METHODS FOR THE ANALYSIS OF TRENDS IN THE AVAILABILITY AND TYPE OF DRUGS SOLD ON THE INTERNET VIA CRYPTOMARKETS

Authors: Anant Mathur, Raimondo Bruno, Nicola Man, Monica J. Barratt, Amanda Roxburgh and Amy Peacock
National Drug and Alcohol Research Centre
University of New South Wales, Sydney

1. General Note

Data presented in the public online interactive visualisation and reported in bulletins are obtained from the weekly 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 darknet markets, also known as cryptomarkets. 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. 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).

Funded by the Australian Government Department of Health under the Drug and Alcohol Program.

2. Definition of 'Cryptomarkets'

Cryptomarkets ('darknet markets') are anonymous online trading platforms that facilitate the purchasing of illicit goods and services via multiple sellers. 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**).

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.

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.

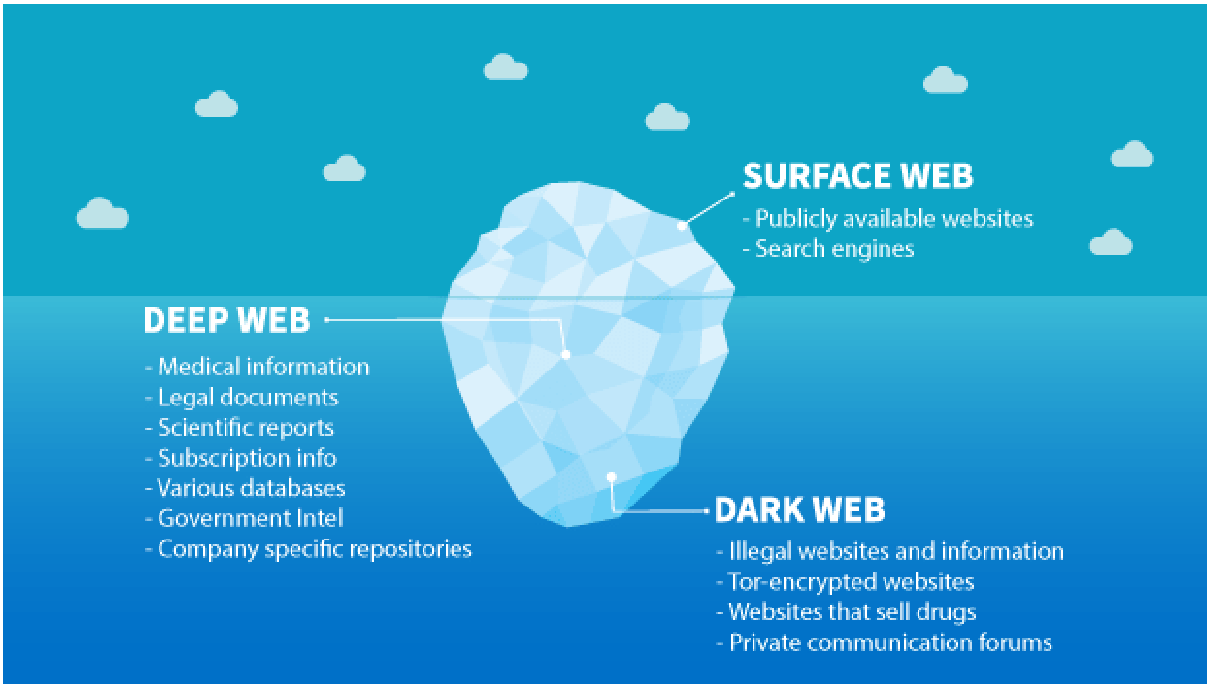

Figure 1. Description of the content available on the internet. For further information, see Barratt & Aldridge (2016). Image: vpnoverview.com

