JMIR Publications
The Leading eHealth Publisher

Drug Abuse Ontology to Harness Web-Based Data for Substance Use Epidemiology Research: Ontology Development Study

Usha Lokala, Francois Lamy, Raminta Daniulaityte, Manas Gaur, Amelie Gyrard, Krishnaprasad Thirunarayan, Ugur Kursuncu, Amit Sheth

Submitted to: JMIR Public Health and Surveillance
on: October 10, 2020

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript **5**
Supplementary Files **35**
Multimedia Appendixes **36**
Multimedia Appendix 0 ..... **36**

Preprint
JMIR Publications

Drug Abuse Ontology to Harness Web-Based Data for Substance Use Epidemiology Research: Ontology Development Study

Usha Lokala[1] MSci; Francois Lamy[2] PhD; Raminta Daniulaityte[3] PhD; Manas Gaur[1] MSc; Amelie Gyrard[4] PhD; Krishnaprasad Thirunarayan[5] PhD; Ugur Kursuncu[1] PhD; Amit Sheth[1] PhD

[1]AI Institute, University of South Carolina Columbia US
[2]Mahidol University, Faculty of Social Sciences and Humanities Salaya, Nakhon Pathom TH
[3]College of Health Solutions, Arizona State Univeristy Tempe US
[4]Trialog Information Technology & Services Ile-de-France FR
[5]Wright State University Dayton US

Corresponding Author:
Usha Lokala MSci
AI Institute, University of South Carolina
Columbia
US

Abstract

Background: Web-based resources and social media platforms play an increasingly important role in health-related knowledge and experience sharing. There is a growing interest in the utilization of these novel data sources for epidemiological surveillance of substance use behaviors and trends.

Objective: The key aims are to describe the development and application of the Drug Abuse Ontology (DAO) as a framework for analyzing web-based and social media data to inform public health and substance use research in the following areas: 1) determining user knowledge, attitudes, and behaviors related to non-medical use of buprenorphine and illicitly manufactured opioids through the analysis of web forum data (PREDOSE); 2) analyzing patterns and trends of cannabis product use in the context of evolving cannabis legalization policies in the U.S through analysis of Twitter and web forum data (eDrugTrends), and 3) assessing trends in the availability of novel synthetic opioids through the analysis of cryptomarket data (eDarkTrends); 4) COVID-19 trends in social media data related to 13 states in USA as per Mental Health America (MHA) reports.

Methods: The domain and scope of the DAO were defined using competency questions from popular ontology methodology (101 ontology development). The 101 method includes 1. determining the domain and scope of ontology; 2. Reusing existing knowledge; 3. Enumeration of important terms in ontology; 4. Defining the classes and its properties and creating instances of the class. The quality of the ontology was evaluated with a set of tools and best practices recognized by the Semantic Web community and the AI community that engage in natural language processing.

Results: The current version of DAO comprises 315 classes, 31 relationships, and 814 instances among the classes. The ontology is flexible and can easily accommodate new concepts. The integration of the ontology with machine learning algorithms dramatically decreased the false alarm rate by adding external knowledge to the machine learning process. The ontology is recurrently updated to capture evolving concepts in different contexts and applied to analyze data related to social media and dark web marketplaces.

Conclusions: DAO provides a powerful framework and a useful resource that can be expanded and adapted to a wide range of substance use and mental health domains to help advance big data analytics of web-based data for substance use epidemiology research.

(JMIR Preprints 10/10/2020:24938)
DOI: https://doi.org/10.2196/preprints.24938

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Preprint JMIR Publications

Original Manuscript

Preprint
JMIR Publications

Drug Abuse Ontology to Harness Web-Based Data for Substance Use Epidemiology Research: Ontology Development Study

Usha Lokala[1], Francois Lamy[2], Raminta Daniulaityte[3], Manas Gaur[1], Amelie Gyrard[4], Ugur Kursuncu[5], Krishnaprasad Thirunarayan[6], Amit Sheth[1]
{nlokala,mgaur}@email.sc.edu, francois.lam@mahidol.ac.th, raminta.daniulaityte@asu.edu, amelie.gyrard@trialog.com, ugur@gsu.edu, t.k.prasad@wright.edu, amit@sc.edu
[1] Artificial Intelligence Institute, University of South Carolina, SC, USA
[2]Faculty of Social Sciences and Humanities, Mahidol University, Thailand
[3] College of Health Solutions, Arizona State University, AZ, USA
[4]Trialog, Information Technology & Services, Paris, France
[5]Georgia State University, GA, USA
[6]Wright State University, OH, USA

Abstract:

Background:

Web-based resources and social media platforms play an increasingly important role in health-related knowledge and experience sharing. There is a growing interest in the utilization of these novel data sources for epidemiological surveillance of substance use behaviors and trends.

Objectives:

The key aims are to describe the development and application of the Drug Abuse Ontology (DAO) as a framework for analyzing web-based and social media data to inform public health and substance use research in the following areas: 1) determining user knowledge, attitudes, and behaviors related to non-medical use of buprenorphine and illicitly manufactured opioids through the analysis of web forum data (PREDOSE); 2) analyzing patterns and trends of cannabis product use in the context of evolving cannabis legalization policies in the U.S through analysis of Twitter and web forum data (eDrugTrends); 3) assessing trends in the availability of novel synthetic opioids through the analysis of cryptomarket data (eDarkTrends); and 4) COVID-19 trends in social media data related to 13 states in the USA as per Mental Health America (MHA) reports.

Methods:

The domain and scope of the DAO were defined using competency questions from popular ontology methodology (101 ontology development). The 101 method includes 1. determining the domain and scope of ontology; 2. Reusing existing knowledge; 3. Enumeration of important terms in ontology; 4. Defining the classes, their properties, and creating instances of the class. The quality of the ontology was evaluated with a set of tools and best practices recognized by the Semantic Web community and the AI community that engage in natural language processing.

Results:

The current version of DAO comprises 315 classes, 31 relationships, and 814 instances among the classes. The ontology is flexible and can easily accommodate new concepts. The integration of the ontology with machine learning (ML) algorithms dramatically decreased the false alarm rate by adding external knowledge to the ML process. The ontology is recurrently updated to capture

evolving concepts in different contexts and applied to analyze data related to social media and dark web marketplaces.

Conclusion:

DAO provides a powerful framework and a useful resource that can be expanded and adapted to a wide range of substance use and mental health domains to help advance big data analytics of web-based data for substance use epidemiology research.

Keywords:

Ontology, knowledge graph, semantic web, illicit drugs, cryptomarket, social media

Introduction

Background

Illicit drug use is a complex social phenomenon generating a variety of public health issues that affect individuals and their communities. In its 2020 report, the United Nations Office on Drugs and Crime (UNODC) estimates that 5.4% of the world population used illicit drugs in 2018 while 0.7% of the whole population suffers from substance use disorder [1]. Individuals suffering from substance use disorder are at risk of experiencing a variety of adverse psychiatric and physical health effects such as unintentional overdoses or disease infections (e.g., HIV, Hepatitis C). Individual drug use also potentially impacts the well-being of others, affecting local communities and neighborhoods (Boardman et al., 2001), which, in turn, creates the contextual conditions and social determinants linked to individual drug use initiation [2]. Although cannabis remains by far the most consumed illicit drug use with more potent forms potentially linked to adverse consequences [3], opioid and amphetamine-type drugs remain more frequently associated with psychiatric and physical harms (NIH, 2020).

Although illicit substance use represents an endemic phenomenon affecting modern societies, recent years have seen radical and rapid changes in terms of the variety of substances available, the growing role played by the Internet, and the decriminalization or legalization of several illicit substances in an increasing number of countries. For example, the EMCDDA has identified and listed approximately 400 Novel Psychoactive Substances (NPS) since 2015 [4], while cryptomarkets located on the Darknet have become increasingly important platforms for the distribution of NPS as well as other illicit or prescription drugs [5,6]. These changes call for more timely methods of data collection allowing the monitoring of both demand and supply sides. In this ever-changing environment, user-generated content on illicit drug use shared on social media represents a rich source of unsolicited and unfiltered self-disclosures of attitudes and practices related to substance use [7]. Furthermore, web-based sources of distribution can be harnessed to provide updates on the illicit drug supply trade and new trends [8].

These unfiltered web-based communications and advertisements offer a rich source of data sensitive to changing and emerging drug use trends and can be used to complement and enhance existing epidemiological surveillance systems.

Semantic web-based approaches play a key role in enhancing and improving big data analytics for such complex domains as substance use. The semantic Web is an extension of the Web in which a set of design principles and technologies have been created to capture the meaning of information [9]. An ontology is defined as a specification of shared concepts and relationships among them, consisting of a schema and a knowledge base of instances [10].

Ontologies also play key aspects in the development of 1) semantic web applications, 2) semantic annotation of data, and 3) tools for querying and reasoning [11]. However, to apply semantic web

tools effectively, there is a need for a domain-specific ontology to represent the main entities of value described in the social media posts and their relationships [12].

There has been a broad range of research developing ontologies for social media data. For instance, the work proposed by Kim et al. [13] aimed to develop an ontology dedicated to obesity for investigating obesity-related social media posts and detecting sentiments, emotions, and opinions posted on specific social media. Their ontology was evaluated by mapping concepts from ontology with similar terms found in tweets related to obesity and is only limited to 8 superclasses related to broader perspectives of any Biomedical ontology. This study is limited to social media posts for improving upon the ontology and the keywords are vastly distributed among the top two obesity types (Abdomen, Thigh) and top three management types (Diet, Exercise, Drug Therapy) and are only limited to the general population in social media.

There are fewer ontologies related to the domain of mental health. For example, Jung et al. [14] proposed to design an ontology using an entity-attribute-value (EAV) triplet data model dedicated to adolescent depression in order to analyze related social media. This ontology was developed using clinical guidelines and unstructured social media posts with 777 terms divided into 'risk factors', 'signs and symptoms', 'screening', 'diagnosis', 'treatment', and 'prevention. This work is mainly limited to the extraction of data solely from adolescent depression-related social media posts.

Several prior ontologies were developed for the analysis of the prescription drug domain. For example, Prescription of Drugs Ontology (PDRO) [15] aims at improving the semantics of drug prescriptions and prospectively enabling the interoperability of prescription data by reusing classes and object properties from Information Artifact Ontology (IAO) [16], the Ontology for Biomedical Investigations (OBI) [17], the Ontology for General Medical Science (OGMS) [18], Ontology for Medically Related Social Entities (OMRSE) [19] and the Drug Ontology (DRON) [20] ontologies. However, these ontologies focus on medical uses of prescribed drugs and do not include concepts or slang terms related to the use of illicit drugs and addiction.

As the opioid crisis has deepened in recent years, efforts to analyze the Opioid research on social media and make policy decisions have intensified. In a recent study, a specific knowledge graph, called ODKG [21] was developed to capture opioid-related drugs and related entities in electronic health records. As the Drug Abuse Ontology (DAO) also contains information about opioid-related drugs, we compared ODKG and DAO in terms of their coverage of relevant entities in opioid-related social media corpus (Twitter) observed that the DAO outperformed ODKG by order of magnitude. Since the DAO was designed to also cover slang terms that are common in social media, the DAO performed well by retrieving 7 Million more tweets compared to ODKG (2 Million) from a resource of 1.2 Billion crawled tweets during COVID19 [22].

The key aims of this paper are to describe the process of development, evaluation, and application of the DAO to facilitate and enhance social media and web-based analytics for substance use epidemiology research. The paper describes the process of DAO development in the context of four research projects, three NIH sponsored studies that aimed to harness web-based and social media data for substance use epidemiology research: 1) PREDOSE project that aimed to characterize user knowledge, attitudes, and behaviors related to non-medical use of buprenorphine and other illicitly manufactured opioids through the analysis of web forum data [23–25]; 2) eDrugTrends project that focused on patterns and trends of cannabis product use in the context of evolving cannabis legalization policies in the U.S through the analysis of Twitter and web forum data [26–30]; 3) eDarkTrends project that aimed to identify availability trends of novel synthetic opioids through the analysis of crypto market data [31–33]; 4) COVID-19 trends in social media data related to 13 states in the USA and its mental health impact.

The terminology related to Machine Learning (ML), Natural Language Processing (NLP), and Ontology Design used in this manuscript is organized alphabetically in Table 1.

