# Detecting Illicit Drugs on Social Media using Automated Social Media Intelligence Analysis (ASMIA)

Paul A. Watters[1] and Nigel Phair[2]

[1] Internet Commerce Security Laboratory (ICSL), University of Ballarat, Ballarat, Australia
p.watters@ballarat.edu.au

[2] Centre for Internet Safety, University of Canberra, Canberra, Australia
phair@suretegroup.com.au

**Abstract.** While social media is a new and exciting technology, it has the potential to be misused by organized crime groups and individuals involved in the illicit drugs trade. In particular, social media provides a means to create new marketing and distribution opportunities to a global marketplace, often exploiting jurisdictional gaps between buyer and seller. The sheer volume of postings presents investigational barriers, but the platform is amenable to the partial automation of open source intelligence. This paper presents a new methodology for automating social media data, and presents two pilot studies into its use for detecting marketing and distribution of illicit drugs targeted at Australians. Key technical challenges are identified, and the policy implications of the ease of access to illicit drugs are discussed.

**Keywords:** illicit drugs; social media; open source intelligence.

## 1 Introduction

Much of the focus in cybersecurity research is on technology enabled crimes, which were previously unknown before the widespread adoption of the Internet1. These types of crimes are those which are dependent on the new technologies, such as distributed denial of service attacks, which have no direct correlate outside the Internet [2]. On the other hand, many traditional crime types have made extensive use of the capability provided by the Internet, making it easier to commit such crimes. These technology enhanced crimes include various types of scams, such as 419 and employment scams, whose distribution previously was only limited by the cost of postage [3]. With the rise of free e-mail and often free Internet access, the cost of sending scam emails is effectively nil.

While traditional crimes have adapted to the new technology, the technologies themselves have also been increasing in scope and reach. The advent of "Web 2.0" technologies, including social media - often accessed by mobile devices - has

[1] See Stabek et al [1] for a classification scheme.

radically transformed the user mix and the functionality available to users. It has also opened up new avenues for security threats to emerge and spread: the Koobface worm is a great example of this [4].

In this paper, we consider how social media technologies could be used to enhance criminal activity, and develop a generic approach to implement an intelligence based approach to understand the extent and nature of the threat. As an example, we investigate how illicit drugs are being quite openly traded through social media. This is in contrast to recent reports that much of this activity is being carried out through layers of anonymisation [5]; as we will see, the hallmarks of Internet based crime - such as fostering threats outside of the target jurisdiction, while servicing customers internationally [6] - are all present with the use of social media in the drugs trade.

In addition to understanding how to automate the gathering and analysis of source material from social media, which may indicate trading in illicit drugs, we further consider how to more actively identify individuals and criminal organisations who may be targeting a specific country. In this sense, we propose to use social media connectedness - which is an asset to organised crime groups - as a means to identify those who are likely to be buying or selling illicit drugs locally [7].

In the following sections, we propose a methodology - known as Automated Social Media Intelligence Analysis (ASMIA) - which we then apply to two different but related problems in understanding how social media is being utilised in illicit drugs trade, especially advertising and distribution. Although ASMIA is used to manually process open source intelligence in the results of the experiments presented in this paper, there is great potential for each stage to be automated, in order to assist analysts in carrying out investigations. The potential for applying data mining techniques to automate the processing of source material has been widely recognized [8].

## 2 Automated Social Media Intelligence Analysis (ASMIA)

ASMIA operates by either passive monitoring or active searching of a number of sources, using query terms which are constructed from noun phrases and verb phrases which are then combined to form search terms. Such a simple approach works well for English, since the role of case – where noun forms are changed to denote its grammatical role – is limited [9].

All terms are then added to a term list, which can then be either monitored or searched for actively. When a term returns a hit, this may relate to a target such as an individual, organisation, drug etc from which entities can be extracted [10] and from which or reasoning can be based, using an expert system or similar approach [11]. The algorithm can be described as follows, with examples related to illicit drug trade detection:

1. Identify data source(s) to be monitored for the presence of one or more terms.

EXAMPLE: Google ads, Facebook ads, Google search, Facebook profiles, Pipl etc

2. Determine whether the data source is suitable for passive monitoring or active searching
   EXAMPLE: A sample of Google ads can be harvested by repeatedly refreshing a search page where a relevant term has been entered.

3. Develop a term list which can be used for searching or monitoring.
   EXAMPLE: List of illegal drug names.

4. A term list may comprise a list of targets (individuals, groups, drugs/chemicals) and their online attributes.
   EXAMPLE: List of persons of interest, known or suspected to be involved in selling drugs online, and links to their Facebook page.

