

Novel psychoactive substances on the Web

1. Background

Over the last decade there has been a rise in the production and consumption of novel psychoactive substances (NPS) inappropriately termed 'legal highs.' These produce the same, or similar, effects to drugs such as cocaine and ecstasy but are not controlled under the Misuse of Drugs Act 1971. They are designed to replicate class A drugs but are structurally different enough to avoid being classified as illegal substances, so you cannot be prosecuted for possessing and using them. Medicine regulation is circumnavigated provided they are not marketed for human consumption, and so they are often sold under the guise of bath salts or plant food, though the design of their packaging and marketing may suggest otherwise.

The term 'legal high' is a misnomer in that often they have been found to contain controlled drugs, whilst creating the misconception that they are safe and approved for use. Little is understood about the effects of NPS on users as the ingredients are often unknown and there have been 68 known deaths linked with NPS use. This is uncharted territory. Whereas the health industry knew what they were dealing with previously with regards to the main types of illegal drugs (Heroin, Cocaine etc.), they, the authorities and the public are uninformed about these new emerging substances. In some cases even the vendors do not know what they are selling. There is little knowledge about NPS and how they are obtained and consumed, even before we consider the Web.

1.1. A Web Science approach to the issue

Working closely with representatives from industry and government, namely The National Crime Agency (NCA), Roke Manor and The Welsh Emerging Drugs and Identification of Novel Substances (WEDINOS) project, this research utilized interdisciplinary skills and innovative research methods in order to aid the understanding of NPS online.

1.2. Aims of the project

The aim of this project was to scope the issue and conduct exploratory research to identify themes and potential trends in the discourse on NPS found on social media and media reports online. We also set out to develop tools and use an array of mixed methods that could potentially be used to investigate online NPS networks in order to better understand the role of the Web in this issue. We were also provided with a set of preferred objectives by the NCA, these were:

1. What are the current top five NPS of choice?

2. Who are the sellers of the top five NPS of choice?
3. What mechanisms are used to promote new NPS on to the market?
4. What new NPS are being discussed - what is now/ what is coming?
5. Most common NPS under discussion

2. Activities undertaken

2.1. Analysis of online media reports

The term Media covers any site that provides 'news' content in the UK or internationally and has published written, video or image content in the context of NPS or "legal highs".

Issues surrounding NPS were afforded most coverage by the BBC news website; largely due to its broad regional coverage. The Huffington Post website also covered many NPS-related stories. Of the media also in print, The Daily Mail had the highest frequency of stories.

Media stories about NPS thematically focused on the state of the NPS market in the UK, regulatory/legislative issues, public safety warnings, and stories related to hospital admissions or the death of an individual from the use of an NPS. Stories would inevitably spike when there was a legislative change - i.e. a particular NPS was banned - or when there were hospital admissions or fatalities.

2.1.2. Objectives

To identify the main themes and narratives focused on by media sources

2.1.3. Findings

There was no clear geographical pattern to NPS coverage, and when the market was referred to, it was largely with reference to the UK as a whole, as opposed to certain regions. There were also a variety of NPS referred to in the articles. Table 1.1. shows the top 10 NPS by frequency of mentions.

Table 1.1. Top 10 NPS by frequency of mentions

Top ten mentioned NPS	Frequency of mentions
Benzo Fury	82
NBOMe	70
Mephedrone	63
AMT	59
Salvia	28
GBL	22
Meow Meow	18
Black Mamba	17
Nitrous Oxide	14
M-Cat PMA	11

The themes that were explored and coded were where the NPS was obtained, whether there was any poly -drug use involved, the number of times a story was referred to, and the overall representation of the story.

Where obtained: In general, there wasn't usually mention as to where the drugs were obtained, but when there was it tended to refer to Head Shops, or the Web, as opposed to individuals per se.

