenafil, dapoxetine, tadalafil, sildenafil
	Corticosteroids for systemic use (H02)	E.g. prednisolone
	Antivirals (J05AP), antiparasitic products,	E.g. choloroquine, daclatasvir
	Muscle relaxants (M03)	E.g., carisoprodol, baclofen
	General and local anaesthetics (N01 excluding N01AH which is classified as opioids)	E.g., benzocaine, lidocaine, propofol
	Analgesics and antipyretics (N02B, N02C)	E.g., clonidine
	Drugs used in addictive disorders (N07B excluding N07BC which is classified as opioids)	E.g., disulfiram, naltrexone
	Respiratory system drugs (R)	E.g., promethazine, lidocaine, salbutamol
	Ophthalmological drugs (S01)	E.g., lidocaine, scopolamine

Note: [a] ATC code: Anatomical Therapeutic Chemical code [b] PIEDs: performance and image enhancing drugs. [c] Other psychotropic medicines were classified under Other medicines in the bulletins before the year 2023 (see Methods of the last bulletin of the year 2022 for the former classification). [d] Some drugs belong under more than one ATC code, e.g. promethazine is under both D04AA and R06AD, and lidocaine is under D04AB, N01BB, R02AD and S01HA. Given they are all classed under "Other medicines" the lack of mutual exclusivity of ATC classes is not an issue. Where mutual exclusivity of ATC code is an issue for our categorisations (e.g. for benzodiazepines and opioids), we have noted the exclusions in the description column.

5.1.1.1. Categorisation of pharmaceutical drugs

Pharmaceutical drugs were categorised according to the Anatomical Therapeutic Chemical (ATC) classification system. The ATC system classifies drugs into groups at five different levels, describing their therapeutic, pharmacological and chemical properties, with the lowest ATC level containing the active chemical substances found in drugs. If a Level 1 pharmaceutical drug exists in the ATC system, then a higher ATC level of that drug was chosen as the broader class at Level 2. The higher level was chosen such that: (1) the class generalises to similarly acting drugs that are also found on cryptomarkets; and (2) the class name suitably describes the drugs falling under that class. The chosen Level 2 class names can correspond to either the therapeutic (2nd highest), pharmacological (3rd highest), or chemical (4th highest) subgroup levels found in the ATC system. The major ATC classes we have used at Level 2 are: benzodiazepines, opioids (excluding heroin), other psychostimulants and nootropics, and PIEDs/weight loss. All other ATC classes are grouped under other medicines at Level 1. Multiple entries of a single substance may exist in different branches of the ATC system. This is due to some chemical substances having multiple therapeutical uses. Our major classes defined above take precedence over other ATC classes, e.g. N01AH are opioid anaesthetic drugs and is categorised under 'opioids (excluding heroin)' instead of under 'other medicines' as 'general and local anaesthetics' (N01A; see **Table 1**). A detailed description of ATC structuring can be found here. See **Figure 4** for an example classification according to ATC.

Figure 4. Example categorisation structure for ATC chemicals.

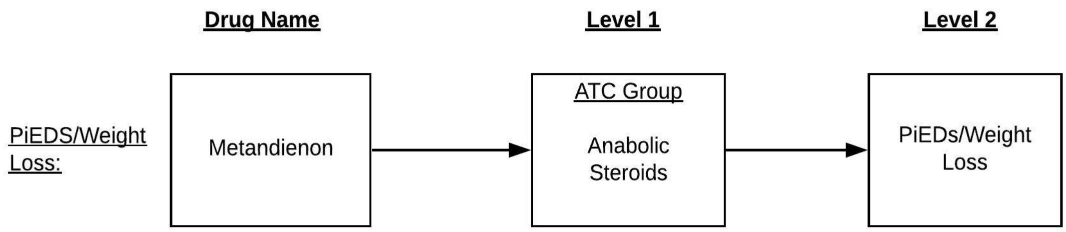


5.1.1.2. Performance and image enhancing drugs (PIEDs) and weight loss drugs.

As drugs that would traditionally fall under PIEDs/weight loss drugs may appear in different branches of the ATC system due to differing therapeutical and pharmacological grouping, it was necessary to apply a suitable ATC level at Level 1 rather than at Level 2. If there was no direct match to an ATC code, we have assigned an appropriate ATC categorisation based on drug indications or functional effect.