5. A term list may be either a simple list or be derived from an ontology or a directed graph.

6. An ontology is a tree which is used to represent knowledge in a form which can be used for reasoning and inference.
   EXAMPLE: An ontology could be constructed showing the possible precursor pathways for all illegal drugs, and these could be added to the term list. Ontologies already exist for legal drugs (eg, http://rxnav.nlm.nih.gov/)

7. A directed graph can be used to represent the relatedness of entities with a weight indicating the strength of the relationship between two entities, which can then be used for inferences.
   EXAMPLE: A group of drug dealers may all ultimately be linked by a common chemical company they use to purchase precursors through several intermediaries.

8. An expert system or reasoning engine to use ontologies and/or directed graphs to make inferences.
   EXAMPLE: An expert system could reason that someone seeking to buy sulfaminic acid and anthranilic acid (both available from alibaba.com) might be aiming to produce ice.

9. Analyse, process and report on prevalence, at some regular internal, the relationships between entities and the scale of the activity.
   EXAMPLE: 1,000,000 Google ads seeded with terms from the term list could be used to report on prevalence on blatant advertising for illicit drugs on an annual basis.

ASMIA uses a pipeline architecture, as shown in Figure 1, with each component being encapsulated, receiving input from the previous stage and passing output to the next stage. Thus, it is very amenable to each component being enhanced without

modifying the architecture at all. This means, for example, that a verb phrase processing component could be enhanced to include stemming, which might improve the accuracy of an ontology-based search, without affecting the activity of other components (other than to seek overall improvements in accuracy and precision).

In the rest of the paper, we present the results of two experiments, both using active search and a range of search terms that are of particular relevance to the illicit drug trade targeting Australians. The preliminary results are then discussed in terms of their relevance for law enforcement, as well as the broader theoretical issues which arise from the implementation of ASMIA.

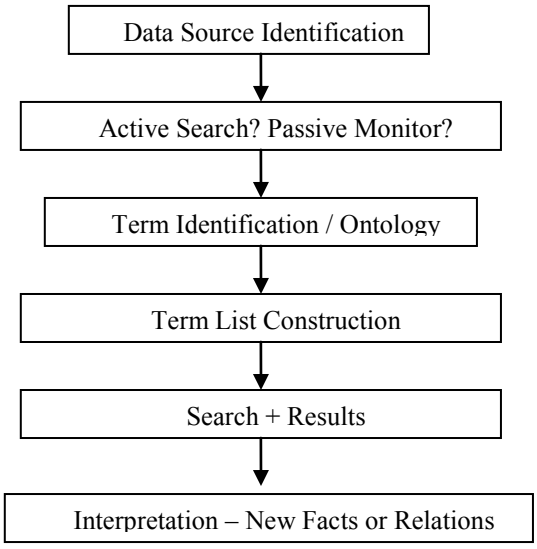


**Fig. 1. ASMIA Pipeline Architecture**

## 3 ASMIA Experiment 1 - Ecstasy and Ketamine Distribution and Advertising

To explore prevalence, Google search was selected as the data source, and active monitoring was used. Two synthetic drugs were selected to represent a range of prevalence in the drug marketplace, being ecstasy and ketamine, which have been used by 9% and only 1% of the Australian population respectively [12]. Street names for these two drugs were identified from the American Council for Drug Education, comprising [Roll, XTC, Adam, and X] for ecstasy and [Special K, Vitamin K] for

ketamine, which were all used as noun phrases (without modifiers)[2]. Verb phrases were selected from synonyms associated distribution [acquisition, bargain, closeout, deal, good deal, investment, purchase, steal, value] or selling [advertise, auction, bargain, barter, be in business, boost, clinch the deal, close, close the deal, contract, deal in, dispose, drum, dump, exchange, handle, hawk, hustle, market, merchandise, move, peddle, persuade, pitch, plug, puff, push, put across, put up for sale, retail, retain, snow, soft sell, soft soap, spiel, stock, sweet talk, trade, traffic, unload, vend, wholesale] as obtained from thesaurus.com. The term list was then searched, using boolean operators, and potential avenues for advertising and distributing synthetic drugs were identified in the results.

There was a wide variety of results, but they could be broadly categorised as described in the following sections.