Poly Drug use: There was little reference to poly drug use, but it was stated, particularly in one case that the user had mixed cocaine and alcohol into the mix with benzo fury

Frequency: Many of the stories – both regulatory and those with victims – were repeated across the news outlets. BBC news was more regional, however, and this is reflected in the higher frequency of deaths/accidents being reported

Representation: There were a number of identified sentiments expressed in the articles. Often, it was to do with the legal status of such products, and the tragedy related with victims of the substances. There was a distinctly different tone used in articles related to

NPS fatalities than to victims of 'illegal' drugs. The victims in the deaths/accidents stories, in these cases were presented in very sympathetic lights – as 'caring', 'hardworking', 'educated', and as victims of a lack of regulation. The suggestion is that they are not considered in the same light as 'junkies', and are victims of 'legality'.

2.1.4. Strengths and limitations

Access to some media articles is restricted via the use of paywalls which require a subscription fee to view the content. Also the manual collection of newspaper archives can be time consuming as some search facilities have limited capabilities. However, this approach does ensure that only valid articles are retrieved.

2.2. Multimodal analysis of websites and top searches of NPS to identify key themes

Multimodality is an interdisciplinary approach that understands communication and representation to be more than about language. It systematically addresses much-debated questions about changes in society, for instance in relation to new media and technologies. Multimodal approaches have provided concepts, methods and a framework for the collection and analysis of visual, aural, embodied, and spatial aspects of interaction and environments, and the relationships between these

Using the top 10 google search results for NPS terms a database was created of ~6500 screenshots of those results including their HTML source code.

2.2.1. Objectives

Potential uses of this include: a resource for multi-modal study of the marketing of NPS, having a qualitative understanding of the industry, such as trust mechanisms, cover narratives ("research"), and to understand the information users are exposed to when searching for NPS. Search engines play a role in determining how a subject is explored and understood by new users, especially in light of what we know about psychology/learning and so the ordering of the first few results could have implications on later decisions.

2.2.2. Findings

One of the key insights highlighted how online sites selling NPS resemble other online industries. Fig 1.1. shows an independent vendor complete with "SMS support" and a mobile phone number, as contact details. Whilst Fig 1.2. shows a larger business which has all the services associated with a legitimate business, such as customer Services, FAQs, shipping and billing.

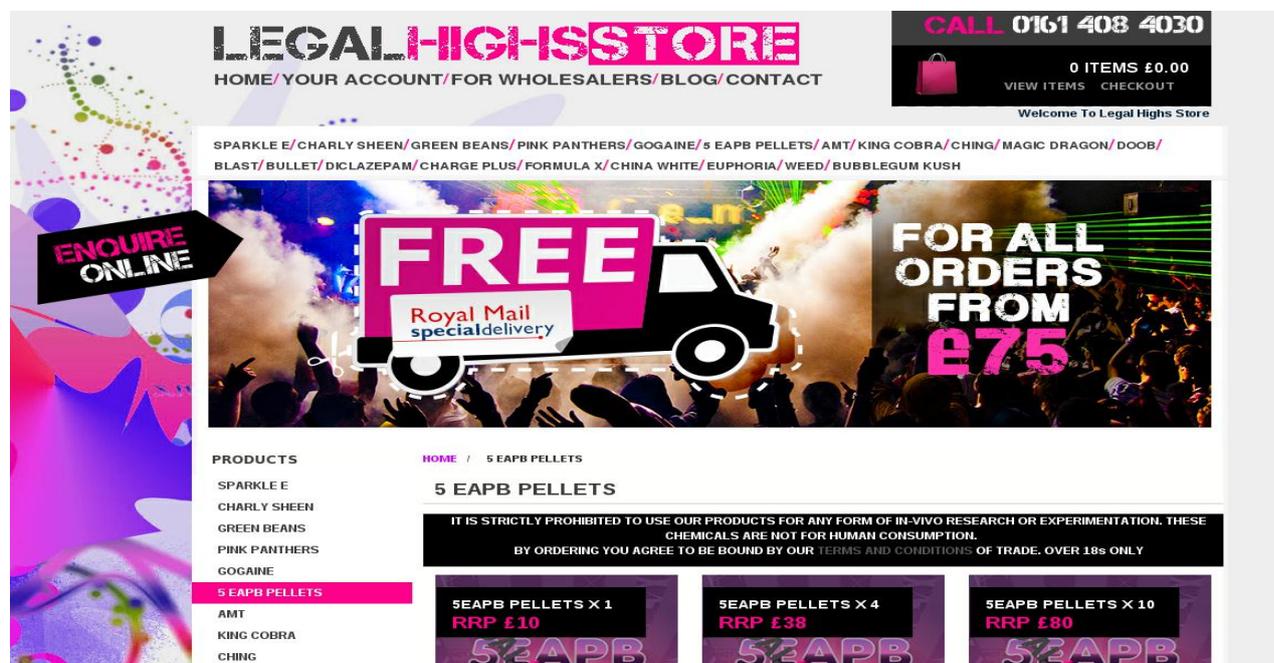
Other features of the online NPS industry include: customer Support (Telephone, Skype [for international customers], complete with service guarantees: *“We attempt to answer all queries within 3 hours,”* price match guarantees, special offers, choice of payment types (Visa, Mastercard, Amex – which potentially imply respectability, credibility and safety), bulk discounts, wholesale options (suggests a combination of direct (B2C) and “dealer” (B2B) sales models), loyalty / Reward schemes, quality control procedures, newsletters and Social Media, and wishlists.