These ATC categories at Level 1 are then grouped to the class 'PIEDs/weight loss' at Level 2. For example, the steroid testosterone is found in the ATC group 'Androgens' (G03B), whereas the steroid oxandrolone is found in the ATC group 'Anabolic steroids' (A14A), both of which do not have a common ATC parent Level. Therefore, the ATC levels 'Androgens' and 'Anabolic steroids' were assigned at Level 1 respectively, instead of the steroid drug name (see **Figure 5**).

Figure 5. Example categorisation structure for PIEDs/weight loss drugs.



5.1.1.3. New psychoactive substances (NPS)

For the purpose of this reporting, we defined NPS as any narcotic drugs or psychotropic substances made available or used from the early to mid-2000s for their psychoactive properties. These drugs are not scheduled or have only been listed since 2015 under the international drug control conventions of 1961 and 1971 and could pose similar threats to public health as substances scheduled under these conventions. There is no standard nomenclature for NPS nor publicly accessible list of all NPS identified globally. For this reason and because of the variety of street names for each NPS, it can be very difficult to identify unique NPS from information provided in the listing. Sources for our list of NPS names includes UNODC, RESPONSE, EMCDDA reports, and presentations at https://www.novelpsychoactivesubstances.org/.

Note that we intend to provide further disaggregation of specific NPS classes at Level 1 (e.g. benzodiazepines, opioids, cannabinoids, phenylamines) in the public online interactive visualisation in the future. Please contact us (drugtrends@unsw.edu.au) if you have queries regarding results for a specific substance.

5.1.1.4. Other non-standard categories

Additional Level 2 classes were created to accommodate non-standard drug listings such as drug paraphernalia and precursors. The 'paraphernalia' class contains injecting equipment, vaping equipment, drug checking equipment, naloxone and other paraphernalia.

Custom-made listings that are intended to be ordered by a specific buyer were identified with the keyword 'custom' in the listing title and excluded. Miscellaneous non-drug items are also identified through keyword matches (e.g. 'recipes', 'tip jar', 'how to ') and excluded.

5.1.2. Applying categorisation of listings

In order to apply this categorisation to the extracted listings, we classify the text contained in the listing title to a Level 1 category. To ensure the largest possible number of listings have been categorised with a high degree of confidence, we have implemented an automated classification procedure comprising of two components: a *rule-based system* and a *machine learning classifier*. Note that both components are performed on all listings in the extracted dataset.

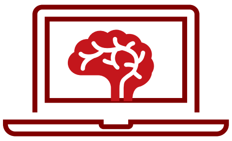
5.1.2.1. Rule-based system

The *rule-based system* checks each listing title for the appearance of certain phrases or words which have been previously mapped to a Level 1 category by a human.

To carry out the rule-based system, individual words contained in the listing title have been matched against a database of drug names and their street and pharmaceutical brand variants developed by the researchers, all of which have an associated Level 1 category. If a single match exists (i.e., there is a unique phrase in our database that is found within the listing title), then, with confidence we can assign the corresponding Level 1 category of that matched phrase. For example, the street name 'xtc' for Level 1 category MDMA is matched to the listing title '5 XTC Pills Bitcoin Stamp 250MG', resulting in the assignment of MDMA at Level 1 for that listing.

The text matching component of classification is implemented using a standard string searching algorithm in Python, which can account for variations in punctuation and capitalisation through the use of regular expression commands.

5.1.2.2. Machine learning classifier



A *machine learning classifier* that has been trained on historically categorised listings is used to classify all listing titles in the extracted dataset. The machine learning component was deemed necessary as (1) it is not feasible to manually categorise thousands of un-matched listings on a regular basis; and (2) it is not feasible to account for all the different ways a drug can be listed with pre-set rules and instructions.

The machine learning model utilises a long short-term memory (LSTM) artificial neural network. Listing titles that were classified by this predictive model with a target predictive accuracy greater than 90% were assumed to be valid. This threshold value was deemed appropriate after a manual check of 200 newly classified un-matched listings. These listings categorised by the predictive model, on average, accounted for 65% of the un-matched listings from February 2014 to January 2020.