### 3.1 True Positives

Actual offers to buy or sell illicit drugs were identified on many occasions. For example, searching for "buy ketamine OR sell ecstasy", for example, returned a link that provided a list of illicit drugs (primarily synthetic) from an Australian classifieds website, which were listed using both street names and full chemical names from "the greatest suppliers of Mephedrone, Heroin, Ketamine, methylone, oxycontin e, MDMA, MDVP". Drugs for sale included (verbatim):

- 4-MEC
- Mephedrone (4-MMC)
- Ecstasy ( Such as : Sky,Gum,Go,Cros,XL)
- Bulytone (bk-MBDB)
- MDAI
- Analgesic Chemical CB1 and CB2
- CP 47 497
- CP-55 940
- HU-210
- HU-331
- Ephedrine HCl powder
- JWH-018 / JWH-200 JWH-250
- TFMPP
- 2C-E, 2C-I, 2C-P, 2C-B, 2C-T-2
- DOC, DOI

[2] http://www.acde.org/

- Bromo DragonFly
- TCB-2
- 5-Meo-DMT
- 4-Aco-DMT
- 4-Ho-MIPT
- 4-Meo-PCP
- Naphyrone
- Heroin.
- Methylone (bk-MDMA)
- BENZO anger.
- Pure Magic,
- Bonsai,
- Smoke
- Chocolate,
- Special New Formula.
- Special gold.
- Methedrone (BK-PMMA, Methoxyphedrine)
- Fluoromethamphetamine 2-(2-FMA)
- Pyrrolidinopropiophenone (a PPP).
- MDPV (Methylenedioxypyrovalerone, MDPK)
- Testosterone
- JWH-073, 1-butyl-3-(1-naphthoyl) indole
- Hydrocodone
- Dimethocaine (Larocaine / DMC)
- Morphine
- JWH-018, 1-pentyl-3-(1-naphthoyl) indole
- Herione
- Fluoroamphetamine 4-(4-FA, 4-FMP, or flux)
- Ketamine

An e-mail address was provided, location stated as Canberra, with delivery taking a maximum of two days. Similar posts from the email address were found from February 2012, so there were six clear months of evidence for trading available.

This list of drugs is extremely broad, representing stimulants, heroin, synethetics, hallucinogenics, analgesics, steroids etc

### 3.2 False Positives

These appeared to arise from word-sense disambiguation errors. For example, searches for "buy ecstasy" provided links to the Rolls Royce Ecstasy car site. Better techniques for word sense disambiguation are required [13].

### 3.3 Trading Advice

Experiences in buying/selling/using illicit drugs appear to be frequently exchanged through social media forums. Examples were found at grasscity.com ("the best counter-culture community"), and several Yahoo groups and 4chan threads. Usernames and photos identifying the individuals involved could potentially be catalogued and cross-referenced within a knowledge base.

### 3.4 "Legal" Alternatives

Advertising of "legal" alternatives to ecstasy was quite prevalent on social meda sites, eg, by clicking a page you would be redirected to another page where you could order "safe and legal" bottles (http://www.buyecstasypills.com/), which are advertised as being shipped to Australia. The "Rave" pills advertised contain Amino Acids, Caffeine, Kava, Citrus aurantium, Pyridoxine HCL, and Riboflavin.

While some of these substances are legal, others may be illegal to possess and/or import under Australian law. For example, since 2007, it has been illegal to import kava without a license unless in accompanied luggage for ceremonial purposes by Pacific islanders. The company involved is unlikely to have TGA approval. Following a progression model of drug addiction, it may be possible to index content on the "follower's" pages to find evidence of use and distribution of illicit drugs. One "herbal" site, for example, has a facebook page with 2,029 followers. This hypothesis needs to be tested.

### 3.5 Market Segmentation

Some advertising was highly targeted to specific demographics in Australia, eg, advertising ketamine and mephedrone on a "tradie" website, with a contact form at the bottom. Other advertising was specifically targeting Australians with location stated as Darwin (email address of distributor and website supplied). The advertiser was also selling ecstasy, MDMA, methamphetamine, MDPV and Fentanyl. The advertiser had 334 postings identified by Google across Australia, New Zealand and other countries, but some further searching revealed that the advertiser was based in Belgium.

### 3.6 Definitions

Some sites provided definitions and associated street names, including methylenedioxymethylamphetamine (MDMA), "X," "beans," "pills," and "rolls" etc without any attempt to engage in sale or distribution.

### 3.7 Usage advice

Some sites provided advice on how to use drugs like ecstasy “safely”, or the medical effects, including "will it kill my baby if I take ecstasy while pregnant"? Again, no attempt was made to directly market or sell a particular product.