Table 1. Descriptions of ML, NLP, and Ontology terms used in this manuscript

Terminology	Description
101 ontology [34]	The 101 ontology is a guideline to create an ontology and offers step by step process. It leverages the authors' experiences developing and maintaining ontologies in several ontology environments like Protege.
BRF-CF	This is a Bootstrap and Bagged Random Forest with Contextual Features. Random Forest is one of the most popular ML algorithms. It is a type of ensemble ML algorithm called bootstrap or bagging.
Class | Data Property | Individual count	These terms are used as the signatures for the imports closure of the active ontology. In other words, the number of distinct classes, object properties, data properties, and individuals are mentioned in the ontology. The numbers here include built-in entities, such as owl: Thing if they are explicitly mentioned in the ontology.
Community Ontology Repository (COR) [35]	This is the repository of ontologies hosted by ESIP members that would let users try out semantic technologies, understand their benefits, and explore possible applications that used semantic resources.
Depression and drug abuse BERT	BERT is a Bidirectional Encoder Representations from Transformers and is a transformer-based ML technique for NLP. We fine-tune BERT models on corpora that are representative of depression and drug abuse.
DBPedia [36]	DBpedia is a crowd-sourced community effort to extract structured content from the information created in various Wikipedia projects.
DSM-5	DSM-5 (Diagnostic and Statistical Manual for Mental Disorders) is the taxonomic and diagnostic manual developed and published by the American Psychiatric Association. It is an authoritative guide for mental healthcare professionals in the diagnosis of mental disorders.
Entity | Concept	The entity is referred to as an encompassing concept for classes, individuals, and properties. Concept and class are simply synonyms.

F1 Score	F1 Score is the weighted average of Precision and Recall. This score takes both false positives and false negatives into account. F1 is usually more useful than accuracy score.
False Positive | True Positive	A false alarm is also known as a false positive. A false positive is a result that indicates a given condition exists when it does not. For example, the model indicates that cannabis can cause pain when it does not cause pain. A true positive is an outcome where the model correctly predicts the positive class. Similarly, a true negative is an outcome where the model correctly predicts the negative class. A false positive is an outcome where the model incorrectly predicts the positive class.
HLF | VLF | FGF	While training a ML model, we organized our feature set into three broad groups: Horizontal Linguistic Features (HLFs), Vertical Linguistic Features (VLFs), and Fine-Grained Features (FGFs). Contextual Features (or embedding of a social media post) with Modulations (CFwM) and without Modulations (CFw/oM) are two additional feature set created using Word2Vec.
Ontology metrics [37]	The metrics list the numbers for structures and representation of ontology in Protege as it is the most widely used tool to create an ontology. Axioms associate class and properties and are a combination of logical and non-logical attributes. The number of distinct classes, object properties, data properties, and individuals reported is focused on the evaluation of the structure of DAO.
Oops (Ontology Pitfall Scanner) | Vapour |Triple Checker [38]	These are Semantic Web (SemWeb) validation or documentation tools that help to improve ontologies. Oops detect common pitfalls in ontology automatically and provide recommendations to fix them.
Owl file	The W3C Web Ontology Language (OWL) is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things.
PerfectO methodology [38]	PerfectO references, classifies, and provides tools to encourage Semantic Web Best Practices to achieve Semantic Interoperability by focusing on ontology improvement.
Precision | Recall	Precision is the proportion of times that when you predict it is positive and it actually turns out to be

	positive whereas Recall is like accuracy over just the positives – it is the proportion of times you labeled positive correctly over the number of times it was actually positive.
Protege	Protege is a free, open-source ontology editor and framework for building intelligent systems.
SEDO [39]	SEDO stands for Semantic Encoding & Decoding Optimization. It is a procedure to modulate the word embedding (vectors) of a word. SEDO modulates the embeddings of each word in the Reddit content of the user based on the proximity of the word to the DSM-5 category.
vanilla BERT	Vanilla BERT is a variation of the attention-based BERT model and provides a pre-trained starting point layer for neural networks.
WebVOWL [40]	WebVOWL is a web application for the interactive visualization of ontologies which is one of ontology visual representations.

Evolution of DAO

As social media and other web resources play an increasingly important role in health-related knowledge and experience sharing [41], there is a need for an ontology explicitly dedicated to the domain of substance use research. DAO was developed to formalize concepts, entities, and relationships relevant to the domains of addictions and mental health to harness its use on social media data. Our approach, built on the integration of Semantic Web technologies, enhances traditional ML and NLP techniques for automatic extraction and representation of relevant data and facilitates analysis and interpretation related to the specific goals of each study.

PREDOSE

This focuses on web forum data related to the non-medical use of buprenorphine [24,25] approved in late 2002 by the U.S. Food and Drug Administration for the treatment of opioid addiction. Use of buprenorphine was defined as non-prescribed when used without medical supervision. Although there is always a level of uncertainty in disambiguating prescribed vs. non-prescribed use in web-based discussions, some of the ques and practices shared by individuals provided indicators about non-prescribed use (e.g., saying that Suboxone was obtained from a friend; that “bupe” was snorted; or that it was cut up and used in smaller amounts; etc.). Buprenorphine (Suboxone, Subutex, etc.) is the only controlled substance that may be prescribed for the treatment of opioid addiction by a licensed physician in an office-based setting. The overall purpose of PREDOSE was to study user-generated Web forum discussions about the illicit use of Suboxone (buprenorphine/naloxone), Subutex (buprenorphine), and other buprenorphine products by applying novel information processing techniques, to facilitate qualitative and quantitative analysis [24]. Along with Twitter and Reddit, we also used three web forums that provided venues for people to freely share drug use experiences and post questions, comments, and opinions about different drugs. One of these online web forums used in our research was Bluelight [42] (please note that in compliance with Institutional Review Board (IRB) guidelines at Wright State University, the names of the other two forums have not been disclosed in this manuscript). Our team has developed a research collaboration with the

Bluelight team and was able to obtain de-identified data updates directly from Bluelight. Data from those forums were collected using custom-built web crawlers. We chose to study buprenorphine because there was at that time (2011-2012), a growing body of evidence that buprenorphine was used and that there was relatively little knowledge about the patterns and trends of its non-medical use in the U.S. Because buprenorphine use is linked to a broader domain of illicit opioid use and addiction, the initial versions of the DAO included detailed representation of the opioid class drugs, including slang and brand name terminology. The DAO developed for the PREDOSE project also included other classes of drugs, such as cannabis and stimulant-type drugs, since polysubstance use is common among illicit opioid users. Figure 1 demonstrates the usage of the DAO ontology within our PREDOSE architecture, which comprises three main modules:

1. Data collection module which collects approximately 1 million posts (1,066,502) from 35,974 users.
2. Automatic coding module which semantically annotated the posts using the DAO ontology.
3. Data analysis and interpretation module to visualize the keywords (e.g., Loperamide, buprenorphine) found within posts and referenced within the DAO ontology.

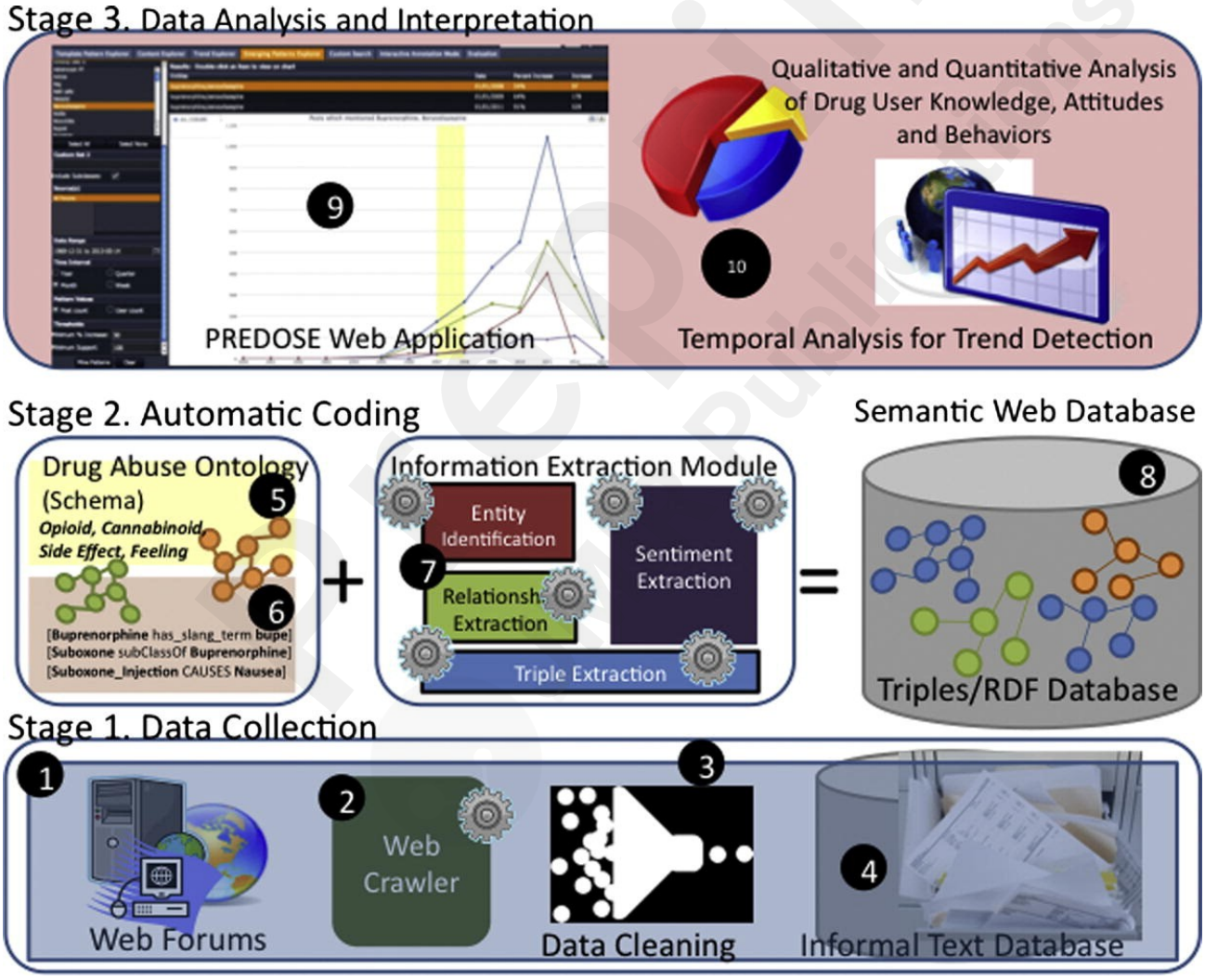


Figure 1. Usage of the DAO ontology within PREDOSE [24]

eDrugTrends

This is our second project that received funding from NIH/NIDA in 2014 [43]. The study focused on social media data related to cannabis and synthetic cannabinoid use in the context of evolving

cannabis legalization policies in the U.S. The aim of this study was to develop eDrugTrends, a comprehensive software platform for semi-automated processing and visualization of thematic, sentiment, spatio-temporal, and social network dimensions of social media data (Twitter and Web forums) on cannabis and synthetic cannabinoid use. The study also aimed to 1) identify and compare trends in knowledge, attitudes, and behaviors related to cannabis and synthetic cannabinoid use across U.S. regions with different cannabis legalization policies using Twitter and Web forum data; and 2) analyze social network characteristics and identify key influencers in cannabis and synthetic cannabinoid-related discussions on Twitter. To address these aims of the eDrugTrends platform, the DAO was expanded further to include a more comprehensive representation of emerging cannabis products, synthetic cannabinoid products, health-related consequences, and mental health conditions.

eDarkTrends

This is the third project using the DAO. It was funded through NIH/NIDA time-sensitive mechanism [44], which started in 2017. EDarkTrends project was oriented towards novel synthetic opioids, such as illicitly manufactured fentanyl (IMF) that have emerged over the past few years and were and still are significant contributors to the increase in unintentional opioid-related overdose mortality in the U.S. [33,45,46]. However, epidemiological surveillance on cryptomarket data was limited at that time. The study's overall goal was to harness cryptomarket data to conduct surveillance of illicit fentanyl, fentanyl analog, and other novel synthetic opioid availability trends over time and identify new substances as they emerge on the Darknet environment. Ultimately, eDarkTrends aimed at providing a powerful tool for epidemiological surveillance enhancing the capacities of early warning systems to capture changes in the fentanyl and other illicit synthetic opioid supply and availability. To address the specific needs of the project, the DAO was further expanded to include a comprehensive and detailed representation of novel illicit synthetic opioid domains (e.g., carfentanil, furanyl fentanyl, U-47700, and MT-45).