The LSTM neural network was trained on a set of 5 million unique categorised listings from February 2014 to January 2020 that were labelled correctly by our rule-based system. To avoid bias, it was ensured this training set included a representative number of all the target Level 1 categories. In order to translate the listing titles into understandable inputs for the neural network, word-embeddings were produced for each of the 52,000 unique words present in all of the historically collected listing titles. These were achieved with a word2vec model programmed in Python that embedded the contextual and sequential relationships found in the various drug listing titles into transformed numerical inputs. The subsequent training and development of the LSTM model was completed through the neural network library, Keras. Cross-validation was performed to assess model validation, and an out-of-sample accuracy of 96% was obtained.

5.1.2.3. Final assignment of Level 1 category to listing

The following algorithm is then applied after implementing the two components of our classification procedure described above.

1. Listings are categorised to a Level 1 category if a single drug could be identified using the *rule-based system*;

2. For the listings with a match to two or more Level 1 categories from the *rule-based system*, common issues include drug-specific equipment containing both a drug name and equipment keyword (e.g., "cannabis vaping equipment"), the ambiguity of MDMA appearance or pill shape names that can refer

to other drugs (e.g. “2c-b pills orange batman”), or a common drug deemed to be similar to an NPS drug is used as a descriptor of the NPS drug (e.g. “flualprazolam (like Xanax)”). To correct for these duplicate cases, the following overriding rules were scripted to allow further categorisation of listings to a Level 1 category.

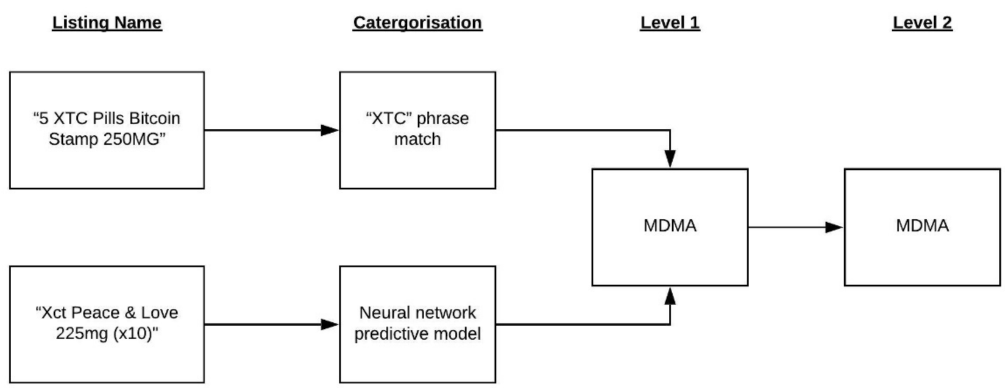
- Vaping equipment is assigned as the Level 1 category when the listing contains another drug name (e.g., “THC e-cig” is assigned to vaping equipment).
- When the listing (e.g. “2c-b orange batman”) is matched to two or more drugs, and MDMA was identified with terms based on appearance or pill shape names (e.g. ‘moonrock’, ‘superman’, ‘mario’, ‘orange batman’), the other drug(s) override or take precedence over MDMA as the identified drug. This is a new addition to the algorithm for data collected from 1st February 2020 onwards.
- The remaining listings with two or more matches were vetted to work out common exception pairs, e.g. listings identifying both flualprazolam and another common benzodiazepine like alprazolam were classified as flualprazolam under NPS as the Level 2 drug class. The majority of these listings used the more common or well-known drug as the descriptor of the drug being sold in the listing.

3. For the remaining listings with more than one Level 1 category identified and where the multiple drugs identified belong to the same Level 2 drug category, they are categorised as a mixed/uncategorised drug under that Level 2 category, e.g. "Mixed/uncategorised benzodiazepines" as a Level 1 drug category under "Benzodiazepines". Where the multiple drugs identified belong under two or more Level 2 drug category, they are categorised into the ‘Mixed/uncategorised drugs’ category at Level 1 under the ‘Other drugs’ category at Level 2. This is a new addition to the algorithm for data collected from 1st February 2020 onwards.

4. For those listings that are not categorised to a Level 1 category by the *rule-based system*, a valid Level 1 category (i.e. with a predictive accuracy greater than 90%) from the *machine learning classifier* is then assigned as the Level 1 category (see **Figure 6**).

5. Custom listings (e.g., “cocaine custom listing”) and listings that are how-to guides for making or growing drugs (e.g. “how to grow cannabis at home”) are excluded even if drug categories were identified for the listing (see **section 5.1.1.4**).

Figure 6. Example of categorisation process applied to two MDMA listings.



