Journal of Intelligence Studies in Business

Using open data and Google search data for competitive intelligence analysis

Jan Černý[a], Martin Potančok[a]*, Zdeněk Molnár[a]

[a]Faculty of Informatics and Statistics, Department of Information Technologies, University of Economics, Prague, Czech Republic

*martin.potancok@vse.cz

PLEASE SCROLL DOWN FOR ARTICLE

JISIB

Using open data and Google search data for competitive intelligence analysis

Jan Černý[a], Martin Potančok[a]* and Zdeněk Molnár[a]

[a]*Faculty of Informatics and Statistics, Department of Information Technologies, University of Economics, Prague, Czech Republic*

Corresponding author (*): martin.potancok@vse.cz

ABSTRACT Open data are information entities that are of significant importance for many institutions, businesses and even citizens as the part of the digital transformation within many fields in our society. The aim of this paper is to provide a competitive environment analysis method using open source intelligence within the pharmaceutical sector and to design the optimal data structure for this purpose. Firstly, we have described the state-of-the-art of open human medicine data within the European Union with a focus on antidepressants and we have chosen the Czech Republic as the primary research territory for demonstrating competitive intelligence analysis. Secondly, we have identified the competitive intelligence and open source intelligence relationship with a new possible contextual analysis method using open human medicine data and Google Search data. Finally, this paper shows the potential of open deep web data within competitive intelligence activities, together with surface web data entities as a low-cost approach with high intelligence value focused on the pharmaceutical market.

---

1. INTRODUCTION

Open data plays a significant role in our present society and is one of the most important digital transformation trends. Moreover, it has become a solid part of the activities of business units that are charged with business analyses, insights and strategy plans (Janssen et al. 2012). The reason can be found in a very broad spectrum of industries and areas where open data has started to be a rational form of result output. As the number and scope of such open datasets has grown enormously to include in the areas of transportation, public services, natural science, education, demography, and last but not least the health sector, it has also become a significant part of many national information policies, shifting from governmental down to local levels. In the USA, a growing trackable significance was evident during the Obama administration after the official Data.gov site was launched (Kostkova et al. 2016). Data is also an essential part of the EU's Digital Single Market strategy, as "The EU needs to ensure that data flows across borders and sectors and disciplines. This data should be accessible and reusable by most stakeholders in an optimal way." (European Commission 2018). Moreover, massive digitalization and increasing information system/information and communication technology (IS/ICT) usage have brought big data challenges and demands for non-traditional analytical methods to uncover global and regional trends (Gandomi and Haider 2015).

This means, therefore, that open data appears to be a strong tool for a spectrum of competitive intelligence (CI) and open source intelligence (OSINT) methodologies at all levels of different industries and organization types. CI can be defined at its basic level as the process of planning, collecting and disseminating data, information and knowledge for the purpose of better decision-making, eliminating risks and uncovering of business opportunities, primarily in an external company environment (Grèzes 2015). The first phase identifies particular business information needs through key intelligence topics (KIT), and then key intelligence questions (KIQ) which analysts use for the collection process as they define information requests (Herring 1999). OSINT consists of a very similar cycle as CI, but it also consists of an open data and information source mining process through the collection phase. In addition, OSINT end-users do not come primarily from business, but from government, military, intelligence services and from the security service sector.

In the present paper we have focused on open human medicine data from two perspectives: government and business. In both cases we wanted to design open data intelligence analysis methods for the public sector, e.g. policy makers, ministers, commissioners and other key persons. Our governmental direction is directed towards setting up the optimal open data structure, and our business direction is directed towards complex business environment analysis. To demonstrate our intent, we have chosen open human medicine data focused on antidepressants. To increase specificity, we have added a Google Search data perspective to gain a territorial dimension for our analysis.

Our two main research questions are: could open data provide significant CI insights within the pharmaceutical industry? and could the surface web search data deliver a territorial perspective to anonymous open human medicine data?

2. LITERATURE REVIEW

There are several studies of how open data could help in the health sector. For example, Bernard et al. (2018) seek possible open source solutions that could be used for the detection, reporting and control of disease outbreaks, and analyze the previous use of similar tools in Ebola and SARS epidemics. One part of this work also considers the ethical level (Oubrich

2011) of these intelligence activities within the context of data ownership. This question is also discussed by Kostkova et al., (2016). Brownstein, et al. (2008) find the power of public information sources in the signal intelligence scope for outbreak-oriented detection activities to be at the local level of information sources such as discussion sites, disease reporting networks, and news outlets (with regards to a very detailed verification process). Google Search data played an important role in the past within the Google Flu Trends project. As Cook, et al. (2011) show in their evaluation, this tool was highly accurate in the prediction of influenza activity in the United States based on user search queries. Akhgar, et al. (2016) demonstrate the complex usage of OSINT methods, however the critical issue is also focused on an early warning system for health hazards. Open innovations (Hughes 2017) and open data (European Commission 2019) initiatives are also more visible in the health sector over recent years. For example, Cantor, et al. (2018) developed a dataset for community-level social determinants of health and strengthened the decision-making process for care planning. Farber (2017) discusses whether data repositories can help find effective treatments for complex diseases. His suggestions for informatics communities consist of methods concerning the provision of an effective data infrastructure with the inexpensive method of data accessibility from different datasets, monitoring the growth of biomedical datasets and finding ways to link data in different repositories.