### 3.8 Biochemistry

Several sites provided research and/or biochemical information about drug structures and precursors, which were scientific in nature, and not attempt was made to buy or sell illicit drugs.

### 3.9 Experiment 1 Summary

In summary, there is strong evidence of quite open advertising of illicit drugs, targeted at Australians, even when searching through limited samples. Also, combining basic noun and verb phrases to create search terms produces reasonable results, but term disambiguation will remain a problem, as will the several categories of content into which a posting could be placed. Thus, the categorization task facing automation is multi-label, unlike binary classification in cognate fields, such as porn site detection [14].

One of the challenges for automation will be to reduce false positives, links to medical information sites etc and focus solely on true positives, and especially those which are targeting specific jurisdictions like Australia. Furthermore, new facts which are extracted from social media need to be integrated more broadly with other knowledge and intelligence from sources such as prisons [15].

## 4 ASMIA Experiment 2 - Social Network Advertising

In this experiment, we examined whether advertising in social media was being targeted to users who list their interests in illicit substances. The procedure was as follows:

- An account was setup with a major social networking site, with the hometown/current location set to Australia.
- A list of amphetamine precursors and other common stimulants in Oceania as noun phrases were entered into a term list (derived from Schloenhardt, [16]).

- The terms were entered sequentially into the search function, and if one or more pages were available for each term, the top page was added to the user's list of interests. This list was used to generate advertising which is considered relevant to that user's interests.
- Over a 3 month period, advertisements that were generated in response to the term list that might be associated with the sale or distribution of illicit drugs was logged.

Step 3 yielded some interesting results: as an example, searching for phentermine (Schedule 4, script only in Australia) resulted in a top hit "community page", which in turn had a user-contributed link to www.ipharmastore.net. This online store sells phentermine for US$226.86 per pack (without the need for an Australian prescription), and they even list a local Melbourne phone number on their website (0399886362). In turn, this number has been reportedly linked to PC servicing scam phone calls (http://whocallsme.com/Phone-Number.aspx/0399886362). The company appears to be selling and distributing the drugs from the Philippines, as their website indicates that they are licensed by the Department of Health, Bureau Food and Drugs (and not by any Australian authority).

While there were many individual pages associated with each of the terms, there was no advertising detected that was targeted towards the sale and distribution of illicit drugs. Perhaps this reflects the greater level of vetting applied to online advertising, as opposed to user-contributed content. For example, Facebook's advertising guidelines clearly prohibit the promotion of illicit or recreational drugs [17].

## 5 Discussion

Social networking clearly provides avenues for the sale, marketing and distribution of illicit drugs and/or for the abuse of legal drugs which have not been prescribed. There is clear evidence that drug dealers are targeting Australians in both of these categories, promoting a disturbingly broad smorgasbord of potentially harmful substances. On the positive side, we found no evidence that paid advertising was being used to promote illicit drugs, underlining the important role that policy decision and enforcement plays in security.

The methodology we have developed to gather and process raw source material for open source intelligence (ASMIA) shows great promise in being able to automate the very tedious aspects of source identification and assistance with linking new facts with existing intelligence product. Yet longstanding problems with word sense disambiguation and deeper-level semantic processing will remain. These remain active areas of research within natural language engineering. In this study, we only examined sites which were in English; a complication would be the cross-referencing and indexation of non-English language sites [18]. Trust in distributed systems remains an open problem [19] given the difficulty in scaling in complexity [20].

In policy terms, the astonishingly rapid rise of social media – especially on mobile devices – and the accompanying introduction of mobile payment systems present

challenges for law enforcement. The suggested introduction of data retention laws for access to websites (including social media) could be a valuable source of data for law enforcement agencies who are trying to understand links between posts, which may only come to an investigator's attention after some period of time. Conversely, rather than invading the privacy of law-abiding citizens by indexing all postings from every social media site, ASMIA could assist in screening out users and sites likely to be of interest to law enforcement. An obvious example from this study would be compiling an index of e-mail addresses which have been posted quite clearly in advertisement for illicit drugs; adding an e-mail address and an attribute indicating that they are involved in marketing or distribution of illicit drugs could be extremely helpful in identifying social media users likely to be of interest.

Regulatory reform in the area of the marketing, sale and distribution of legal pharmaceuticals – without a prescription – needs to be considered. According to the Therapeutics Goods Administration, such medicines can be bought into Australia under the traveller's exemption (incuding opioid analgesics) or the Personal Importation Scheme. Given that such drugs are being promoted through social media directly to Australians it is likely that this sales channel will grow over time. Thus, we suggest that further powers be given to Customs and Quarantine agencies to use ASMIA to maintain and index of known suppliers and customers within Australia (perhaps using data derived from social media) to assist in the prevention of a more organized approach to the sale and distribution of these substances within Australia.