We continually match and review uncategorised listings to lists or databases of drug names. Some pharmaceutical drugs have been identified from their generic names in the uncategorised listings in this bulletin and the previous bulletin. **Table 2** shows the 160[2] Level 1 drugs that were further identified using new single-term drug names curated from selected drug classes in the ATC database. These pharmaceutical drugs will be added to the drug term database so that future listings of these drugs can be categorised by the *rule-based system*. These new listings will be used to update the historical categorised training set in the future, and the machine learning model will be re-trained on this updated dataset. This process will allow future unmatched drug listings containing newly identified drugs to be categorised by the predictive model. Please note that the new terms have not been applied to the training set as yet because of the time and effort it takes to train and evaluate the machine learning classifier.

Table 2. Frequency of pharmaceutical drug names identified with new drug terms from selected classes of single-term drugs in the ATC database

Drug class	Feb 2020 – Sep 2022	Oct 2022 – Jan 2023	Feb 2023 – May 2023	Total
Benzodiazepines				
Clobazam	217	21	6	244
Estazolam	92	31	11	134
Loprazolam	48	6	0	54
Quazepam	1	0	0	1
New psychoactive substances (NPS)				
Tofisopam	163	47	25	235
Opioids (excluding heroin)				
Pentazocine	2	12	0	14
Other psychostimulants & nootropics				
Caffeine	68	104	84	256
Citicoline	16	6	13	35
Pramiracetam	2	0	0	2
Vinpocetine	17	6	0	23
Other psychotropic medicines				
Agomelatine	39	28	65	132
Amantadine	28	17	51	96
Amisulpride	189	47	75	311
Amobarbital	0	2	0	2
Apomorphine	8	0	0	8
Biperiden	11	0	0	11
Brexpiprazole	0	6	13	19
Bromocriptine	14	9	8	31
Buspirone	0	0	56	56
Cabergoline	187	170	173	530
Carbamazepine	154	25	13	192
Chlorpromazine	38	24	24	86
Citalopram	246	56	45	347
Clomipramine	59	27	62	148
Clozapine	4	22	53	79
Cyamemazine	12	0	0	12
Desvenlafaxine	1	2	13	16

[2] This excluded the new Level 1 categories for mixed/uncategorised drugs, e.g. 'Mixed/uncategorized PIEDS/weight loss drugs', 'Mixed/uncategorized antibacterials', 'Mixed/uncategorized anxiolytics', 'Mixed/uncategorized psychostimulants and nootropics'.

DOI: 10.26190/r9tc-ba06

DNeT Drugs and New Technologies

Dexmedetomidine	8	0	0	8
Donepezil	0	0	1	1
Doxepin	52	27	71	150
Duloxetine	231	52	66	349
Emylcamate	53	28	47	128
Entacapone	7	0	0	7
Escitalopram	761	204	182	1147
Etifoxine	10	6	0	16
Fluvoxamine	148	53	69	270
Galantamine	3	0	0	3
Haloperidol	48	5	0	53
Imipramine	18	20	64	102
Lamotrigine	85	28	33	146
Levetiracetam	100	27	30	157
Levodopa	80	12	9	101
Levomepromazine	43	0	6	49
Loxapine	175	12	8	195
Melatonin	7	64	12	83
Melitracen	0	0	8	8
Methohexital	0	2	0	2
Milnacipran	12	7	6	25
Nortriptyline	58	14	56	128
Olanzapine	0	0	126	126
Oxcarbazepine	1	0	6	7
Paroxetine	246	66	90	402
Perampanel	6	1	4	11
Pramipexole	128	18	0	146
Primidone	0	2	6	8
Prochlorperazine	1	0	0	1
Promazine	110	28	14	152
Prothipendyl	21	0	6	27
Rasagiline	2	0	0	2
Reboxetine	14	6	0	20
Risperidone	104	27	10	141
Rivastigmine	0	0	6	6
Ropinirole	20	12	1	33
Secobarbital	58	36	10	104
Selegiline	31	6	0	37
Suvorexant	0	14	15	29
Tianeptine	609	162	55	826
Topiramate	23	0	26	49
Trifluoperazine	4	0	6	10
Vilazodone	1	15	74	90
Vortioxetine	19	7	0	26
Ziprasidone	1	14	56	71
Zuclopenthixol	8	0	0	8
Other medicines				
Acamprosate	1	0	0	1