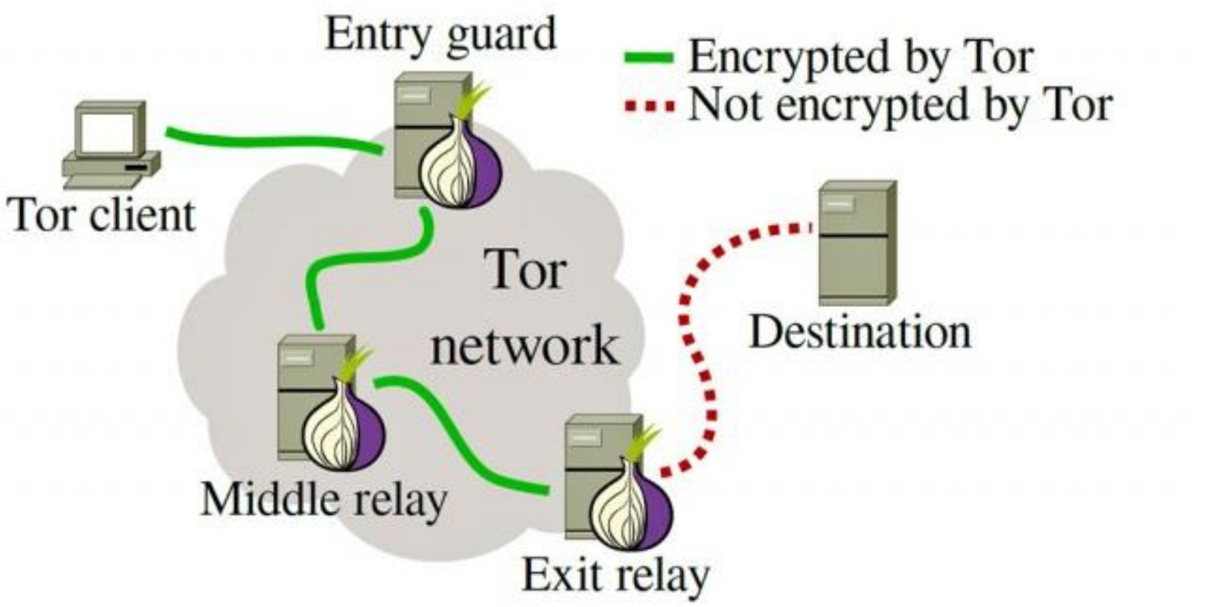

Figure 2. Tor network structure. Image: The Conversation

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).