The analysis also indicated the significance of the sales narrative. For example sites are legally obligated to emphasis to customers that products are ‘not fit for human consumption’ and often sites convey this message in narratives such as:

“All products sold are strictly not for human consumption or any form of in-vivo research. Please read our terms and conditions before ordering.”
“This product is for research and forensic purposes only.”

However, in some cases the accompanying images on websites selling NPS suggest a completely different message and appear to be actively promoting consumption. Attention is drawn to pictures of ‘party-goers’ and festivals and brightly coloured wording promoting the names of the NPS, rather than to the ‘not fit for human consumption’ disclaimer that has been conveniently put at the bottom or some other obscure place not of immediate interest, on the page. Such wording is also put in monochrome which is not easily visible, as in Fig 1.3. below.

Fig 1.3. NPS website (significance of sales narrative)



In addition such narratives are continued on forums, even though discussion of NPS in context of human consumption is not illegal. Is there the possibility that the “research” narrative, such as NPS being viewed as research chemicals, shapes the role users see themselves in? Therefore they may be creating a narrative around selves as ‘researchers’. What are the implications of that?

2.2.3. Strengths and limitations

Ethics plays a large role regarding accessing the information on public forums, and prior ethical approval is required. Collecting publicly available data without the subject’s knowledge or consent is a form of covert research. There has been much discussion about what is public and private online and it is acknowledged that users may have alternating views about their privacy expectations. It is difficult to determine the extent to which information is viewed as publicly accessible by the author. Covert research is controversial if deception occurs, for example when a researcher intentionally misinforms or does not fully disclose relevant information to subjects in cases when informed consent is required. Researchers justify gaining access to ‘deviant; research settings by deception on the basis that their work contributes to the public good. Online researchers might justify this deception by stating that it is the only way to obtain information on an important issue. The risk of harm to the research subject is usually negated by anonymising the data.

However, anonymity is also problematic because even if names and identifying information are removed from research reports, data collected from the Web is easily traceable via search engines, especially verbatim quotes. In addition, complicated ethical considerations can arise where data aggregators or search tools make information accessible to a wider public than what might have been originally intended. This is an issue related of statistical data concern rather than textual data, though publication of textual data may also allow the original source to be traced via search engines.

2.3. Investigation of NPS conversations on Twitter

This involved using modified entity recognition techniques to try and identify the names of NPSs automatically. The aim was to identify phrases to search for that will possibly contain the name of an NPS but might not. For example some might be quite direct: X is a legal high, Is there a legal high called X but they are perhaps unlikely to occur so explicitly and others might be much more tangential:

I ended up in hospital yesterday after taking X
I had an allergic reaction to X
X gave me a real buzz

The idea was to have a bank of these closed phrases that we searched for and we put each X into a hashmap when we found the phrase. Each time an instance of X matched a phrase we increased the count in the hashmap. The more phrases the instance matches, the more likely it is to be an NPS. In theory, it is relatively scale independent in that you might only need half a dozen mentions of an NPS to detect it and it is also agnostic in terms of the semantics of the name, i.e. it isn’t looking for just proper nouns or lexical units of a certain type.