Aciclovir	190	26	61	277
Allopurinol	10	0	0	10
Amoxicillin	0	0	309	309
Azithromycin	1237	221	206	1664
Betamethasone	0	3	0	3
Cefadroxil	0	0	14	14
Cefalexin	16	6	0	22
Cefdinir	0	0	14	14
Cefditoren	0	0	14	14
Cefixime	60	51	44	155
Cefpodoxime	0	0	24	24
Ceftibuten	0	0	14	14
Ceftriaxone	1	0	0	1
Cefuroxime	15	6	19	40
Celecoxib	50	0	3	53
Cinnarizine	6	0	0	6
Ciprofloxacin	334	88	88	510
Clarithromycin	18	6	4	28
Clindamycin	46	0	14	60
Deflazacort	9	0	8	17
Dexamethasone	30	18	0	48
Diamorphine	22	0	0	22
Doxycycline	645	156	179	980
Emtricitabine	37	0	3	40
Entecavir	0	3	13	16
Eperisone	0	3	0	3
Erenumab	2	0	0	2
Ergotamine	19	0	0	19
Ertapenem	7	0	0	7
Erythromycin	1	0	0	1
Febuxostat	0	0	13	13
Fluconazole	305	104	100	509
Gemifloxacin	0	0	13	13
Gentamicin	10	6	7	23
Gonadorelin	31	2	8	41
Hydrocortisone	4	6	0	10
Ibuprofen	23	4	0	27
Indometacin	0	6	0	6
Isoniazid	23	6	0	29
Isotretinoin	1405	191	183	1779
Itraconazole	7	0	8	15
Ketorolac	15	2	8	25
Lamivudine	0	0	8	8
Levofloxacin	32	38	23	93
Linezolid	0	0	8	8
Lymecycline	5	0	8	13
Mepivacaine	1	0	0	1
Meropenem	10	6	6	22

Methenamine	0	0	6	6
Methoxyflurane	19	0	0	19
Methylprednisolone	17	6	0	23
Metronidazole	434	120	105	659
Minocycline	12	6	8	26
Moxifloxacin	8	0	8	16
Nefopam	128	60	44	232
Neomycin	14	6	0	20
Nitrofurantoin	35	5	7	47
Ofloxacin	0	0	15	15
Ondansetron	135	22	48	205
Oseltamivir	6	12	75	93
Oxytetracycline	1	0	0	1
Oxytocin	121	26	21	168
Phenacetin	34	0	3	37
Posaconazole	0	0	4	4
Prednisolone	424	67	91	582
Rifampicin	5	6	0	11
Riluzole	0	0	6	6
Ritonavir	31	13	11	55
Rizatriptan	18	0	0	18
Roxithromycin	296	55	62	413
Sofosbuvir	56	3	32	91
Sparfloxacin	0	0	6	6
Sumatriptan	71	35	71	177
Tenoxicam	0	2	6	8
Tetrabenazine	0	0	4	4
Tetracaine	25	24	8	57
Tetracycline	64	11	3	78
Tinidazole	13	18	8	39
Tolperisone	0	0	6	6
Trilostane	9	0	0	9
Trimethoprim	39	0	0	39
Vancomycin	10	6	6	22
Varenicline	18	0	3	21
Viminol	197	26	2	225
Voriconazole	3	0	12	15
Zolmitriptan	10	1	13	24

In summary, drug listings from February 2020 to May 2023 were categorised into 342 Level 1 classes, and 25 Level 2 classes. (NB: All NPS drugs were collapsed into one Level 1 category.) Approximately 80% were categorised by the *rule-based-system*, and about 10% were classified by the neural network model with a prediction score of >90%. Of the total extracted listings, about 10% were left uncategorised as they were not categorised from the rule-based system nor did they achieve the 90% prediction score from our machine learning model classification. On excluding the custom and miscellaneous listings (<0.5%), about 90% of the extracted listings were categorised (see section 5.1.2.3 in the methods document of the previous bulletin for further details). For a summary of listing categorisation of historical data before 1st February 2020, please refer to the methods bulletin for cryptomarket listings from February 2014 to January 2020.

Uncategorised listings consist of misspelt drugs, non-drug items, unidentified drugs, and drugs described by street names that have not been included in our database. These listings have been excluded from reporting in the bulletin and online visualisation.

5.2. Deduplication of listings

We deduplicated listings with identical drug listing title and vendor username, and that are located on the same market in the same snapshot. Removing these duplicate listings avoids recounting repeated listings that often arise within a marketplace (promotional offers, server-side errors) or when listing pages are re-scraped after a timeout.