## References


1. A. Stabek, P.A. Watters and S. Brown, "The Case for a Consistent Cyberscam Classification Framework (CCCF)", *Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing*, Brisbane, 2009, pp. 525-530.
2. G. Urbas and K.K.R Choo, "Resource materials on technology-enabled crime", in *Technical and background paper No 28*, Canberra: Australian Institute of Criminology. http://www.aic.gov.au/publications/current%20series/tbp/21-40/tbp028.aspx
3. A. Stabek, P.A. Watters and R. Layton, "The Seven Scam Types: Mappng the Terrain of Cybercrime", *Proceedings of th 2nd Cybercrime and Trustworthy Computing Workshop*, Ballarat, 2010, pp. 41-51.
4. A. Tanner, G. Warner, H. Stern and S. Olechowsi, "Koobface: The evolution of the social botnet", *Proceedings of the eCrime Researchers Summit (eCrime),* 2010, pp. 1-10.
5. E. Ormsby, "The drug's in the mail", 2012, in The Age, retrieved from http://www.theage.com.au/victoria/the-drugs-in-the-mail-20120426-1xnth.html
6. P.A., Watters, S., McCombie, R. Layton and J. Pieprzyk, "Characterising and Predicting Cyber Attacks Using the Cyber Attacker Model Profile (CAMP)", *Journal of Money Laundering Control*, 2012, 15(4)
7. J. Miron, "Violencem, guns, and drugs: A cross-country analysis", *Journal of Law and Economics*
   Vol. 44, No. S2, 2001, pp. 615-633.
8. D. Bradbury, "Data mining with Linkedin", *Computer Fraud & Security*, 2011, 10, pp. 5-8.
9. A. Blake, *Case*. Cambridge University Press, 2001.
10. Etzioni, M. Cafarella, D. Downey, A. Popescu, T. Shaked, S. Soderland, D. Weld, A. Yates, "Unsupervised named-entityextraction from the Web: An experimental study", *Artificial Intelligence*, 165(1), 2005, pp. 91–134

11. Y. Duan, J.S. Edwards and M.X. Xu, "Web-based expert systems: benefits and challenges", Information & Management, 42(6), 2005, pp. 799–811.
12. DEEWR, "Redi: A drug information resource for Australian school communities", 2011, retrieved from http://www.deewr.gov.au/schooling/programs/redi/Pages/default.aspx
13. P.A. Watters, "Discriminating English word senses using cluster analysis", *Journal of Quantitative Linguistics*, 2002, 9(1), pp. 77-86.
14. S. Ho. and P.A. Watters, "Structural and statistical approaches to filtering Internet pornography", in *Proceedings of the IEEE Conference on Systems, Man & Cybernetics*, The Hague, Netherlands, 2004.
15. J. Prichard, Y. Foon, Y. Lai, P. Kirkbride, R. Bruno, C. Ort, S. Carter, W. Hall, C. Gartner, K. Phone and J. Mueller, "Measuring drug use patterns in Queensland through wastewater analysis", *Trends and Issues in Crime and Criminal Justice* (422) , 2012, pp. 1-8.
16. A. Schloenhardt, "The market for amphetamine-type stimulants and their precursors in Oceania", *Research in Public Policy Series no 81*, Canberra, Australian Institute of Criminology, 2007.
17. Facebook. "Facebook Advertising Guidelines", retrieved from https://www.facebook.com/ad_guidelines.php
18. P.A. Watters and M. Patel, "Semantic processing performance of Internet machine translation systems", *Internet Research 9 (2)*, 2009, pp. 153-160.
19. H. Tran, P.A. Watters, M. Hitchens and V. Varadharajan, "Trust and authorization in the grid: a recommendation model", *Proceedings of International Conference on Pervasive Services*, 2005, p. 433-436.
20. P.A Watters, F. Martin and Z. Schreter, "Quadratic dose-response relationship between caffeine (1, 3, 7-trimethylxanthine) and EEG correlation dimension", *Psychopharmacology, 136*, 1997, 264-271.

## Acknowledgements

Paul A. Watters is supported by the Australian Federal Police, Westpac Banking Corporation, IBM and the State Government of Victoria. Nigel Phair is supported by a National Drug and Law Enforcement (NDLERF) grant.