We also manually created the closed phrases by taking the existing list of NPSs and using that to identify phrases in which they commonly occurred. This then generated our bank of phrases to use in identifying new and unknown NPSs.

2.3.1. Objectives

To undertake some interesting analysis of social media data from Twitter to try and identify new and emerging NPS.

If we analyse big enough datasets, we will also be able to begin to establish reliabilities of phrases in identifying NPSs and focus our bank on mostly likely phrases, or even weight the phrases based on reliability.

To create a dashboard that watches streams of social media data and tries to flag up new words or phrases when they cross a threshold that can be determined based on the known NPSs.

2.3.2. Findings

Following the first data from twitter around the search term “legal high” it quickly became apparent that users of NPS were not using “legal high” when they were discussing taking them. This led to an investigation of forums where users were discussing their experiences and recommendations for research chemicals they were taken. A difference was noted in the terms used by sellers and users and a list of slang terms and brand names was compiled. These could be broadly categorised into types – such as research chemicals, botanicals, salvia. Using these slang terms and brand names some basic twitter searches were conducted to find out how users were talking around drug taking and compiled some sentences around those searches.

Many terms are used elsewhere innocently, eg balloons. “Taking some balloons to the party” can mean Nitrous Oxide or just balloons. Same as baths salts “got a tonne of new baths salts for tonight!!!!” – are they getting high or having a lovely soak in the tub. This was found with all the slang terms. There was a crossover with drug taking language and everyday (legal) items. Eg “got jacked up on” – can be followed by baths salts or red bull. In addition a new term - ‘gravel’ was discovered, this relates to new drug in the USA a mix of three others, very new and not being talked about over here. But searches for gravel bring up people doing DIY or road works which is not entirely useful.

2.3.3. Strengths and limitations

There are considerations regarding discourse as slang terms need keeping up with. What a slang term refers to can vary temporally and geographically. Other potential avenues for slang terms include The Urban dictionary -This lacks an API but if it had one then terms that were labelled as relating to drugs would be worth understanding, and Wikipedia - Wikisauri are often talked about but never generated. Wikipedia contains numerous slang terms

linking through to their technical names. These slang terms tend to only be the most common and regional details tend to be discouraged.

Further research is needed to identify slang terms for the UK and its various regions. It would appear serious users (or researchers) discuss drug use on forums. Further research into use and drugs may be better off centred around these forums where users discuss new types of drugs and how “researchers” experience them. Co-occurrence of these slang terms with other related terms may be a useful approach to help narrow down context and provide more useful information from this type of social media analytics.

2.4. Web Observatory approaches

The Web Observatory infrastructure at UoS currently stores around one billion tweets taken from a *double garden-hose* Twitter stream. This represents approximately 20% of the total tweets generated from Twitter in any given time period, in contrast a standard Streaming API call provides around 1% of feeds from a given time period. The WebObs tweet data is stored without the full complement of fields available from Twitter in a Hadoop data store. Access to this store is through an endpoint and *Hue* - a front-end for DB admin.

Little is known about online engagement with NPS websites and substances on social media. The NCA currently do not observe social engagement relating to NPS on the Web or use the Web to augment existing data with contextual information openly available online. The NCA are interested in ‘lean’ browser-based tools that can be developed rapidly and would be highly adaptable to different data sources. The visualisation and interaction with this data online would support and facilitate decision making and insights.

2.4.1. Objectives

To triangulate the two datasets supplied by the NCA (website URLs and substances) with social media data (Twitter) and other third party data from the Web.

1. Generate a ‘watchlist’ of terms from origin data (NCA supplied)
2. Observe Twitter’s streaming content for terms in a ‘watch list’.
 - a. Information diffusion observations
 - b. Relationship network growth
3. Get contextual data about websites from NCA list

2.4.2. Findings

Three modular tools were developed during the week:

1. Twitter stream extractor

The application is constructed with Node.js, using MongoDB as a data store. An array of terms are inserted into the app which extracts relevant Tweets from the Twitter Streaming API.