5.2.1. Deduplication of listings for quantity variants

Cryptomarkets vary in their ability to show differing quantities of the same product. Some markets allow for sale of different quantities within the same listing (**Figure 7A**), whereas other markets require separate listings for each quantity of the same drug (**Figure 7B**).

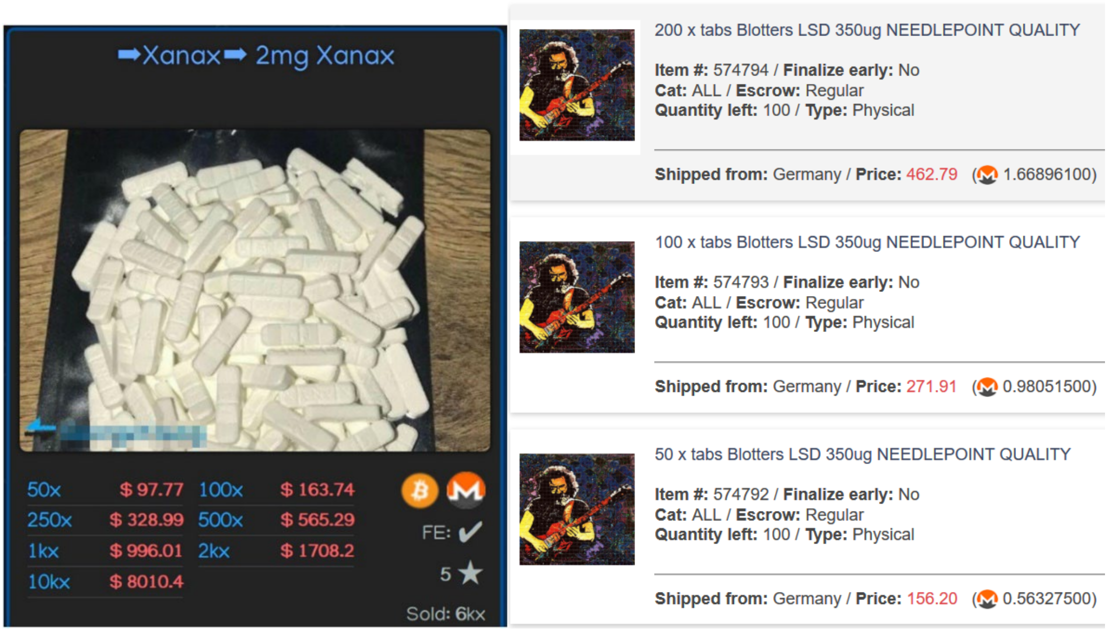
In our previous reporting, we would have identified one listing from the example in **Figure 7A**, and three listings from the example in **Figure 7B**. In this bulletin, we have addressed this issue, and counted different quantities of the same product as a single listing (e.g., both **Figure 7A** and **Figure 7B** would only count as 'one' listing, respectively).

Where multiple listings of the same product varying in quantity have been identified in a given snapshot, the listing of that underlying product with the lowest quantity is kept, and listings of additional quantities are removed. In the examples above, the lowest quantity (i.e. 50x for both listings) would be kept. This removal of quantity variants further reduces the number of listings by about 50% (see **Table 4** in the Methods document of the last bulletin of the year 2022).

DOI: 10.26190/r9tc-ba06

Figure 7. Listing varying quantities of the same product in one listing (left image A), as opposed to listing varying quantities across multiple listings (right image B).

A B



5.3. Creation of twice monthly snapshots from historical data

For the weekly data from before 1st August 2022, twice monthly snapshots of the cryptomarket data were taken from the first week of data in each of the two-week period starting on the 1st and 15th of each month. Where data for the cryptomarket was missing or incomplete, data for that cryptomarket was sourced from the second week of data in the two-week period if available.

6. Data analysis and interpretation of results

We report on the number of listings in a given snapshot as the main metric. This is aggregated by market name, Level 2 category and/or month in our reporting.

6.1. Number of listings

The number of listings is the sum of listings deduplicated for quantity variants (sections 5.2 and 5.2.1) in each snapshot. We can interpret the total drug-related listings over time as indicative level of 'market size' by cryptomarket or by drug class. Percentage of listings for each drug class of the total listings is considered the total 'market share'.