Perer and Gotz (2013) and Hu, et al. (2016) have illustrated how data-informed and data-driven decisions can be supported by data visualization in the health sector. To achieve appropriate data visualization several principles should be followed (e.g. simplify, compare, explore) (Few 2012).

2.1 Survey and preparations

The quality of open datasets is an aspect of many discussions at all levels of the policy-making process. Policy makers put high pressure on the availability and frequency of the "update" aspect, but we would like to raise concern over the poor open data structure concept with regards to quality. Within this context, we highlight our recent study directed toward open human medicine data in the European Union (Cerny et al. 2018). During the first three months, we made a large survey

of national medicine control offices and their open data policies. The results showed significant differences.

Further, we have designed a method for the new CI approach and demonstrated the context for open human medicine data and Google Search data. There were a number of reasons for this step. To begin with, according to our secondary survey, five billion Google queries are conducted per day, and even a small sample of this amount could lead to significant insights. The second reason is that national control offices provide datasets strictly anonymously, with no territorial information. And, through the Google Trends application, we were able to mine the information-seeking behavior of surface web users and get the following data entities:

- The searcher interest rate

- The territorial origin of the searchers (region, city)

- Trend keywords connected to our desired terms

3. KEY INTELLIGENCE QUESTIONS

After we defined the research questions, we continued and narrowed our information needs through the following key intelligence questions (KIQ):

- KIQ1: Is antidepressant use increasing in the Czech Republic?

- KIQ2: What is the most prescribed antidepressant on the market?

- KIQ3: Who is the key player in the specific market?

- KIQ4: What is the market share of antidepressants on the specific market?

- KIQ5: How can Google Search data help to determine territorial information seeking behavior regarding antidepressant-oriented queries?

4. MATERIAL AND METHODS

4.1 Medicine data structure

We have analyzed data accessibility from national control agencies that are in charge of regulatory and distribution policies in the European Union, Switzerland, Norway and Turkey. When we contacted each agency, we collected information about time response, level of content relevancy feedback and the human factor based on their ability to help regarding the requested data collection. We were aware that the results from this primary research are strictly qualitative and could be misleading, so we broadened the timeframe of the research to three months. E-mail communication has been chosen as the first method of contact, however, in specific cases phone communication was also needed.

Secondly, we went through all the possible information sources, e.g. official websites, repositories and FTP servers, and monitored whether the open human health datasets are available and how they are handled with respect to their format. If the datasets did not exist online, we concentrated on a search system interface that could be used to generate datasets with the required data fields. If there was no evidence of the existence of open data, we contacted the person in charge of communication to gather information about the state of the open data policy. As our aim is to gain market insights regarding antidepressants, we have focused on the specific data entities that could lead to quality business analysis. Table 1 suggests the fields we have monitored and that, in our opinion, could uncover specific market trends. Here we explain why our suggested data fields should be considered to be key information elements for advanced CI business analysis.

Table 1 Designed data field for a complex business analysis.

ATC	Anatomical Therapeutic Chemical classification system
CODE_MED	Specific code of the medicine
NAME	Name of the medicine
ADDITIONAL_ INFORMATION	Additional information to the name
PRODUCER	Registration holder
COUNTRY_ORIGIN_ PRODUCER	Country of a registration holder
NUMBER_OF_ PACKAGES_YEAR	Number of packages / year
PRICE_NOSUR CHARGE_EXCL_VAT	Price per package excl. a surcharge and VAT
TOTAL_SUM_NOSUR CHARGE_EXCL_VAT	Total sum / all packages / excl. a surcharge and VAT
PRICE_SURCHARGE_ INCL_VAT	Price per package incl. a surcharge and VAT
TOTAL_SUM_SURCH ARGE_INCL_VAT	Total sum / all packages / incl. a surcharge and VAT
NUMBER_DDD	Defined daily doses / package
TOTAL_DDD	Defined daily doses / total
DDD_1000INH_DAY	Defined daily doses / 1000 inhabitants

Firstly, the ATC code (WHO 2018) is the internationally respected classification in the pharmaceutical field. We can demonstrate its role in our case study. As shown below, we have chosen the N06A group, but if the specific active component is needed for analysis, we could narrow it down and be more specific.