COVID-19

Additionally, we applied DAO on COVID-19 social media data analysis to analyze the social media data related to the pandemic. The intent is that COVID-19 has alleviated community-wide depression, and has led to increased drug use [47]. The impact of COVID-19 on mental health was investigated in recent studies [48–50]. For this, we proposed a novel framework for assessing the spatiotemporal-thematic progression of depression, drug use, and informativeness of the underlying news content across the different states in the United States [51]. DAO is used along with the MeSH terms hierarchy in the Unified Medical Language System (UMLS), the Diagnostic and Statistical Manual for Mental Disorders (DSM-5) lexicon (Gaur et al., 2018) which are collectively referred to as Mental Health and Drug Abuse Knowledgebase (MHDA-Kb) to spot additional entities.

Methods

The ontology was manually developed by the domain expert co-authors (FRL and RD) who used a range of sources including 1) key epidemiological data sources and reports accessible through the National Institute on Drug Abuse [52], Drug Enforcement Agency [53], European Monitoring Centre for Drugs Addiction [54], and RxNorm [55]. 2) prior peer-reviewed publications related to illicitly manufactured opioids, cannabis, and other drugs [56–59]; and 3) ongoing manual assessment and examination of web-based, social media sources related to selected substances [23,25,60]. Sources of types 1 and 2 provided primary concepts while sources of type 3 were important in identifying alternative concepts including synonyms and street names. To develop the DAO, we followed the well-known 101 ontology development methodology [61]. The 101 method includes 1. determining the domain and scope of ontology; 2. reusing existing knowledge; 3. Enumeration of important terms

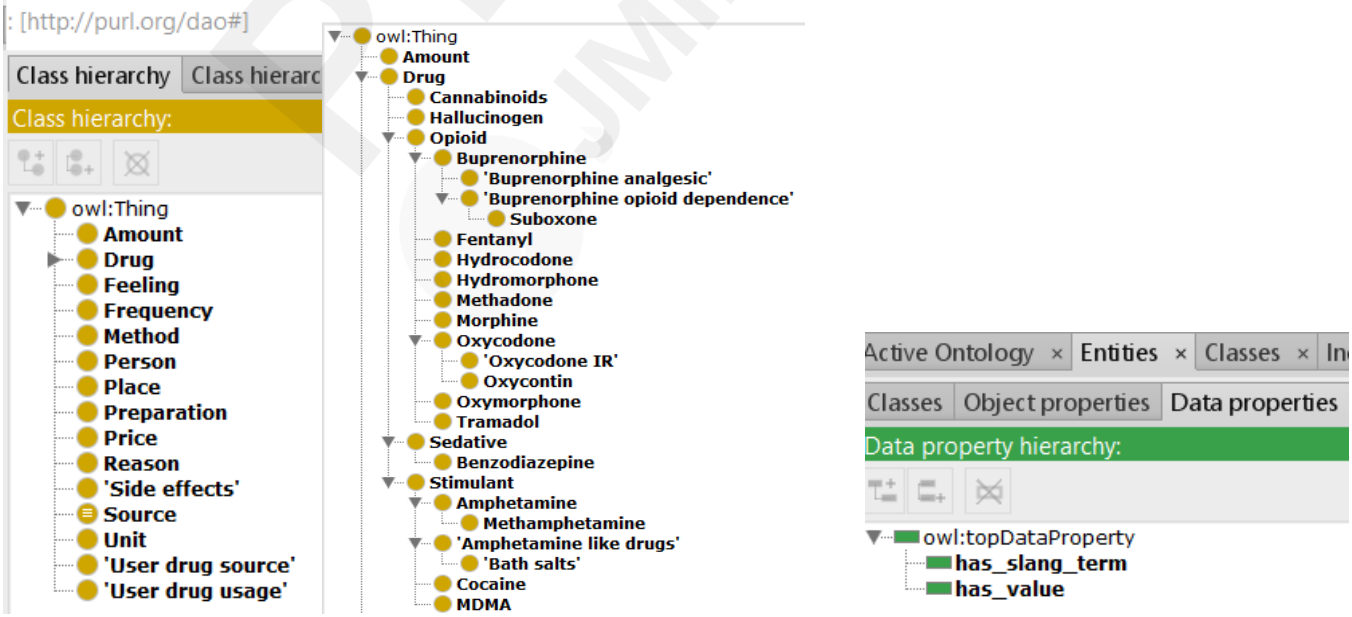
in ontology; 4. defining the classes and their properties; and 4) creating instances of the class.

Design

Figure 2 gives an overview of the DAO ontology. Protege [62], a popular ontology editor, was used to build the ontology as a tree of subclasses. The ontology was designed as a catalog of concepts related to substance use. Hence, classes of psychoactive substances (e.g., cannabinoids, opioids) were created and populated with subtypes of substances (e.g., morphine, fentanyl). Each substance was defined by its name and, when applicable, information regarding its pharmaceutical/brand name ("has_brand_name"), slang/street name ("has_street_name"), and chemical designation ("has_chemical_formula") were added. This latter information was collected through different sources: pharmaceutical/brand names were based on existing medical or pharmacological dictionaries; slang/street names were based on the domain knowledge of the second and third authors (RD and FRL); chemical designations concerned mostly Synthetic Cannabinoid Receptor Agonists (SCRA) and were based on academic literature as well as on seizure data (e.g., NFLIS, Europol). The DAO was also enhanced with concepts and slang terms related to those concepts regarding unit (e.g., "caps," "ml," "bottle"), purity, and form of preparation (e.g., "crush," "eyeball") to enable the identification and analysis of triple in text content [63]. For example, one instance of the drug 'Morphine' is 'Poppy_Tea' which has the slang terms 'Pod' and 'Poppy_Pods' used on social media.

Instantiation

This is defined as creating instances of classes in a hierarchy. The instance of a class has its own class and fills a value. The instance has its own properties. For example, 'Fentanyl' belongs to the class 'Opioid' and has its own properties like 'has_brand_name', 'has_synonym', 'has_slang_term' etc. The DAO ontology reuses instances from the DBpedia dataset [64] (e.g., Buprenorphine). Figure 3 is the WebVOWL (web application for the interactive visualization of ontologies) representation of DAO focused on the entity Cannabis derived from the visual data web [65]. Figure 2 shows the tree of drug names implemented as a web ontology format (owl) file within the DAO ontology. In Figure 2, Entities, Object Properties, Instances, and Data Properties are represented in yellow, green, and purple tags respectively which clearly depict the nature of classes, instances, hierarchies, and relationships for each entity.



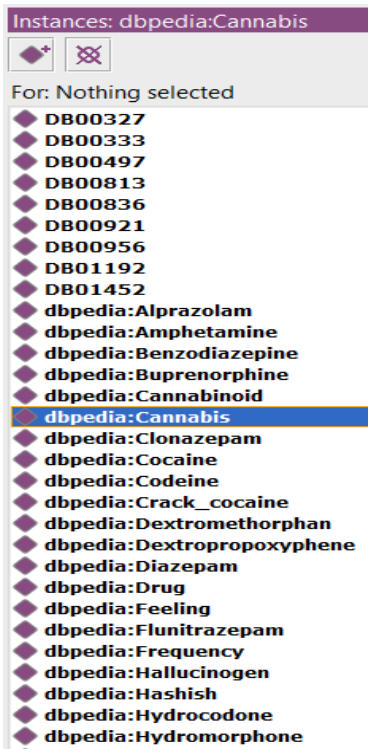


Figure 2. DAO ontology in Protégé (concepts, object properties, data properties, and instances)

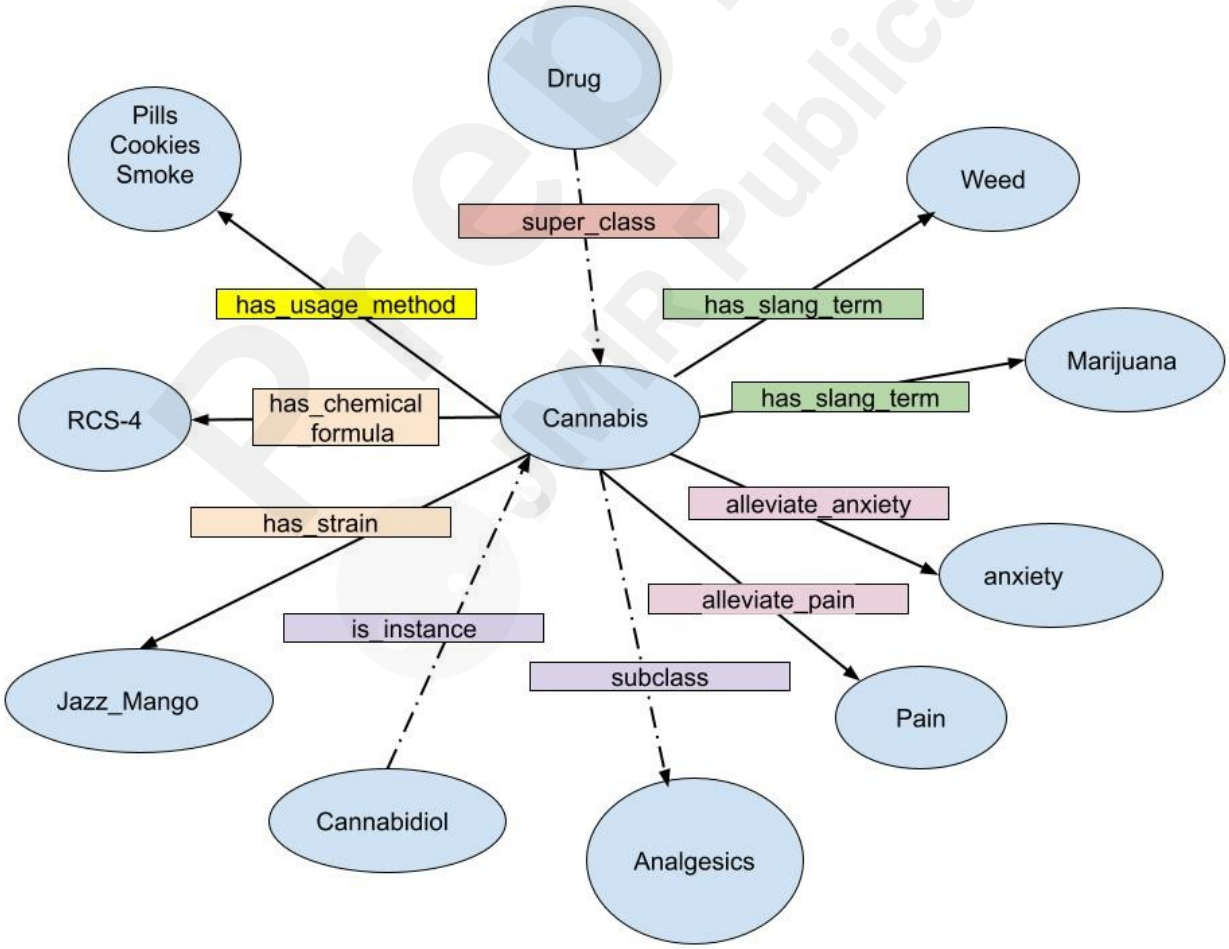


Figure 3: WebVOWL representation of the DAO ontology, focused on the Cannabis concept

Results

Evaluation

The DAO ontology has been evaluated following the semantic web best practices recognized by the International Semantic Web Conference (ISWC) Resource Track guidelines [66], which provide the following criteria: 1) Impact; 2) Reusability; 3) Design and technical quality, and 4) Availability. We have also followed the PerfectO methodology [38], which synthesizes a set of additional best practices and eases their achievements with the step-by-step tutorial [67]. We discuss the results of applying the following criteria to our DAO:

- Impact and Reusability: The DAO has been exploited in 4 scenarios as mentioned earlier. Automatic documentation can be provided with the LODE tool [68] and the DAO documentation is available here [84].