DOI: 10.26190/r9tc-ba06

6.1.1. Average monthly percent change in number of listings by market and by Level 2 drug class and in percentage of listings by Level 2 drug class

In this bulletin, we computed average monthly percent change (AMPC) in number of listings (i.e. market size) as a Poisson count variable and in percentage of listings (i.e. market share) as a proportion using the command line version of the Joinpoint regression program version 4.9.0.0 (National Cancer Institute, 2022). The regression modelling is performed over the 12-months of data from October 2021 to September 2022. Because regular time points are required for a time series model, missing scrapes for number of listings by market were interpolated for the analysis. The model estimates the optimal number of joinpoints (which is the time point at which the rate of change changes) by determining the best-fitting model using the weighted Bayesian Information Criterion (BIC). A maximum of 3 joinpoints is used in the model. Where the number of data points (i.e. number of snapshots) is less than 17, the recommended maximum number of joinpoints is used for the model.

6.1.2. Relative percent difference in percentage of listings of a Level 2 drug between individual markets and overall market

The relative difference in market share of a particular drug (expressed as a percentage of total listings in the market) between an individual market and the overall market is:

$$\frac{\text{Percentage of drug in individual market} - \text{Percentage of drug in overall market}}{\text{Percentage of drug in overall market}} \times 100\%$$

The p-value for the difference is calculated via normal approximation to binomial proportions using the prop.text function in R. Given the multiple comparisons used in these analysis, we used a more strigent threshold of $p<0.01$ in this analysis.

6.1.3. Caveats on analysis of number of listings

- ***Number of listings is only an approximation of total drug availability via cryptomarkets.*** Data on number of listings can be interpreted as an approximation of global drug cryptomarket availability only. These data are limited due to the inability to monitor and identify all active cryptomarkets, and the inability to extract from and crawl markets as soon as they appear. However, as large markets are given priority for monitoring, these figures show reasonable estimates for trends in drug availability on cryptomarkets. It is important to note that drug listings via cryptomarkets likely only comprise a small proportion of the total drug markets, which also include street-based selling, social supply and other digitally mediated drug trading (e.g. app-based). According to the World Drug Report 2019, monthly drug-related revenue of the then eight largest darknet markets amounted to 0.1-0.2 percent of overall drug retail. In addition, we have not identified scam or out-of-stock listings (e.g. some listings with prices that appear to be unrealistically high are not excluded).
- ***Inferences of sale volume cannot be made from the number of listings.*** The number of listings does not capture any information regarding the number of sales of any listing, and therefore, they cannot be translated to any metric that reflects the sale volume of a market or specific drug. These metrics capture the availability of a certain drug only on a specific market. To gain information regarding the volume of actual purchases, other features such as customer feedback and vendor ratings must be analysed. Due to the varying nature of what is visible on differing markets and our limited scope of data extraction, a universal analysis can only be framed in terms of the number of listings. We refer the reader to Ball et al., 2019 for scraping procedures that implement more exhaustive crawling, and Aldridge and Décary-Hétu, 2016b for analyses of transactions through the proxy of feedback comments and/or drug listing ratings, and we acknowledge the broader literature on these topics (too extensive to list here).

6.2. Number of vendors

The number of vendors is the sum of unique vendors observed in a snapshot selling a specific drug category. For this measure, a vendor is considered unique only within the same market only; that is, the same vendor may be counted multiple times across different markets. This measure maintains the same interpretability as counting the number of listings. For conciseness, the bulletin does not report on number of vendors but it is presented in the interactive online visualisation.