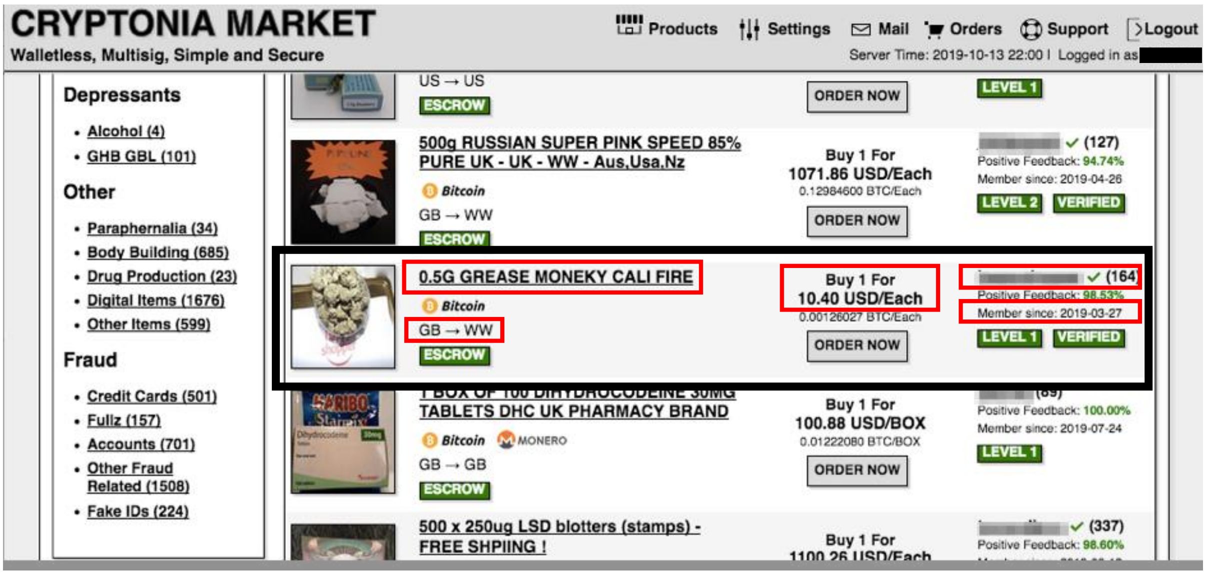

Figure 3. The landing page from 'Cryptonia Market' (2019)

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.

For a historical record of marketplaces monitored by DNeT, we refer the reader to our interactive timeline, where the start and end dates refer to the first and last time points of data collection for that market.

3. Current Methods of DNeT

3.1 Overview of approach

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 relation to the drugs that are listed for sale; and (3) classifying these listings into specific drug categories.

It is important to caveat that previous reporting on drug listings on cryptomarkets from DNeT (i.e., in 2017 and earlier; available here) were based on different approaches to categorisation and reporting. For this reason, reporting may not be comparable over time. We have applied the below automated approaches to historical data: for this reason, we recommend referring to most recent data available in the public online interactive visualisation. The exception to consistency in processes used over the course of the project comprises the method of crawling the cryptomarkets: **prior to 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.**

[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.

Categorisation was achieved via a dynamic lookup table compiled using previously identified terms and their associated categories.

This project has institutional approval and ethical approval from the University of New South Wales Human Research Ethics Committee (HC180004). Work on this project was conducted on a dedicated computer and a record was kept of when staff engaged with the hidden web. Due to anonymised online details, no individual results were gathered and data are published in tabular, aggregate form only.

3. 2 Identification and crawling of cryptomarkets

Active cryptomarkets and their Tor links are identified through surface-web websites such as darkfail.net and the recently defunct deepdotweb.com. 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.

Marketplaces are included in monitoring if they:
- Facilitate trade of illicit goods and services via multiple sellers;
- Display more than one hundred drug listings;
- Are displayed in the English language.

In order to regularly access a cryptomarket, a dedicated account is created upon its initial monitoring. Since August 2018, custom web crawling 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 certain markets (e.g., Dream Market), crawling was 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 weeks, the market is assumed to be defunct.

Automated crawling is conducted on a weekly basis (monitoring is ongoing) on a stand-alone computer, and can take minutes to hours to complete, 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.

3.3 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.

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. These include the drug listing name, vendor pseudonym and the price in bitcoin or dollars. **Figure 3** shows an example listing page with the features to be extracted outlined in red.

3.4 Drug classification

3.4.1 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, and categories conflict across marketplaces. Drug listings are also often miscategorised (e.g., cannabis listings within a cluster of listings categorised as 'stimulants') due to server-side issues. 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. 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).

Chemical substances 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. 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. In such cases, the therapeutic level was chosen such that it would align best with the effect if the substance was consumed extra-medically (i.e., outside the bounds of a doctor's prescription). A detailed description of ATC structuring can be found here. See **Figure 4** for an example classification according to ATC.

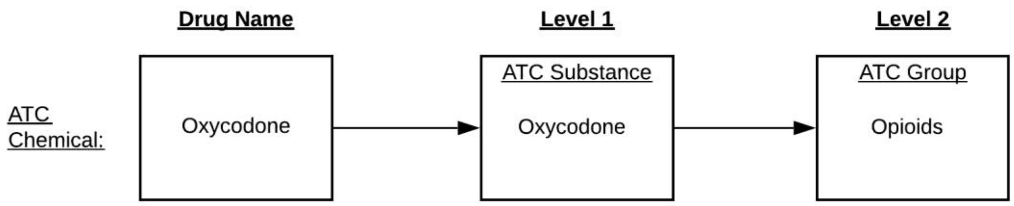

Figure 4. Example categorisation structure for ATC chemicals.