2. Website WHOIS

Website URLs are looked up against WHOIS and other registration data about the operating company, location, hosting and people associated with a website.

3. Visual Analytics for NPS / Social Media data

A collection of data visualisations that facilitate insights using multidimensional filtering to explore the heterogeneous data sources.

2.4.3. Strengths and limitations

The use of Twitter for this type of investigation work shows potential, but only as part of a more complete set of discrete data sources that serve to contextualise the limited data that the NCA currently have on how NPS are sold, discussed and promoted on the open Web.

Using the Website Whois data it was possible to see the several NPS sites identified by the NCA were owned and operated by the same company. This relates directly to the NCA's second objective - who are the sellers of the top five NPS of choice?

There are issues with Whois information in that it can cost money to be able to query regularly. Whois information is also not formatted in a consistent way to make parsing it straightforward. There is additional Information available in DNS and redirects data however, DNS will indicate web server IP addresses and mail server information that may allow correlation between sites.

This was achieved using Python and will need to be converted to a Node.js app to fit with the rest of the application suite.

<http://www.dnspython.org/> Hosting websites on services such as the Amazon cloud makes it hard to distinguish individual operations without further work. Many sites use Gmail for handling their mail also which means that the source domain / mailservers are hidden.

Volume of data on legal highs too large – perhaps split by timezone to determine UK tweets, split by sentiment to distinguish advertisements from media responses – advertisements will have positive sentiment whereas the media will talk negatively <https://github.com/loadfive/Knwl.js> <https://github.com/thinkroth/Sentimental>

Analysis of the "legal high" tweets showed that whilst there are no conversations (thus no diffusion of messages) that occur, what does diffuse is the use of URL's and other Web related content (e.g. Images, videos, or specific blog articles). The diffusion of links/urls/media was constructed using a software application which analyses the content of the corpus of tweets and provides the diffusion (or chain) of urls within the legal high tweets. The results provide another perspective on the network of information exchange on Twitter, revealing that a large proportion of the highly diffusion content within the Legal Highs tweets contained URLs.

By applying network analysis techniques to resulting network chains, the analysis shows that there exists a number of smaller, disconnected networks which are assembled around the sharing of URLs relating to websites which are selling/discussing the legal highs. As a visual aid, a video has been constructed (link:

<https://www.youtube.com/watch?v=G2F12wVI288> demonstrating the temporal growth of these diffusion networks, which shows the temporal analysis/growth of the data, plus highlighting the different referral chains that occur. From the video, what becomes clear is the 'viral' nature of content within certain tweets, which can be measured by both their speed and volume of retweets. Future aims to use previous datasets such as those already collected in order to identify and flag potential content which may be deemed as viral within the legal high context. This method may prove to work as an early indicator for content that needs to be monitored or inspected in more detail using more manual processes.

3. Conclusions

What we have identified is that as an early predictor of NPS trends the Web can be a very useful tool, given the right analytical methods, datasets, and technical platforms. The research week has increased social capital and enabled the creation of a network of individuals with a variety of skills to tackle this social problem. In a very short space of time we have identified key areas for further research and designed appropriate methods to investigate them.

The expectation is that with the continuation of this project we can aid the understanding of the online NPS market and develop suitable technological and methodological approaches to observe the Web. A long-term goal that can be expected from this project is a set of computer-supported observational tools.

With the support of the NCA and WEDINOS we are able to research what is essentially an unexplored area that falls between policing and health care, namely that of the open online drug market which is distinct from the illegal online drug market. Vendors and consumers are discussing NPS in easily observable channels. Within this research we can then identify online trends such as demographics, high areas of activity, and pathways into NPS activity.

Project members

Wendy Hall
Lisa Sugiura
Paul Booth
Richard Gomer
Laura Hyrjek
Becki Nash
Kieran Rones
Ramine Tinati
Mark Weal
Craig Webber
Abby Whitmarsh