  Table 2. DAO Ontology metrics: The ontology metrics view displays entity and axiom count for the axioms in the active ontology [37]

- Design, Technical quality, and Availability: The design of the ontology is available online as a graph

Metric		Count	Description
Ontology Metrics			
	Axioms	4876	combined logical and non-logical axiom count
	Logical axiom count	3478	the number of logical axioms
	Declaration axioms count	1185	the number of declaration axioms
	Class count	316	the number of distinct classes, object properties, data properties and individuals that are mentioned in the ontology
	Object property count	12	
	Data property count	13	
	Individual count	845	
Class Axiom			
	SubClassOf	313	the number of SubClassOf axioms in the ontology. A subclass axiom states that a class is a subclass of another class
Individual Axioms			
	Data Property Assertion	2317	A data property assertion states that the individual a is connected by the data property expression (DPE) to the literal.
	ClassAssertion	830	A class assertion states that the individual is an instance of the class expression.
	AnnotationAssertion	213	An annotation assertion states that the annotation subject as an anonymous individual with the annotation property and value

visualization with WebVOWL [85]. We have improved the ontology with the Oops (Ontology Pitfall Scanner) tools which automatically detect common pitfalls and provide recommendations to fix them. Oops loaded with DAO can be tested online [86]. A Vapour tool integrated with the DAO ontology [87] has been used to check dereferencing URI and content negotiation. Finally, TripleChecker checks the use of existing ontologies if it has been correctly used within our DAO ontology [88].

- Ontology Metrics: The DAO ontology is also evaluated as shown in Table 2 with respect to several ontology metrics [69]. The metrics list the numbers for structures and representation of ontology in Protege as it is the most widely used tool to create ontology [70]. Axioms associate class and properties and are a combination of logical and non-logical axioms [71]. The number of distinct classes, object properties, data properties, and individuals reported in Table 2 is focused on the evaluation of the structure of DAO.

The below sections demonstrate the results with DAO in different platforms and how DAO evolved with each use case.

DAO within PREDOSE

Figure 4 describes how the texts are automatically annotated using DAO. In the text shown in Figure 4, we identify drug entities, dosage, time interval, route of administering the drug, etc. In DAO, Buprenorphine is defined as the subclass of ‘Subutex’ and ‘Suboxone.’ It has the slang terms ‘Bupe’ and ‘Bupey’. The term ‘Bupe’ identified in the text would not have been possible without defining it as a slang term in DAO. The DAO is capable of mapping units (e.g., mg → MILLIGRAM) and slang terms (e.g., bupe – Buprenorphine), based on a lexical lookup in the ontology. Similarly, the other concepts such as route of administration ‘injected’ are also identified in the text. In NLP-related tasks such as lexical, semantic, and syntactic analysis of textual data, adding ontology works as an external source of knowledge in identifying triples and entities in data. Conceptualizing the domain in data acts as a prior requirement for processing further information (Lexicon and Rule-based Grammar) about it [72] is shown in Figure 5. While evaluating 601 web forum posts with DAO, we achieved 84.9% precision and 72.5% recall in information extraction tasks. In particular, out of 3639 annotations, 2640 were predicted correct (true positives), while 683 slang terms are incorrect (false positives). As far as the recall is concerned, only 999 out of 3639 annotations are missed (false negatives) [24]. For Triple Extraction with DAO, we achieved 33% precision across 197 evaluated triple patterns (66 were correct, and 131 were incorrect). For Relation Extraction with DAO, we achieved 36% precision across 183 phrases (66 were correct, and 117 were incorrect). Another finding (Figure 6) is that our analysis of Web forums with DAO found that Loperamide was widely used as a treatment for withdrawal symptoms related to Opioid addiction where Buprenorphine and Methadone are commonly prescribed. Three toxicology studies following this work led to an FDA warning in 2016 [23,73]. A video demo [75] on the PREDOSE platform is available online. The PREDOSE platform indicates a need for additional enhancements in information extraction and automated data coding techniques.

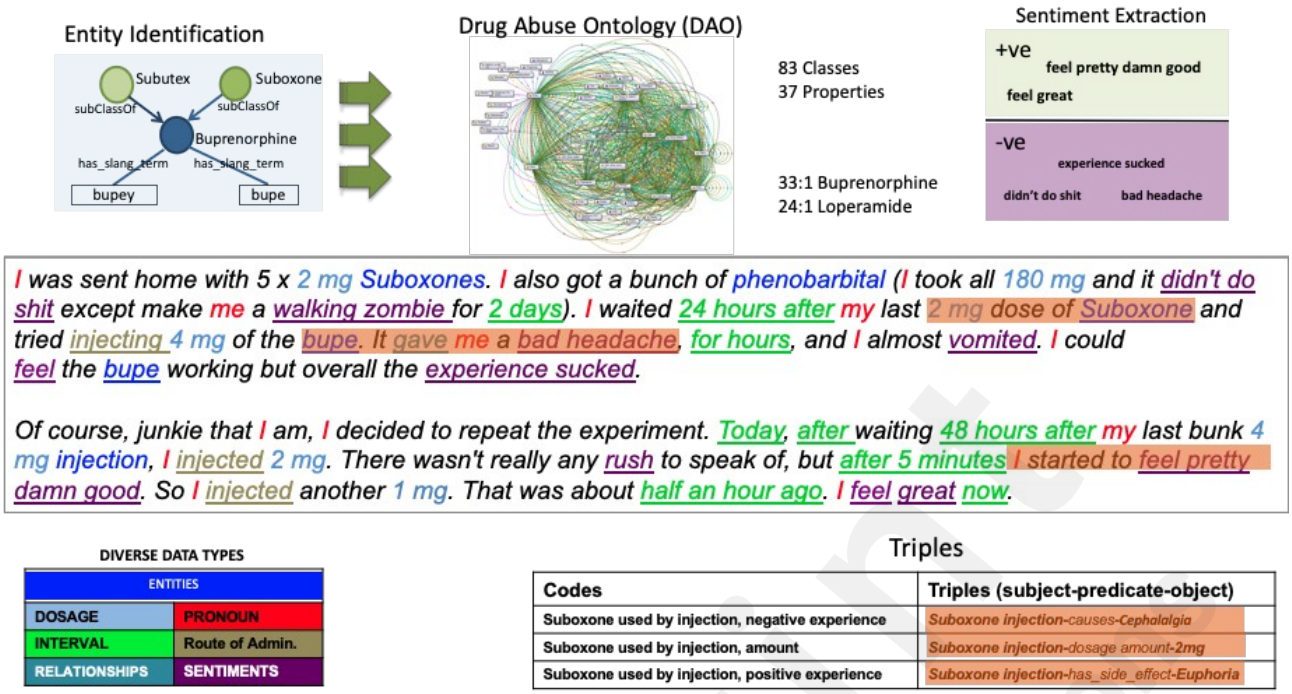


Figure 4. Automatic annotation of texts with DAO [74]

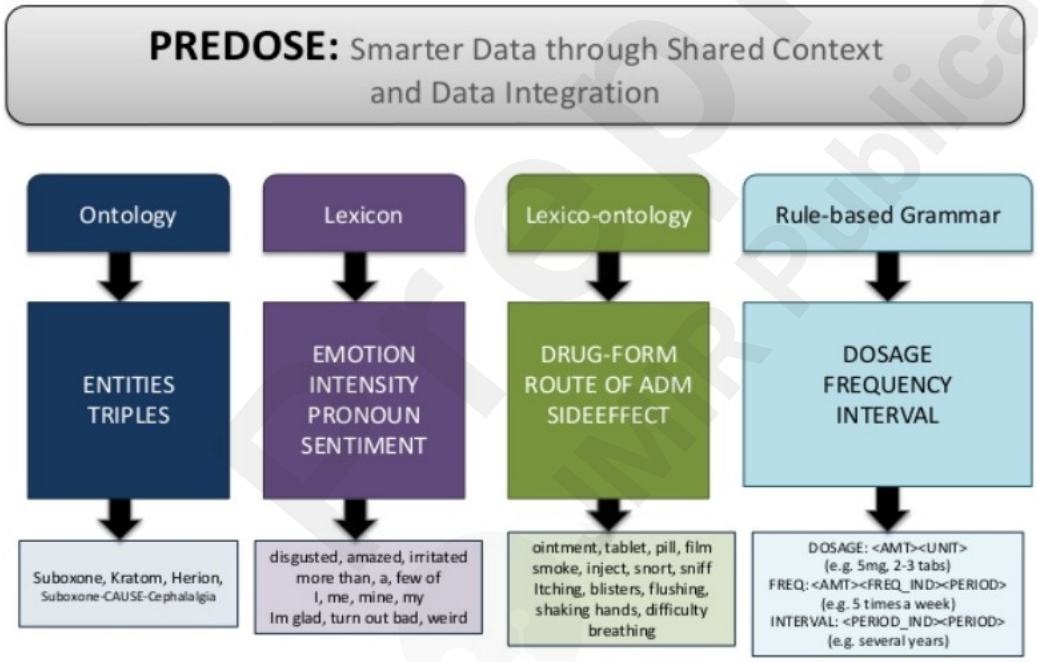


Figure 5. Benefits of ontologies with lexicons and rule-based grammar[74]

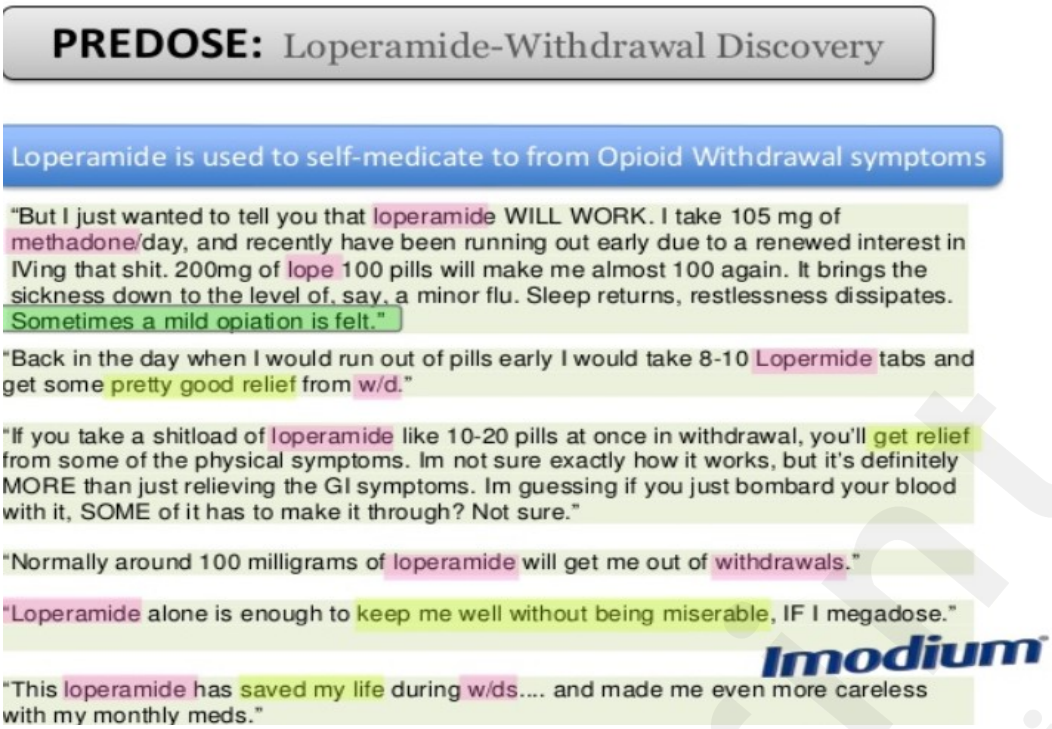


Figure 6. Loperamide Discovery and its use in self-medication for Opioid withdrawal [23]

EDrugTrends (monitoring drug trends on social media)

The eDrugTrends project aimed to analyze trends in knowledge, attitudes, and behaviors related to the use of cannabis and synthetic cannabinoids on Web forums and Twitter [24,26–29]. Figure 7 shows the application of the DAO ontology within the eDrugTrends architecture which includes four stages: 1) Data collection; 2) Data processing; 3) Data access tools for exploration and visualization; 4) Quantitative, Qualitative analyses and Interpretation. From the social science/substance use epidemiology perspective, the data processing and information extraction stages correspond with the coding task that prepares raw data for further analysis and interpretation. During data processing, the DAO came into the picture by playing an important role in identifying entities in data which are exact names/synonyms/slang terms/street names of a drug. We generated embedding vectors by using DAO for domain-specific word embedding models and we built a ML model to classify users by their types (Individual, Agency, and Retailer) on Twitter through the classification of their marijuana-related conversations [26]. We achieved this by using multimodal embeddings extracted from People, Content, and Network views achieving an 8% improvement over the empirical baseline [26]. We evaluated our approach in this work by using the Average F1-score for each user type Individual (P), Informed Agency (I), and Retailer (R). The F1 scores for individual classes of P, I, and R were 95%, 42%, and 73%, respectively. The descriptive statistics of the training set at a Twitter user account level used for this study which involved semantic filtering [76] utilizing DAO are shown in Table 3.