There are two primary exceptions to our two-tier hierarchy for Level 2 drug categories: performance and image enhancing drugs (PIEDs) and weight loss substances, as well as new psychoactive substances (NPS).

3.4.1.1. Performance and image enhancing drugs (PIEDS) and weight loss substances. 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', whereas the steroid oxandrolone is found in the ATC group 'Anabolic steroids', 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**).

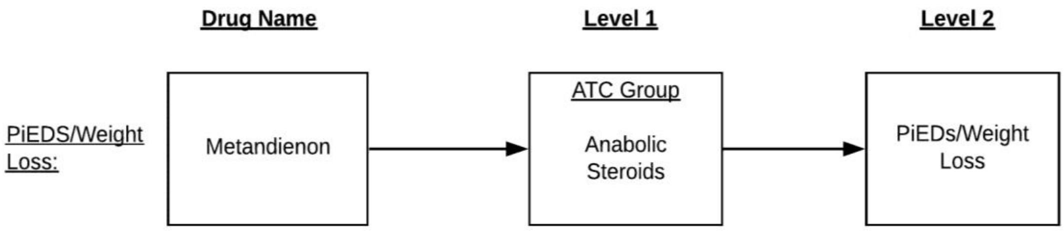

Figure 5. Example categorisation structure for PIEDs/Weight loss drugs.

3.4.1.2 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.

Consequently, we opted to specify sub-classes at Level 1 rather than the drug name, comprising: cannabinoids, empathogens-entactogens, hallucinogens/dissociatives/psychedelics, sedative-benzodiazepines, sedative-opioids and sedative-hypnotics[2]. These sub-classes were then grouped to the broader class 'NPS' at Level 2. For example, the drugs AB-PINACA and AB-CHMINACA were grouped as 'cannabinoids' at Level 1, then 'NPS' at Level 2 (see **Figure 6**). 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 greater disaggregation of specific NPS in the public online interactive visualisation in future. Please contact us (drugtrends@unsw.edu.au) if you have queries regarding results for a specific substance.

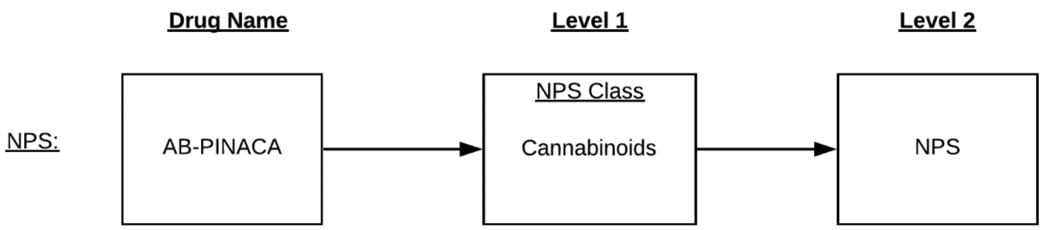

Figure 6. Example categorisation structure for NPS.

[2]Currently NPS are not reported at this level of disaggregation but we aim to do so in future reporting.

3.4.1.3 Other non-standard categories. Additional Level 2 classes were created to accommodate non-standard drug listings such as drug paraphernalia and precursors. The 'drug paraphernalia' class contains injecting equipment, vaping equipment, drug checking equipment, naloxone and other paraphernalia. Custom-made listings that are intended to be ordered by specific buyers were given their own Level 2 class as no information was made available pertaining to the drug being sold. Moreover, listings where multiple drugs are identified e.g. "MDMA plus free hash" were classified as 'other drugs' at Level 2.

Table 1 provides a complete list of drug classes chosen for reporting in the online visualisation and bulletin. ATC codes for Level 2 have been included.