6.3. General caveats to interpretation of findings

- ***Categorisation of listings may be subject to fallacy.*** While 92% of historically collected data has been categorised and included in reporting, we cannot guarantee perfect accuracy due to human errors in our drug database (e.g. phrases that may be shared between two different drugs). Moreover, we cannot guarantee no misclassifications from our predictive model, however this is limited by our predictive accuracy threshold of 90%. We have excluded listings unable to be categorised from our presentation of results. We are currently updating our drug categorisation dictionaries by extracting terms from our uncategorised listings and matching them to our current and new data sources on NPS, pharmaceutical drugs and their brand names, and names of cannabis strains. The categorisation algorithm is also continually improved on for future reporting.
- ***Inferences regarding illicit manufacture of medicines cannot be captured from our findings.*** We cannot distinguish between illicit and pharmaceutically manufactured substances (e.g., benzodiazepines produced by legitimate pharmaceutical companies to pharmaceutical standards versus those manufactured illegally or tampered with). Occasionally, this information can be deciphered from available information contained in listing title (e.g., determining whether fentanyl is sold in the form of a transdermal patch) however, it is not feasible currently to implement this kind of finer processing for every drug.
- ***Findings reflect purported substance content as opposed to objective information on contents.*** We can only categorise substances based on the information contained in a drug listing. The advertised contents may or may not align with the actual contents (e.g., drug listed as MDMA may actually contain MDA only, or a mix of both MDMA and MDA).
- ***Time series may be variable due to market fluctuations.*** Time series for certain markets may exhibit large variation in the short-term due to market discruption or instability. Statistical smoothing has not been applied to remove these abrupt phenomena from reporting, however we hope to integrate this in future reporting.
- ***There are missing data***. The reported data contains missing data where complete crawling was not undertaken. Most often, this is due to difficulties with accessing the market (e.g. DDoS attack).

DOI: 10.26190/r9tc-ba06

Glossary

Term	Definition/Description
Anatomical Therapeutic Chemical (ATC)	A unique code assigned to a medicine according to the organ or system it works on and how it works. The classification system is maintained by the World Health Organization (WHO).
Average monthly percent change	The relative percent change in number or percentage of listings per month estimated using the Joinpoint regression program version 4.9.0.0 (National Cancer Institute, 2022) (see **section 6.1.1**).
Beautiful Soup	A Python package for parsing HTML and XML documents.
Cryptocurrency	Digital currency in which encryption techniques are used to regulate the generation of units of currency and verify the transfer of funds.
Cryptomarket	An online marketplace that facilitates the purchasing of illicit goods and services via multiple sellers, and provides its users with anonymity via its location on the hidden web.
Hidden web (aka darknet)	Encrypted, anonymous services built on the Tor Internet service and similar services that are not indexed by conventional search engines.
Denial-of-service attacks (DDoS)	Cyber-attacks in which the web-service is attacked with superfluous requests in an attempt to overload server systems and prevent some or all of the legitimate requests from being fulfilled.
Keras	An open-source neural-network library written in Python.
Market share	This is defined as the percentage of drug-related listings by drug class.
Market size	This is defined as the number of drug-related listings per snapshot, overall (i.e. in the total market), by market or by drug class.
Number of listings	Sum of listings observed in a snapshot, belonging to a specific market or drug class. For this measure, identical listings and their quantity variants from the same vendor and market within the same snapshot are removed (see **section 5.2**).
Number of vendors	Sum of vendors observed in snapshot selling a specific drug category. For this measure, a vendor is considered unique only within the same market only; that is, the same vendor may be counted multiple times across different markets.
Regular expressions	A special sequence of characters that helps you match or find other strings or sets of strings, using a specialized syntax held in a pattern.
Scrapy	A package integrated into Python for parsing HTML and XML documents.
Selenium	A free (open source) automated testing suite for web applications across different browsers and platforms.
Snapshot	A snapshot of the cryptomarkets is taken on a twice monthly basis, in the two weeks starting on the 1st and 15th of each month.
Surface web	Internet content that can be accessed through search engines.
Tor Network ('The Onion Router')	An open source privacy network that permits users to browse the web anonymously by re-routing their IP address through various encrypted nodes.
Word embedding	A learned representation for text where words that have the same meaning have a similar representation.

Funding and Copyright

Recommended Citation

Man N, Linghu Q, Bruno R, Barratt MJ, Sutherland R, Peacock A. Trends in the availability and type of drugs sold on the internet via cryptomarkets, June 2022 - May 2023. Drug Trends Bulletin Series. Sydney: National Drug and Alcohol Research Centre, UNSW Sydney; 2023. Available from: http://doi.org/10.26190/r9tc-ba06

Acknowledgements

We would like to acknowledge the following individuals for their contribution to this project: Anant Mathur, Max Pedersen, Rajat Katyal, Dr Amanda Roxburgh, Joe van Buskirk, A/Prof Lucinda Burns, A/Prof Timothy Dobbins, Dr Courtney Breen, Sundresan Naicker and Rosie Swanton.

Related Links

- Data visualisations: https://drugtrends.shinyapps.io/cryptomarkets
- For more research from the Drug Trends program go to: https://ndarc.med.unsw.edu.au/program/drug-trends