Table 3. Descriptive Information of user accounts on Twitter extracted using DAO [26]

Features ('#' is 'number of')	Personal Accounts	Retail Accounts	Informed Agency	Total
# Tweets	9836	1928	338	12102
# profile pictures	4394	476	111	4981

# users with description	3884	461	108	4453
# Retweets	955	24	964	1943
# Mentions	94	6	307	407

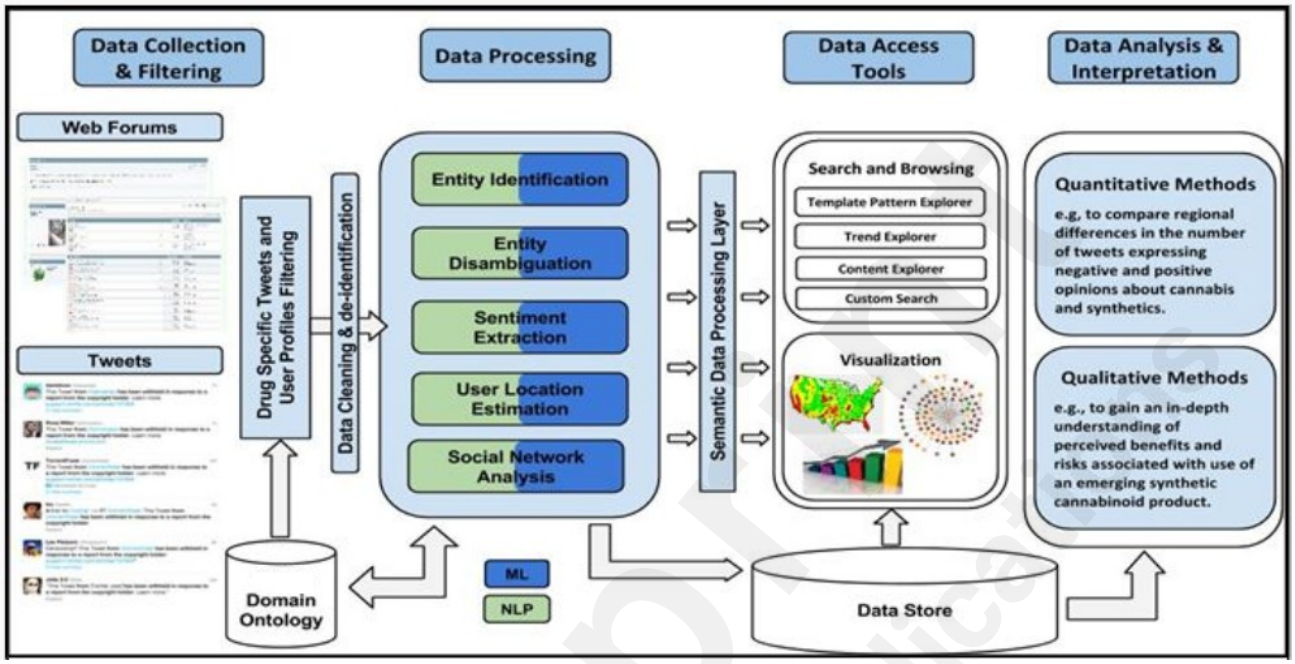


Figure 7: Architecture of the eDrugTrends project [77]

Enhancing DAO with DSM-5

The motive of enhancing DAO with DSM-5 is to provide actionable information to clinicians about the mental health of a patient in diagnostic terms for web-based intervention. We chose Reddit data for this study as the concepts, instances, and relations associated with drugs are semantically connected to mental-health communications on social media, especially on Reddit. In our Reddit corpus, the drug use-related categories form a substantial portion 48% (corpus size is 2.5M posts from 15 Mental Health subreddits by 268,104 users) of the dataset in size. However, the DAO still lacked concepts directly related to mental-health diagnostic disorders as defined in DSM-5 that are present in ICD-10 [78], SNOMED-CT [79], and DataMed [80]. In a recent study [39] on matching mental conditions of user posts on Reddit to DSM-5 diagnostic disorders, we enhanced the DAO with knowledge derived from DSM-5 which includes 20 chapters (see Table 5), consistent with ICD-10 and NIH's Research Domain Criteria [81] for mental health. Then the enhanced DAO includes representations of mental health disorders and related symptoms that were developed following the DSM-5 classification [82]. For example, references for "Cannabis Use Disorder" include terms such as "addicted to cannabis," "addicted to Marijuana," and "Jazz_mango addict." References to the feeling of "anxiety" or "anxious" include such terms as "antsy", "worried", and "agitated." These lay terms were added to the DAO manually using synonym dictionaries and by manually examining Reddit conversations related to depression, anxiety, and other mental health conditions.

The DAO curated and enhanced by DSM-5 concepts was used in a weakly supervised setting to label Reddit posts with DSM-5 categories. On a comparative analysis with the state-of-the-art research by Park and Conway [83], Saravia [84], and Gktosis [85], we observed that expansion of DAO with DSM-5 helped improve the accuracy of our entity identification tools (reduced false

positives by 92%). These results are shown in Figure 8. We further assessed the meaningfulness of the prediction through a reliability assessment with a domain expert, which gave an agreement score of 84%. Additionally, the incorporation of slang terms from DAO to match and process the informal social media data improved both the coverage and recall. These results are shown in Table 4. Thus, we demonstrated that semantic weighting of contextual features from the content using DAO and DSM-5 knowledge could significantly improve the robustness of the AI system. Since the online content is mapped to a clinically acceptable vocabulary, the system brings in explainability. Further, Table 5 shows the improvement in number of concepts extracted from DAO being captured in our Reddit Corpus that relate to DSM-5 20 chapters before and after adding slang Terms.

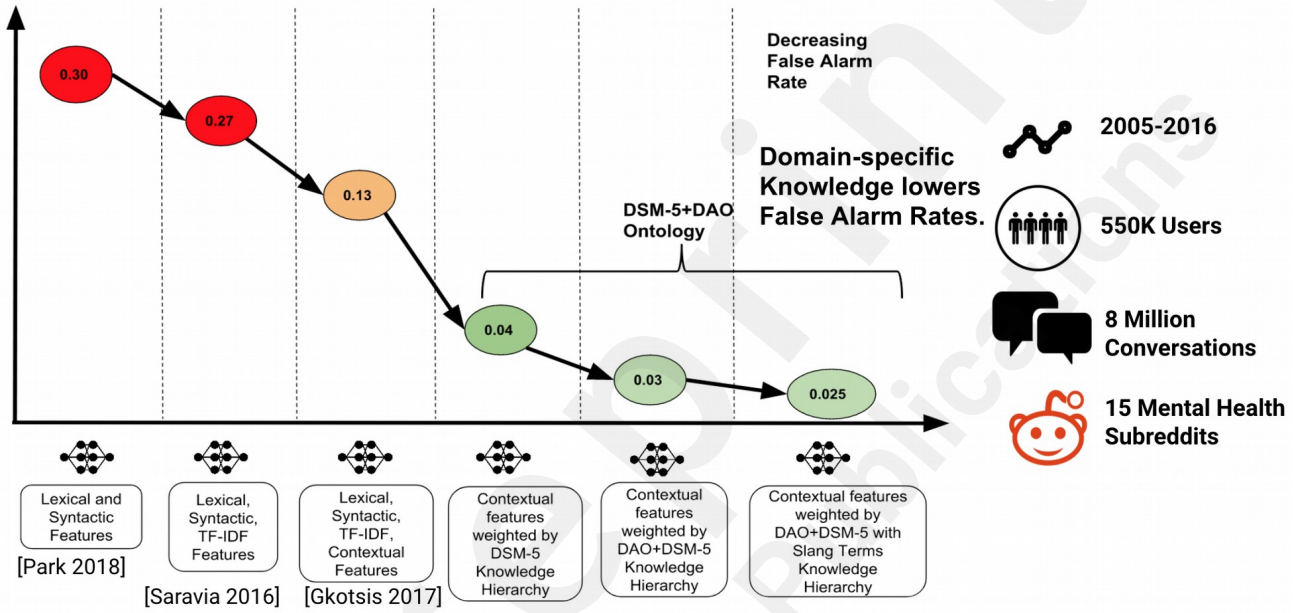


Figure 8: Results illustrating that Domain-specific Knowledge bases lower False Alarm Rates in identifying DSM-5 categories to tag posts in mental health subreddits.

Table 4. Ablation study on Contextual Features (CF) and their modulation using SEDO Weights generated from either DSM-5 or its enrichment using DAO and slang terms. Note: This table demonstrates the improvement of models with enhanced DAO.

Method (with HLF, VLF, and FGF)	Precision	Recall	F1-Score
BRF with CF	0.60	0.54	0.57
BRF-CF (SEDO Weights generated from DSM-5 Lexicon without DAO)	0.87	0.77	0.82
BRF-CF (SEDO Weights generated from DSM-5 Lexicon with DAO without Slang Terms)	0.87	0.80	0.83
BRF-CF (SEDO Weights generated from DSM-5 Lexicon without DAO with Slang Terms)	0.85	0.82	0.83
BRF-CF (SEDO Weights generated from DSM-5 Lexicon with DAO with Slang Terms)	0.88	0.83	0.85

The base model for the ablation study is Balanced Random Forest with Horizontal Linguistic Features (HLF: Number of definite articles, words per post, first-person pronouns, pronouns, subordinate conjunctions), Vertical Linguistic Features (VLF: Number of POS-tags, Similarity between the posts, Intra-Subreddit Similarity, Inter-Subreddit Similarity), and Fine-grained Features (FGF: Sentiment, Emotion, and Readability Scores).

Table 5. Demonstration of improvement in the number of DSM-5 category related concepts being captured before and after including DAO [39]

DSM-5 Category	# DSM-5 related concepts captured without DAO	# DSM-5 related concepts captured with DAO
Dissociative Disorder	20	20
Anxiety Disorder	40	87
Substance Use & Addictive Disorder	39	123
Schizophrenia Spectrum	77	77
Sleep-Wake Disorder	14	19
Paraphilic Disorders	14	14
Gender Dysphoria	15	15
Neurodevelopmental Disorders	25	53
Sexual Dysfunctions	23	23
Personality Disorders	76	98
Trauma & Stressor Related Disorder	25	28
Disruptive, Impulse, Control & Conduct Disorder	34	34
Psychotic Disorders	85	87
Bipolar & Related Disorders	75	84
Elimination Disorders	18	18

Preprint JMIR Publications

Depressive Disorders	71	107
Obsessive-Compulsive Related Disorder	43	60
Feeding & Eating Disorders	32	39
Neurocognitive Disorders	80	80
Suicidal Behavior/Ideation	34	47

EDarkTrends (monitoring drug trends on cryptomarkets)

The DAO also plays an essential role in identifying the relevant entities and analyzing data from the Darknet cryptomarkets (e.g., Agora, Dream Market, Empire Market) in order to quantify and assess the availability of fentanyl, fentanyl analogs, and other novel synthetic opioids on the cryptomarkets [23-24]. The snapshot of Darknet Marketplace is shown in Figure 9. The terms and slang terms associated with instances populating the DAO's opioid subclass, as well as the dosage (e.g., gram, ml, once) and form (e.g., tablet, powder) classes were compiled as Regular Expressions and used as expression patterns in the dedicated Named Entity Recognition (NER) algorithm specifically designed for Darknet data [33]. The DAO was inductively augmented with abbreviations and terms specific to the cryptomarket environment (e.g., fuff for fluoro-furanyl fentanyl or FE for Finalize Early) to ensure that only relevant data about novel synthetic opioids were collected. The NER allows capturing the types and quantities of novel synthetic opioids advertised on crypto markets: for example, the NER would provide the following information about the advertisement "*FENTANYL TRANSDERMAL PATCHES 100mcg/h*" as Class: Fentanyl-type; Name: Fentanyl; Dosage: 0.0001g/h; Form: transdermal. The results regarding the average number of fentanyl, fentanyl analogs, and other non-pharmaceutical synthetic opioids advertised on cryptomarkets identified are shown in Table 6. The crawls considered to obtain these results are the dark web posts collected from Agora and Dream market in the years 2015 and 2018 [33]. We also classified vendors on darknet markets (Dream, Tochka, and Wall Street are the marketplaces used for this study) using DAO. The summary of our findings related to unique vendors, substance, location, vendor descriptions, and number of withdrawal transactions are shown in Table 7.