Table 1. Grouped drug classes chosen for reporting in the online visualisation

Drug Class	ATC Code	Description
Alcohol	-	-
Cannabis	-	All forms of cannabis (plant, oil, seeds, etc)
Cocaine	-	-
DMT	-	DMT only, excluding plant sources
E-cigarette	-	-
GHB/GBL/1,4-BD	-	GHB or GBL or 1,4-BD
Hallucinogenic mushroom	-	-
Heroin	-	-
Ketamine	-	-
LSD	-	-
MDA	-	-
MDMA	-	-
Meth/amphetamine	-	Any illicit amphetamine or methamphetamine (includes speed). This excludes substances identified as pharmaceutical stimulants (see below).
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
Paraphernalia*	-	Drug checking equipment, injecting equipment, naloxone, vaping equipment and other paraphernalia
PCP	-	-

Drug Class	ATC Code	Description
PIEDS/weight loss*	-	Performance and image enhancing drugs (e.g., anabolic steroids, androgens) and weight loss products
Tobacco	-	
Psychostimulants and nootropics*	N06B	Pharmaceutical stimulants (e.g., dexamphetamine, methylphenidate, modafinil)
Benzodiazepines*	N05BA	Benzodiazepines (e.g., alprazolam, diazepam). Note that certain novel benzodiazepines (e.g., etizolam) have been classified as NPS
Opioids*	N02A, N07BB	Pharmaceutical opioids (e.g., oxycodone, fentanyl etc), drugs used in addictive disorders (e.g., methadone) and opium. Note that certain novel synthetic opioids (e.g., acetylfentanyl) have been classified as NPS
Hypnotics and sedatives*	N05CM, N05CA	Hypnotics and sedatives (e.g., methaqualone, clomethiazole) and barbiturates (e.g., pentobarbital, phenobarbital, thiopental)
Inhalants*	-	Alkyl nitrites and nitrous oxide
Other medicines*	J05AP, G04BE, M03, R, N05A, N06A, N01A, N03A, N02B, N05B, N06DX, N01B, N02C, H02, S01, D04A	Antivirals, erectile dysfunction drugs, muscle relaxants/anti-inflammatories, respiratory system drugs, antiparasitic products, antipsychotics, antidepressants, general and local anaesthetics, antiepileptics, other analgesics and antipyretics, anxiolytics, anti-dementia drugs, corticosteroids, ophthalmologicals, antihistamines
Other drugs	-	Precursors (e.g., ephedrine) and drug listings with multiple drugs (e.g., “Cocaine+Heroin”)
Categories excluded from reporting	-	Miscellaneous non-drug items (e.g., hotel listings) and custom drug listings (custom-made listings intended to be ordered by specific buyers)

Note: *Indicates drug classes containing multiple Level 1 drugs which we hope to report on in future. Refer to description column for details of included Level 1 drugs. Note that there were originally 46 Level 2 categories however, for the purpose of interpretability in displaying findings, we have further clustered to reduce the number of classes.

3.4.2 Applying categorisation of listings

In order to apply this categorisation to the extracted listings, we classify the text contained in the listing name to a Level 1 drug. To ensure the largest possible number of listings have been categorized with a high degree of confidence, we have implemented an automated classification procedure comprising of two parts.

The first component consists of a *rules-based system* that checks each listing name for the appearance of certain phrases or words which have been previously mapped to Level 1 by a human.

The second component consists of a *machine learning classifier* that has been trained on historically categorized listings (see **Figure 7** and **Figure 8**). The machine learning model attempts to classify the listing names that were unable to be categorized by the rules-based system. The machine learning component was deemed necessary as (1) it is not feasible to manually categorize thousands of un-matched listings on a weekly 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.

To apply the rules-based system, individual words contained in the listing name 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 name), 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 name '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.

Matches to two or more Level 1 drugs occurred for on average 6% of historical listings from February 2014 to January 2020, as an example. The most common double matches were inspected and, while some listings did contain multiple drugs as expected, the majority of cases were listings where extra rules were required to set the correct drug at Level 1.

The most common issue arose from drug names nested within longer drug names, e.g., ‘5-MeO-DMT’ and ‘DMT’. Other less common issues include branded MDMA pill names found in the listing name of non-MDMA drugs, and drug-specific equipment containing both a drug name and equipment keyword, e.g., ‘cannabis vaping equipment’.