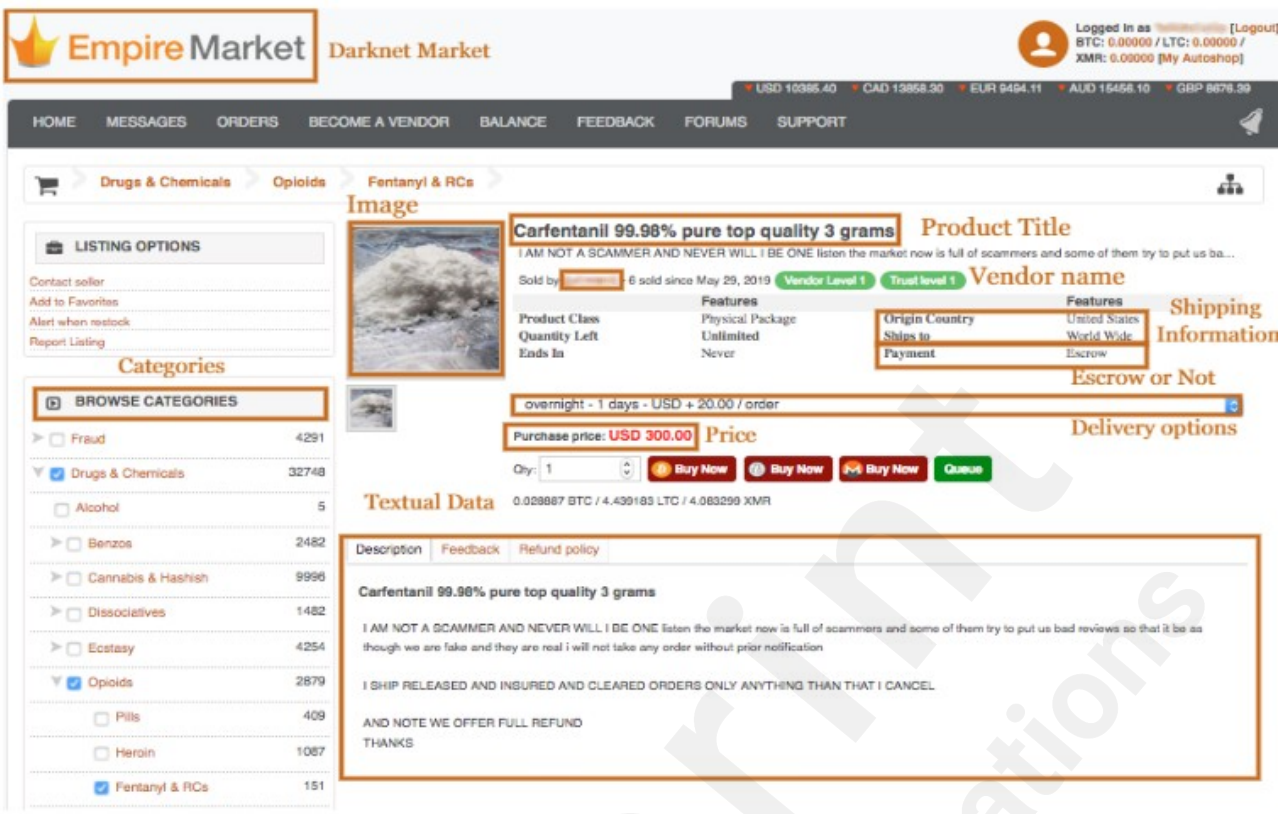


Figure 9. Snapshot of Darknet Marketplace [31]

Table 6. Average number of fentanyl, fentanyl analogs and other non-pharmaceutical synthetic opioids advertised on cryptomarkets extracted with DAO [32]

Preprint JMIR Publications

Type of substances	Average number of ads per day, by month (number of crawls)				
	AGORA			Dream market	
	03/2015	04/2015	05/2015	03/2018	04/2018
Fentanyl[a]	130	174	139	207	216
Fentanyl analogs					
Acetyl fentanyl	44	39	41	3	1
Butyr fentanyl	12	10	17	6	7
Carfentanil	0	0	0	12	5
Furanyl fentanyl	0	0	1	31	39
MethoxyAcetyl fentanyl	0	0	0	14	14
4-FluoroIsoButyr fentanyl	0	0	0	19	16
3-MethoxyMethyl fentanyl	0	0	0	2	2
Total, fentanyl analogs	56	49	59	87	84
Other non-pharmaceutical synthetic opioids					
U-47,700	5	4	5	0	3
W-18	5	4	5	0	0
MT-45	9	8	9	0	0
AH-7921	0	0	1	0	0
U-48,800	0	0	0	1	7
U-49,900	0	0	0	0	1
U-4TDP	0	0	0	0	4
U-50,488	0	0	0	8	4
MPF-47700	0	0	0	0	5
Total, other NP synth opioids	19	16	20	9	24
Other opioids[b]	827	1061	1152	3211	3137
Total (any opioids)	1033	1300	1370	3512	3460

[a]Includes mentions of Fentanyl, China White Heroin, Synthetic Heroin, and mentions of pharmaceutical fentanyl, such as Duragesic, fentanyl patches, fentanyl transdermal system

[b]Includes mentions of heroin, opium, morphine and other types of pharmaceutical opioids (e.g., hydrocodone, oxycodone, hydromorphone) excluding pharmaceutical fentanyl

Table 7. Summary of Dataset extracted from Darknet Markets using DAO [31]

Marketplace	Withdrawal # Transactions	Bitcoin	USD Equivalent	Unique # Vendors	Unique # Substance	Unique # Location	Unique # Description
Dream	261	99.1503695	$197,589.12	1448	852	356	16800
Tochka	2990	0.70483642	$5,072.33	408	313	44	1829
Wall Street	7755	2.572515	$18,729.40	466	290	29	1723

COVID Scenario

We performed a spatio-temporal analysis of the psychological impact of novel coronavirus using ~1.2 Billion tweets from January 01 to April 10, 2020 [86,87]. The concepts related to addiction and mental health in COVID-19 data were semi-automatically recognized using the entities and slang terms mentioned in DAO. Approximately 90 related concepts and 140 slang terms were used to

extract tweets mentioning illicit drug use, alcoholism, and pharmacological drug misuse. Further, suicide risk factors (such as insomnia, and depression) were observed in the tweets extracted using DAO. Likewise, we study the negative media exposure from ~700K news articles published during the COVID-19 by fine-tuning the BERT model with DAO [51]. The three months (January, February, and March) in the year 2020 are considered for our earlier study as this period has a huge COVID-19 spread as per the Mental Health America report (MHA) [88]. We used 10 of the 13 states recognized as high spread areas by this report. The 3 states that are not included in Table 8 are Washington, Wyoming, and Idaho. These 3 states are not included as the related data is not present in our dataset cohort. In this work, we reported the state-wise labels (i.e., depressive, drug abusive, informative) for each month using deep learning models vanilla BERT, depression BERT, and drug use BERT as shown in Table 8. The definitions of these deep learning models are described in Table 1. This study is followed by analyzing Social Quality Index (SQI) which aggregates mental health components (Depression, Anxiety), Addiction, and substance use disorders considering tweets in the period March to April 2020. The SQI Index and tweets for states IL, NY, MD, AZ, NM, and MA are shown in Figure 10 [87].

Table 8. Evaluation of BERT models for Mental Health America states over three months (January, February, and March 2020) [51,87]

MHA states with Depression and Drug use	vanillaBERT (2020 Months)	Druguse-BERT (2020 Months)	Depression-BERT (2020 Months)
Tennessee	Feb, Mar	Feb, Mar	Feb, Mar
Alabama	Feb	Feb, Mar	Feb
Oklahoma	Mar	Feb, Mar	Feb, Mar
Kansas	Feb	Jan, Feb	Jan, Feb
Montana	Mar	Feb	Feb, Mar
South Carolina	Mar	Mar	Feb, Mar
Alaska	Feb, Mar	Jan, Feb, Mar	Feb, Mar
Utah	Mar	Mar	Mar
Oregon	None	Feb	None
Nevada	Feb	Feb	Feb

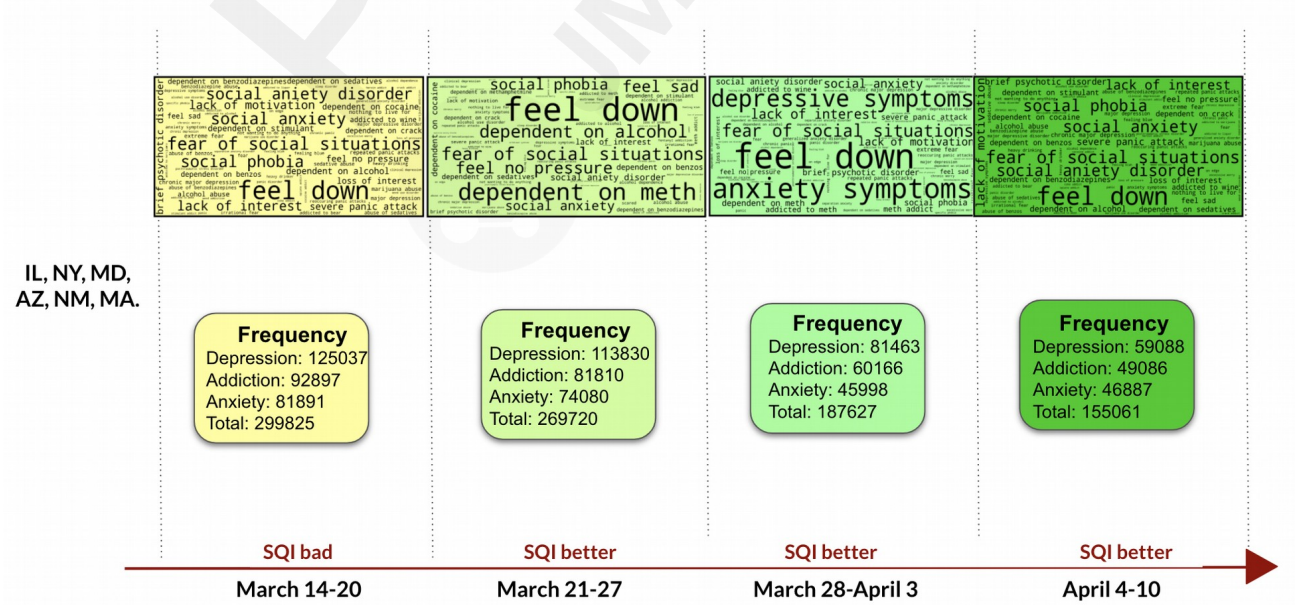


Figure 10. SQI Pattern of improvement in conditions as the decline in the number of tweets on

Depression, Addiction, and anxiety

Discussion

DAO is an ongoing project that can be continuously improved and expanded to handle additional topic areas and emerging substance use issues and trends. DAO development requires intensive, hands-on involvement from experts in the field of substance use research (domain experts). We acknowledge a limitation to our approach in that our DAO development team did not include persons with lived experiences of substance use disorders. In the future, it would be important to also involve individuals who use drugs to help, develop and refine DAO sections and terms. DAO can provide a tool and a framework for interdisciplinary collaborative teams to carry this work forward. The DAO ontology has been proven effective in several scenarios as demonstrated in Section 3. Table 9 summarizes the evolution and improvement of the ontology use according to the needs of the projects. The public health findings described in this document of associated projects with a focus on person, place, and/or time are referenced in Table 9.