To correct for these duplicate cases, the following overriding rules were scripted:

- For drug names nested within longer drug names, the longer drug name is assigned to the listing (e.g., ‘5-MeO-DMT’ is assigned to 5-MeO-DMT not DMT).
- Logo or drug shapes that have been assigned to MDMA are overridden by the presence of another drug name, (e.g., “Tesla 2-CB” will be overruled as 2C-B).
- Custom listings containing a drug name (e.g., ‘cocaine custom listing’) are assigned as custom listings (see **Table 1**).
- Vaping equipment is assigned as the Level 1 class when the listing contains another drug name (e.g., “THC e-cig” is assigned to vaping equipment).

These rules were automatically implemented, bringing the percentage of listings with two or more matches to less than 1% from February 2014 to January 2020, as an example. Due to significant number of unique listings that were resolved by the above rules, the remaining percentage was left uncategorised.

The remaining 20% of listings from February 2014 to January 2020 that could not be matched to our database were classified at Level 1 through a long short-term memory (LSTM) artificial neural network. Listing names that were classified by this predictive model with a target percentage 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 date to January 2020.

The LTSM neural network was trained on a set of 5 million unique categorized listings from February 2014 to January 2020 that were labelled correctly by our rules-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 names 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 names. These were achieved with a word2vec model programmed in Python that embedded the contextual and sequential relationships found in the various drug listing names into transformed numerical inputs. The subsequent training and development of the LTSM 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.

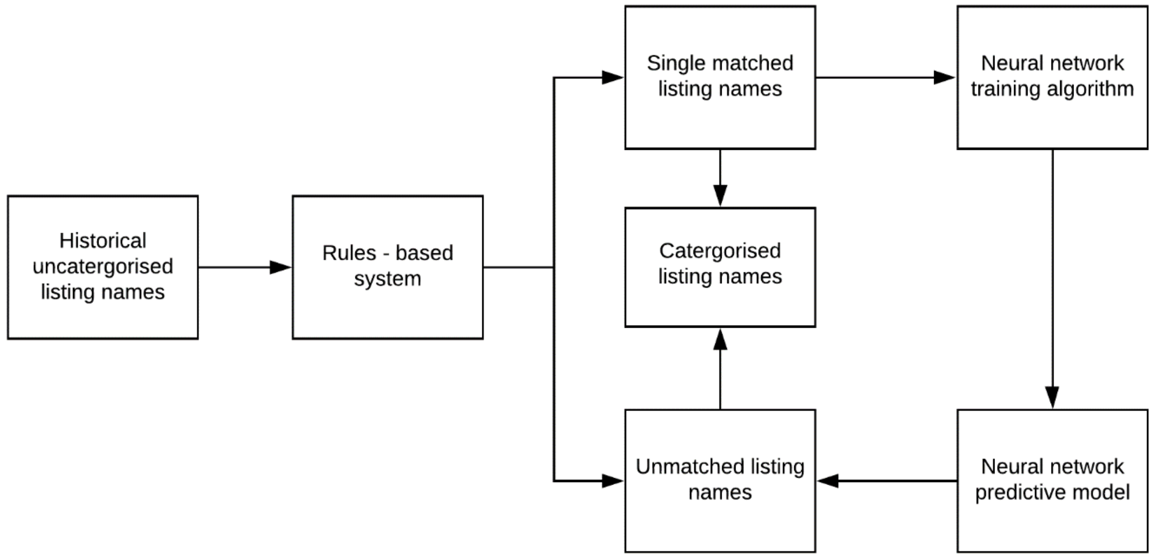

Figure 7. Categorisation process applied to historical data.