Table 9. Summary of the DAO ontology implemented in projects

Domain	Related Publications	Manuscript Section	Data Type	Findings Reference
Buprenorphine, Loperamide, Other Opioids	Cameron et al. [24,39]	PREDOSE[24]	Web forum data	Figures 4, 5
User types in Marijuana-related posts on Social Media	Kursuncu et al [26], Lamy et al. [29]	eDrugTrends[26–29,89]	Twitter data, Web forums, BlueLight	Table 3
Depression DSM-5	Gaur et al. [39]	eDrugTrends[90]	Web forums, Reddit, Twitter	Tables 4, 5
Fentanyl/Fentanyl analogs, Clustering of dark web vendors	Usha et al.[33], Kumar et al.[31], Lamy et al. [32]	eDarkTrends[91]	Social Media, cryptomarket	Tables 6, 7
COVID	Gaur et al. [51,87]	COVID-19:Public Health Study [92]	Social Media	Table 8, Figure 10

In this study, we developed and evaluated DAO as a framework for identifying concepts, entities, and relationships of interest in social media posts. The DAO developed in this study comprises 315 classes, 31 relationships, and 814 instances with 2 to 4 levels deeper. Our ontology is designed for studying social media data, dark web data, and web forums. The DAO is mainly employed for knowledge extraction and is broadly applicable for these platforms.

The superclasses of our ontology integrate all the concepts regarding health conditions, individual-related, network-related, society (public policies), sources (dealers, Internet, medical, self_produced), Spatio-temporal, and substance-related classes. The integrated ontology developed in this study is suitable for analyzing social media posts, and dark web posts to understand the network-related characteristics, location/time issues, identifying new trends, synonyms, slang items, and new drugs.

Our ontology incorporates terminology not only extracted from DSM-5 but also various terms and slang used on social media and other web posts. The terminology with all the medical terms, synonyms, and slang terms representing all the substances enabled a rich collection of terms in social media and dark web data. Our ontology also helps in doing the topic discovery, and entity extraction on social media/dark web data. Along with this, we use ontology to extract information in the description of each product in dark web marketplaces to identify substances that are being sold that are not known before like synthetic drugs, research chemicals, synthetic cannabinoids, and synthetic heroin.

Following well-known software development methodologies (e.g., agile methodology), the ontology is constantly being updated according to the needs of current addiction-based research. DAO stands as a machine-processable resource that describes a collection of addiction domain-related objects and classes and is growing with the needs of the new ongoing projects. For instance, the current ontology is being enriched with knowledge from the dark web. In future work, the ontology will be linked to other ontologies (e.g., MEDDRA [93], a Medical Dictionary for Regulatory Activities) to design the Drug Abuse Knowledge Graph (DAKG). Another research contribution would be to automatically update the DAO ontology with new concepts and properties, inspired by the algorithm that allows users to interactively build topic-specific ontologies using suggestions retrieved from a knowledge graph [94].

Acknowledgments

This work was supported in part by the National Institute on Drug Abuse (NIDA) grant 5R01DA039454-02 “Trending: Social Media Analysis to Monitor Cannabis and Synthetic Cannabinoid Use”; the National Science Foundation (NSF) award 1761931 “Spokes: MEDIUM: MIDWEST: Collaborative: Community-Driven Data Engineering for Substance Abuse Prevention in the Rural Midwest”; the NIDA Grant 5R21DA044518-02 “eDarkTrends: Monitoring Darknet Markets to Track Illicit Synthetic Opioid Trends”; and the National Institutes of Health (NIH) Grant R21 DA030571-01A1 “A Study of Social Web Data on Buprenorphine Abuse using Semantic Web Technology”. Any opinions, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect that of the NSF, NIH, or NIDA.

Multimedia Appendix 1: [Figures]

References

1. UNODC World Drug Report 2020: Global drug use rising; while COVID-19 has far reaching impact on global drug markets. In: United Nations : Office on Drugs and Crime [Internet]. [cited 9 Apr 2022]. Available: https://www.unodc.org/unodc/press/releases/2020/June/media-advisory---global-launch-of-the-2020-world-drug-report.html

2. Spooner C, Hetherington K. Social determinants of drug use. National Drug and Alcohol Research Centre, University of New South Wales …; 2005.

3. Hall W, Lynskey M. Assessing the public health impacts of legalizing recreational cannabis use: the US experience. World Psychiatry. 2020. pp. 179–186. doi:10.1002/wps.20735

4. European drug report 2019: Trends and developments. [cited 15 Feb 2021]. Available: https://www.emcdda.europa.eu/publications/edr/trends-developments/2019_en

5. Kruithof K, Aldridge J, Hétu D, Sim M, Dujso E, Hoorens S. The role of the “dark web” in the trade of illicit drugs. 2016. doi:10.7249/rb9925

6. Aldridge J, Décary-Hétu D. Hidden wholesale: The drug diffusing capacity of online drug cryptomarkets. Int J Drug Policy. 2016;35: 7–15.

7. Kursuncu U, Gaur M, Lokala U, Thirunarayan K, Sheth A, Arpinar IB. Predictive Analysis on Twitter: Techniques and Applications. In: Agarwal N, Dokoohaki N, Tokdemir S, editors. Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining. Cham: Springer International Publishing; 2019. pp. 67–104.

8. Soska K, Christin N. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. 24th ${USENIX} security symposium ({USENIX}$ security 15). 2015. pp. 33–48.

9. Cardoso J, Sheth A. The Semantic Web and Its Applications. In: Cardoso J, Sheth AP, editors. Semantic Web Services, Processes and Applications. Boston, MA: Springer US; 2006. pp. 3–33.

10. UscholdMichael, GruningerMichael. Ontologies and semantics for seamless connectivity. ACM SIGMOD Record. 2004 [cited 16 Feb 2020]. Available: https://dl.acm.org/doi/abs/10.1145/1041410.1041420?casa_token=G-uVPZXXNIEAAAAA:xT0ERwJkvNnvvpj1WUUTO-C_bkUjN7sTdk7RJYxmeUJ-jmA9fgzOkvczRvVHsmB_PbhmKbLCqEGU

11. Grimm S, Abecker A, Völker J, Studer R. Ontologies and the Semantic Web. In: Domingue J, Fensel D, Hendler JA, editors. Handbook of Semantic Web Technologies. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. pp. 507–579.

12. Horrocks I. Ontologies and the semantic web. Commun ACM. 2008;51: 58–67.

13. Kim AR, Park H-A, Song T-M. Development and Evaluation of an Obesity Ontology for Social Big Data Analysis. Healthc Inform Res. 2017;23: 159–168.

14. Jung H, Park H-A, Song T-M. Ontology-Based Approach to Social Data Sentiment Analysis: Detection of Adolescent Depression Signals. J Med Internet Res. 2017;19: e259.

15. Ethier J-F, Barton A, Taseen R. An ontological analysis of drug prescriptions. Appl Ontol. 2018;13: 273–294.

16. Ceusters W. An information artifact ontology perspective on data collections and associated representational artifacts. Stud Health Technol Inform. 2012;180: 68–72.

17. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The Ontology for Biomedical Investigations. PLoS One. 2016;11: e0154556.

18. Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. Summit Transl Bioinform. 2009;2009: 116–120.

19. Hicks A, Hanna J, Welch D, Brochhausen M, Hogan WR. The ontology of medically related social entities: recent developments. J Biomed Semantics. 2016;7: 47.

20. Hanna J, Joseph E, Brochhausen M, Hogan WR. Building a drug ontology based on RxNorm and other sources. J Biomed Semantics. 2013;4: 44.

21. Kamdar MR, Hamamsy T, Shelton S, Vala A, Eftimov T, Zou J, et al. A Knowledge Graph-based Approach for Exploring the U.S. Opioid Epidemic. arXiv [cs.CY]. 2019. Available: http://arxiv.org/abs/1905.11513

22. Motlagh FG, Shekarpour S, Sheth A, Thirunarayan K, Raymer ML. Predicting public opinion on drug legalization: social media analysis and consumption trends. Proceedings of the 2019 IEEE/ACM

International Conference on Advances in Social Networks Analysis and Mining. New York, NY, USA: Association for Computing Machinery; 2019. pp. 952–961.

23. Daniulaityte R, Carlson R, Falck R, Cameron D, Perera S, Chen L, et al. “I just wanted to tell you that loperamide WILL WORK”: A web-based study of extra-medical use of loperamide. Drug Alcohol Depend. 2013;130: 241–244.

24. Cameron D, Smith GA, Daniulaityte R, Sheth AP, Dave D, Chen L, et al. PREDOSE: a semantic web platform for drug abuse epidemiology using social media. J Biomed Inform. 2013;46: 985–997.

25. Daniulaityte R, Carlson R, Brigham G, Cameron D, Sheth A. “Sub is a weird drug:” A web-based study of lay attitudes about use of buprenorphine to self-treat opioid withdrawal symptoms. Am J Addict. 2015;24: 403–409.

26. Kursuncu U, Gaur M, Lokala U, Illendula A, Thirunarayan K, Daniulaityte R, et al. What’s ur Type? Contextualized Classification of User Types in Marijuana-Related Communications Using Compositional Multiview Embedding. 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI). 2018. pp. 474–479.

27. Daniulaityte R, Nahhas RW, Wijeratne S, Carlson RG, Lamy FR, Martins SS, et al. “Time for dabs”: Analyzing Twitter data on marijuana concentrates across the U.S. Drug and Alcohol Dependence. 2015. pp. 307–311. doi:10.1016/j.drugalcdep.2015.07.1199

28. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. “When ‘Bad’ is ‘Good’”: Identifying Personal Communication and Sentiment in Drug-Related Tweets. JMIR Public Health Surveill. 2016;2: e162.

29. Lamy FR, Daniulaityte R, Sheth A, Nahhas RW, Martins SS, Boyer EW, et al. “Those edibles hit hard”: Exploration of Twitter data on cannabis edibles in the U.S. Drug Alcohol Depend. 2016;164: 64–70.

30. Daniulaityte R, Lamy FR, Smith GA, Nahhas RW, Carlson RG, Thirunarayan K, et al. “Retweet to Pass the Blunt”: Analyzing Geographic and Content Features of Cannabis-Related Tweeting Across the United States. J Stud Alcohol Drugs. 2017;78: 910–915.

31. Kumar R, Yadav S, Daniulaityte R, Lamy F. eDarkFind: Unsupervised Multi-view Learning for Sybil Account Detection. Proceedings of The. 2020. Available: https://dl.acm.org/doi/abs/10.1145/3366423.3380263?casa_token=tx4QUY96qc4AAAAA:GXVQF5CVyjq6iVcOlDVfcbGvkutJHQPz_rqTZDBYfWpaEDYyGbCiZ5X0MICH6Tk5wHvH1IkxYOT-

32. Lamy FR, Daniulaityte R, Barratt MJ, Lokala U, Sheth A, Carlson RG. Listed for sale: Analyzing data on fentanyl, fentanyl analogs and other novel synthetic opioids on one cryptomarket. Drug Alcohol Depend. 2020;213: 108115.

33. Lokala U, Lamy FR, Daniulaityte R, Sheth A, Nahhas RW, Roden JI, et al. Global trends, local harms: availability of fentanyl-type drugs on the dark web and accidental overdoses in Ohio. Comput Math Organ Theory. 2018; 1–12.

34. Ontology101. [cited 10 Apr 2022]. Available: https://protegewiki.stanford.edu/wiki/Ontology101

35. COR by ESIPFed. [cited 10 Apr 2022]. Available: http://esipfed.github.io/cor/

36. About DBpedia. In: DBpedia Association [Internet]. [cited 9 Apr 2022]. Available: https://wiki.dbpedia.org/about

37. Ontology Metrics. [cited 9 Apr 2022]. Available: http://protegeproject.github.io/protege/views/ontology-

metrics/

38. Gyrard A. PerfectO: Semantic Web Best Practices. [cited 20 Feb 2020]. Available: http://perfectsemanticweb.appspot.com/

39. Gaur M, Kursuncu U, Alambo A, Sheth A, Daniulaityte R, Thirunarayan K, et al. “Let Me Tell You About Your Mental Health!” Proceedings of the 27th ACM International Conference on Information and Knowledge Management - CIKM ’18. 2018. doi:10.1145/3269206.3271732

40. Link V, Lohmann S, Marbach E, Negru S. WebVOWL - Web-based Visualization of ontologies. [cited 10 Apr 2022]. Available: http://vowl.visualdataweb.org/webvowl.html

41. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Given LM, et al. Social media use among patients and caregivers: a scoping review. BMJ Open. 2013;3. doi:10.1136/bmjopen-2013-002819

42. Bluelight.org. [cited 15 Feb 2021]. Available: https://www.bluelight.org/xf/

43. eDrugTrends | Center for Interventions, Treatment, and Addictions Research (CITAR) | Boonshoft School of Medicine | Wright State University. [cited 20 Feb 2020]. Available: https://medicine.wright.edu/citar/edrugtrends

44. eDarkTrends | Center for Interventions, Treatment, and Addictions Research (CITAR) | Boonshoft School of Medicine | Wright State University. [cited 20 Feb 2020]. Available: https://medicine.wright.edu/citar/edarktrends

45. Scholl L. Drug and Opioid-Involved Overdose Deaths — United States, 2013–2017. MMWR Morb Mortal Wkly Rep. 2019;67. doi:10.15585/mmwr.mm6751521e1

46. Wilson N, Kariisa M, Seth P, Smith H 4th, Davis NL. Drug and Opioid-Involved Overdose Deaths - United States, 2017-2018. MMWR Morb Mortal Wkly Rep. 2020;69: 290–297.

47. Panchal N, Kamal R, Orgera K, Cox C, Garfield R, Hamel L, et al. The implications of COVID-19 for mental health and substance use. Kaiser family foundation. 2020. Available: https://abtcounseling.com/wp-content/uploads/2020/09/The-Implications-of-COVID-19-for-Mental-Health-and-Substance-Use-_-KFF.pdf

48. Garfin DR, Silver RC, Holman EA. The novel coronavirus (COVID-2019) outbreak: Amplification of public health consequences by media exposure. Health Psychol. 2020;39: 355–357.

49. Holmes EA, O’Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry. 2020;7: 547–560.

50. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. General psychiatry. 2020;33. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/pmc7061893/

51. Alambo A, Gaur M, Thirunarayan K. Depressive, Drug Abusive, or Informative: Knowledge-aware Study of News Exposure during COVID-19 Outbreak. arXiv [cs.SI]. 2020. Available: http://arxiv.org/abs/2007.15209

52. National Institute on Drug Abuse. Drugabuse.gov. [cited 11 Feb 2021]. Available: https://www.drugabuse.gov/

53. Home. [cited 11 Feb 2021]. Available: https://www.dea.gov/

54. EMCDDA home page. [cited 11 Feb 2021]. Available: https://www.emcdda.europa.eu/emcdda-home-page_en

55. RxNorm. 2004 [cited 11 Feb 2021]. Available: https://www.nlm.nih.gov/research/umls/rxnorm/index.html

56. Daniulaityte R, Carlson RG, Kenne DR. Initiation to Pharmaceutical Opioids and Patterns of Misuse: Preliminary Qualitative Findings Obtained by the Ohio Substance Abuse Monitoring Network. J Drug Issues. 2006;36: 787–808.

57. Daniulaityte R, Carlson RG. "To Numb Out and Start to Feel Nothing": Experiences of Stress among Crack-Cocaine Using Women in a Midwestern City. J Drug Issues. 2011;41: 1–24.

58. Daniulaityte R, Falck R, Carlson RG. Illicit use of buprenorphine in a community sample of young adult non-medical users of pharmaceutical opioids. Drug Alcohol Depend. 2012;122: 201–207.

59. Daniulaityte R, Carlson RG, Kenne DR. Methamphetamine use in Dayton, Ohio: preliminary findings from the Ohio Substance Abuse Monitoring Network. J Psychoactive Drugs. 2007;39: 211–221.

60. Lamy FR, Daniulaityte R, Zatreh M, Nahhas RW, Sheth A, Martins SS, et al. "You got to love rosin: Solventless dabs, pure, clean, natural medicine." Exploring Twitter data on emerging trends in Rosin Tech marijuana concentrates. Drug Alcohol Depend. 2018;183: 248–252.

61. Dl NFNM. Ontology Development 101: A Guide to Creating Your First Ontology. Stanford University, Stanford CA, http://protege. stanford. edu/publications …; 2001.

62. Musen MA, Protégé Team. The Protégé Project: A Look Back and a Look Forward. AI Matters. 2015;1: 4–12.

63. Lin Y, Liu Z, Sun M, Liu Y, Zhu X. Learning Entity and Relation Embeddings for Knowledge Graph Completion. AAAI. 2015;29. Available: https://ojs.aaai.org/index.php/AAAI/article/view/9491

64. Lehmann J, Isele R, Jakob M, Jentzsch A. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. Semant Pragmat. 2015. Available: https://content.iospress.com/articles/semantic-web/sw134

65. Link V, Lohmann S, Marbach E, Negru S, Wiens V. WebVOWL. [cited 20 Feb 2020]. Available: http://visualdataweb.de/webvowl/#iri=http://purl.org/dao#

66. Call for Resources Track Papers « ISWC 2018. [cited 20 Feb 2020]. Available: http://iswc2018.semanticweb.org/call-for-resources-track-papers/

67. 01 - Step By Step Tutorial to Improve ontologies (e.g., FIESTA-IoT, IEEE 1872.1 Autonomous robotics, DAO, empathi, asthma) - Semantic Web Best Practices. In: Google Docs [Internet]. [cited 20 Feb 2020]. Available: https://docs.google.com/presentation/d/1eZ8-OtTnt7iNmkFvJuSrILF8iaKbiW5yjBuS1ybyxq0/edit?usp=sharing

68. LODE - Live OWL Documentation Environment. [cited 21 Feb 2020]. Available: http://www.essepuntato.it/lode

69. Ontology Metrics. [cited 15 Feb 2021]. Available: http://protegeproject.github.io/protege/views/ontology-metrics/

70. García J, García-Peñalvo FJ, Therón R. A Survey on Ontology Metrics. Knowledge Management, Information Systems, E-Learning, and Sustainability Research. Springer Berlin Heidelberg; 2010. pp. 22–27.

71. Web ontology language (OWL) abstract syntax and semantics: Abstract syntax. [cited 15 Feb 2021]. Available: https://www.w3.org/TR/2002/WD-owl-semantics-20021108/syntax.html

72. Estival D, Nowak C, Zschorn A. Towards ontology-based natural language processing. Proceeedings of the Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology. Association for Computational Linguistics; 2004. pp. 59–66.

73. Daniulaityte R, Carlson RG, Falck RS, Cameron DH, Udayanaga S, Chen L, et al. A Web-Based Study of Self-Treatment of Opioid Withdrawal Symptoms with Loperamide. 2012 [cited 8 Apr 2022]. Available: https://corescholar.libraries.wright.edu/knoesis/624/

74. Sheth A, Perera S, Wijeratne S, Thirunarayan K. Knowledge will Propel Machine Understanding of Content: Extrapolating from Current Examples. Proc IEEE WIC ACM Int Conf Web Intell Intell Agent Technol. 2017;2017: 1–9.

75. Knoesis Center. PREDOSE Demo. Youtube; 16 Sep 2013 [cited 16 Feb 2020]. Available: https://www.youtube.com/watch?v=gCFPzMgEPQM

76. Sheth A, Kapanipathi P. Semantic Filtering for Social Data. IEEE Internet Comput. 2016;20: 74–78.

77. eDrugTrends | Center for Interventions, Treatment, and Addictions Research (CITAR) | Boonshoft School of Medicine | Wright State University. [cited 20 Feb 2020]. Available: https://medicine.wright.edu/citar/edrugtrends

78. International Classification of Diseases, Version 10 | NCBO BioPortal. [cited 20 Feb 2020]. Available: https://bioportal.bioontology.org/ontologies/ICD10

79. SNOMED CT | NCBO BioPortal. [cited 20 Feb 2020]. Available: http://bioportal.bioontology.org/ontologies/SNOMEDCT

80. bioCaddie Core Development Team. Home - DataMed | bioCADDIE Data Discovery Index. [cited 20 Feb 2020]. Available: https://datamed.org

81. Research domain criteria (RDoC). In: National Institute of Mental Health (NIMH) [Internet]. [cited 10 Apr 2022]. Available: https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc

82. Alambo A, Gaur M, Lokala U, Kursuncu U, Thirunarayan K, Gyrard A, et al. Question Answering for Suicide Risk Assessment Using Reddit. 2019 IEEE 13th International Conference on Semantic Computing (ICSC). 2019. pp. 468–473.

83. Park A, Conway M. Harnessing Reddit to Understand the Written-Communication Challenges Experienced by Individuals With Mental Health Disorders: Analysis of Texts From Mental Health Communities. J Med Internet Res. 2018;20: e121.

84. Saravia E, Chang C, De Lorenzo RJ, Chen Y. MIDAS: Mental illness detection and analysis via social media. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2016. pp. 1418–1421.

85. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJP, Dobson RJB, et al. Characterisation of mental health conditions in social media using Informed Deep Learning. Sci Rep. 2017;7: 45141.

86. Manas Gaur, Vedant Khandelwal, Ugur Kurşuncu, Vishal Pallagani. Measuring Spatio-Temporal Psychological Impact of Novel Coronavirus through Social Quality Index. [cited 27 Jun 2020]. Available: https://youtu.be/XzYrn0PEzNk

87. Gaur M, Kursuncu U, Khandelwal V, Pallagani V, Shalin V, Sheth A. Psychidemic: Measuring the

Spatio-Temporal Psychological Impact of Novel Choronovirus with a Social Quality Index. Computing Research Association Annual Conference. 2020. Available: https://scholarcommons.sc.edu/aii_fac_pub/529/

88. Ranking the states. [cited 11 Feb 2021]. Available: https://www.mhanational.org/issues/ranking-states

89. Daniulaityte R, Carlson RG, Golroo F, Wijeratne S, Boyer EW, Martins SS, et al. " Time for dabs": Analyzing Twitter data on butane hash oil use. 2015. Available: https://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=2434&context=knoesis

90. eDrugTrends | Center for Interventions, Treatment, and Addictions Research (CITAR) | Boonshoft School of Medicine | Wright State University. [cited 20 Feb 2020]. Available: https://medicine.wright.edu/citar/edrugtrends

91. eDarkTrends | Center for Interventions, Treatment, and Addictions Research (CITAR) | Boonshoft School of Medicine | Wright State University. [cited 20 Feb 2020]. Available: https://medicine.wright.edu/citar/edarktrends

92. Covid19 - knoesis wiki. [cited 8 Apr 2022]. Available: http://wiki.aiisc.ai/index.php/Covid19

93. Mozzicato P. MedDRA. Pharmaceut Med. 2009;23: 65–75.

94. Böhm K, Ortiz M. A Tool for Building Topic-specific Ontologies Using a Knowledge Graph. Description Logics. 2018. Available: http://ceur-ws.org/Vol-2211/paper-10.pdf

84. http://cor.esipfed.org/ont/~ushanri/DAO

85. http://visualdataweb.de/webvowl/#iri=http://cor.esipfed.org/ont/~ushanri/DAO

86. http://oops.linkeddata.es/response.jsp?uri=http://cor.esipfed.org/ont/~ushanri/DAO

87. http://linkeddata.uriburner.com:8000/vapour?uri=http://cor.esipfed.org/ont/~ushanri/DAO

88. http://graphite.ecs.soton.ac.uk/checker/?uri=http://cor.esipfed.org/ont/~ushanri/DAO

89.http://perfectsemanticweb.appspot.com/documentation/SemanticWebBestPracticesForDummies.pdf

Supplementary Files

Preprint JMIR Publications

Multimedia Appendixes

Glossary.
URL: http://asset.jmir.pub/assets/b96631887ed5c0b51adc4e5bc0a0193d.docx

Preprint JMIR Publications