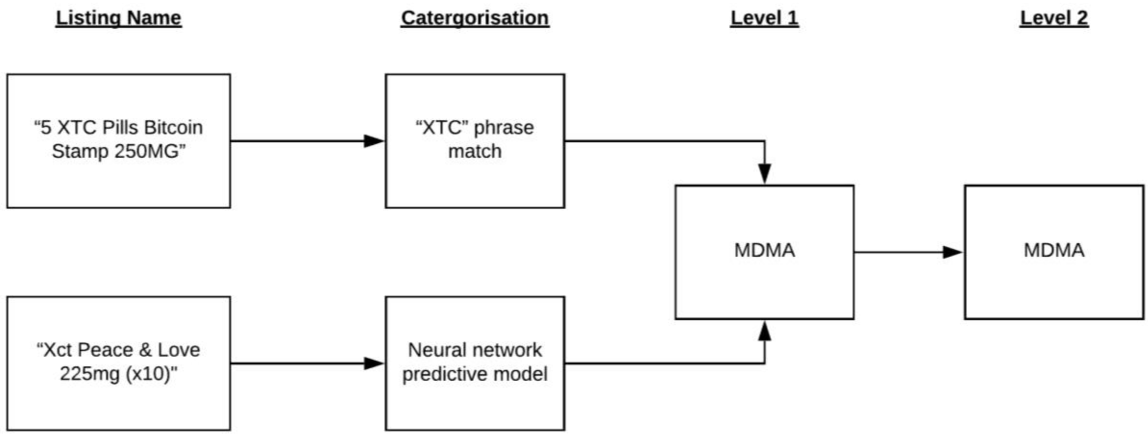

Figure 8. Example of categorisation process applied to two MDMA listings.

In summary, drug listings from February 2014 to January 2020 were categorized into 608 Level 1 classes, and 46 Level 2 classes. Of these listings, 79% on average were categorized by a rules-based-system, and 20% on average were attempted to be classified by a neural network model. Of those total listings, on average 7% were left uncategorized as they did not achieve the 90% prediction score by our machine learning model. If we add the approximately 1% of the unaccounted listings with two or more matches, 8% or 1,098,240 listings in total were left uncategorized (**Table 2**).

Table 2. Categorisation breakdown for historical listings from February 2014 to January 2020.

Categorised listings (92%)	
Rules-based system	79%
Predictive model	13%
Uncategorised listings (8%)	
Unaccounted listings with two or more matches	1%
Listings that fell below the 90% predictive score threshold	7%

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 bulletin and online visualisation.

Uncategorised listings will be reviewed on a regular basis and searched for new drugs. If a new drug is identified, then the drug database will be updated, such that future listings of that drug can be categorised by the rules-based system. These new listings will be appended to the historical categorised training set, 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. Once retrained, the updated predictive model will not be re-applied to historical data.

4. Data Analysis and Interpretation of Results

Data in the public online interactive visualisation and bulletin have been disaggregated by Level 2 class and market name. The data can be viewed through two metrics: the number of listings observed in given week or the number of unique vendors observed in a given week.

[3]Currently, regular review of uncategorised listings not undertaken. This review is to be implemented after first release of bulletin.

4.1 Number of drug listings as an outcome

The number of listings is the sum of listings observed in each weekly crawl, belonging to a specific market or drug class. For this measure, duplicate listings within the same weekly scrape are removed. Duplicate listings in this case are defined as listings with identical names that are sold by the same vendor, contain the same quantity of drug, and are located on the same market. Removing these duplicate listings (as in the current case) avoids recounting repeated listings that often arise within a marketplace (promotional offers, server-side errors). **We can interpret total drug-related listings over time at the market level as indicative level of 'market size', while number of listings over time at the drug level is indicative of the availability of that drug. Percentage of listings for each drug of the total listings is considered the total 'market share'.**

4.2 Number of vendors as an outcome[4]

The number of vendors is the sum of unique vendors observed in each weekly crawl, 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 is advantageous as it does not over represent cases where a drug sold by a single vendor is listed multiple times with different quantities. This measure maintains the same interpretability as counting the number of listings; however, as a single vendor may list multiple drugs, the sum of unique vendors listing two or more drugs will be often greater than the true value. For conciseness, only figures displaying number of listings are included in bulletins.

4.3 Measures of change over time

Our regular reporting includes statistics that compare the change in drug listings observed in a month to that of a month one or twelve months prior. These statistics include:

- The change in total *market size* (either across all markets or for a single market). This is expressed as the percentage change in the average **number** of weekly listings in the months of interest (e.g., December versus November).
- The change in *market share* for a drug class (either across all markets or for a single market). This is expressed as the change in **percentage** of total listings for each drug class, across all marketplaces in the months of interest (e.g., December versus November).
- The change in the number of listings for a drug class (*market size of* drug class). This is expressed as the relative change in the average **number** of listings observed per week, for each drug class, in the months of interest (e.g., January 2019 versus January 2020).

[4] Currently, number of vendors as an outcome have been excluded from the online visualisation. This outcome may be included in future reporting.

Note, interpolated data has been used to calculate these values in order to avoid discrepancies occurring from missing data (see below for information on interpolation).

4.4 Key caveats to interpretation of findings

- ***Data are only an approximation of total drug availability via cryptomarkets.*** Certain figures contained in the online visualisation display the total number of listings and vendors observed on all monitored cryptomarkets at a given time. These totals 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.

- ***Inferences of sale volume cannot be made from the presented data.*** The number of listings and number of unique vendors 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 or the number of vendors. 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).

- ***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 name (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).

- ***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 threshold percentage of 90%. We have excluded listings unable to be categorised from our presentation of results.

- ***Some weekly data points reflect averages from multiple collections.*** During the period of earlier data collection before fully automated crawling, multiple data collection points may have been taken within the same week. In such cases, the number of listings and number of unique vendors for each drug have been averaged across the multiple collections in the week.

- ***Time series may be variable due to market fluctuations.*** Time series for certain markets may exhibit large variation from week to week due to server-side issues or market 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. Data has been imputed between these time points by linear interpolation. Interpolated data has been indicated by dashed lines in the online visualisation and in bulletins.

- ***Reporting on specific substances will be expanded in future.*** The online visualisation currently includes only Level 2 broader categorisation, and not the finer-detailed drugs at Level 1. We hope to include greater disaggregation by drug type in future reporting.

4.5 Reporting on findings

Bulletins summarising most recent data captured in the previous year will be released on a quarterly basis and the online visualisation will be updated on a quarterly basis with minimal lag (where possible) since last scraping. We have also released a summary output which provides the timeline of markets monitored and key findings from February 2014 to January 2020.

Glossary

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 anonymous online trading platform that facilitate the purchasing of illicit goods and services via multiple sellers.
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.
Number of listings	Sum of listings observed in a week, belonging to a specific market or drug class. For this measure, duplicate listings (defined as listings with identical names and same quantity of drug by a single vendor on a single market) within the same week are removed.
Number of vendors	Sum of unique vendors observed in week 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.
Selenium	A free (open source) automated testing suite for web applications across different browsers and platforms.
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

The National Drug and Alcohol Research Centre (NDARC) and the Drug Trends program are funded by the Australian Government Department of Health under the Drug and Alcohol Program.

Acknowledgements

We would like to acknowledge the following individuals for their contribution to this project: Joe Van Buskirk, A/Prof Lucinda Burns, Dr Courtney Breen, Sundresan Naicker and Timothy Dobbins.

Recommended citation

Mathur, A., Bruno, R., Man, N., Barratt, M. J, Roxburgh, A., Van Buskirk, J., & Peacock, A. (2020). Methods for the analysis of trends in the availability and type of drugs sold on the internet via cryptomarkets. Drug Trends Bulletin Series. Sydney: National Drug and Alcohol Research Centre, UNSW Sydney.

Related Links

- Data visualisations: https://drugtrends.shinyapps.io/cryptomarkets
- Drugs and New Technologies (DNeT) project: https://ndarc.med.unsw.edu.au/project/drugs-and-new-technologies-dnet
- For more information on NDARC research, go to: http://ndarc.med.unsw.edu.au/
- For more information on the ATC classification system, go to: https://www.whocc.no/atc/structure_and_principles/
- For more research from the Drug Trends program go to: https://ndarc.med.unsw.edu.au/program/drug-trends

Contact us

Email: drugtrends@unsw.edu.au