- **N** NERVOUS SYSTEM
- **N06** PSYCHOANALEPTICS
- **N06A** ANTIDEPRESSANTS
- **N06AA** Non-selective monoamine reuptake inhibitors
- **N06AB** Selective serotonin reuptake inhibitors
- **N06AF** Monoamine oxidase inhibitors, non-selective
- **N06AG** Monoamine oxidase A inhibitors
- **N06AX** Other antidepressants

The lowest level of the classification is further divided into specific medicines and this could be a crucial factor for resolving the situation when the datasets do not include commercial medicine names. For example, class N06AA (non-selective monoamine reuptake inhibitors) covers subclass N06AA01 (desipramine) along with information about the daily defined dose (DDD). In this scenario we would use MeSH Browser (U.S. National Library of Medicine 2019) to uncover commercial names, e.g. Pertofran, Norpramin among others. The specific code of the medicine supports ATC codes as the existence confirmation identifier of the specific medicine. The name of the medicine, its additional information and the producer, together with the country of origin, are the basic identifiers of any possible commercial entity analysis. The significance of the market activity of a given producer, or possibly of a specific medicine, uncovers the total number of prescribed packages with their total cost with no surcharge and excluding value added tax. Additional price fields are used for the price comparison of individual medicines.

4.2 Google data structure

Further, our intention was directed towards the process that could verify our open human medicine data CI market analysis results. If we were able to get detailed market data about pharmaceutical companies, we would also need to add territorial information, which is crucial because of the strict anonymity of open human medicine data. Through the Google Trends application data, we were able to mine the information-seeking behavior of surface web users and obtain the following data entities:

- The searcher interest rate with retrospectivity to the year 2004
- The territorial origin of the searchers (region, city)
- Trend keywords connected to our desired terms
- We have structured the Google Search data sets as follows:
- Country
- Search term (the keyword antidepressant in a given national language and in English)
- Week (in a specific year)
- Number of searches in given country
- Region
- Number of searches in given region

5. RESULTS

5.1 Open human medicine data analysis results

The data collected reflect the present level of open human health data quality and accessibility. We went through all three levels of the collection process and found significant differences. The biggest issues we faced could be identified as the different data structure in each of the countries together with language barriers leading to difficulties as to when data should be used in a whole-region analysis. Some datasets were complex (e.g. the Czech Republic and Slovenia), while others provided only simple insights into specific medicines, e.g. Wales or Slovakia, and others, e.g. Bulgaria or Greece, had no data. Although some countries had neither open repositories nor data files accessible, a few of them did provide a specific search interface that could be used for searching, filtering and exporting open medicine data. This approach is advantageous because the exported files already include the requested class of the medicine. Excluding France, we could define the classes in the search forms with the specific ATC code. Poland, Croatia and Lithuania especially have powerful search interfaces.

The third level of the data collection phase found significant differences between the information services of the agencies. Table 2 summarizes the response time, with comments. Where references are mentioned, the agency provided links to repositories, or to search interfaces.

Table 2 Survey of agency information service time response.
R = days until response.

Entity	R	Institution Remarks
Austria	1	AGES (1 day), BASQ did not respond
Bulgaria	5	BDA (no data availability)
Croatia	-	HALMED (no response)
Czech Republic	1	SÚKL (reference to the Czech datasets)
Estonia	5	REAM (did not provide datasets with requested fields)
Finland	-	FIMEA (no response)
France	-	AMELI (no response)
Hungary	20	OGYÉI (references)
Germany	7	Several institutions contacted. Only paid datasets
Italy	-	AIFA (no response)
Latvia	5	ZVA (reference to the search interface)
Malta	3	Medicine Authority Malta (limited data availability)
Netherlands	4	CBG-MEB (do not provide requested data publicly)
Norway	7	NORPD (references)
Poland	5	URPL (no requested data availability)
Portugal	17	Infarmed (provided data only for study purposes)
Romania	5	ANM (references)
Turkey	6	TITCK (references)
Slovakia	2	ŠÚKL (only paid datasets)
Slovenia	-	JAZMP (no response)
Spain	3	AEMPS (no cooperation)
Sweden	15	LMF – Läkemedelsverket (references)
Switzerland	2	Interpharma (only limited data)
United Kingdom	18	MHRA (contacted several times, references)

During the collection process, we dealt mainly with data structure and data quality obstacles and did not get relevant support for our open data CI analysis research. The file formats and structure field values were different in every country. Moreover, the data quality implied high time costs in preparation for data analysis, especially when we dealt with the company and medicine name differences in each of the analyzed countries.

For the purpose of this paper we have chosen the open data CI analysis possibilities in the dataset from the Czech Republic. Firstly, the Czech dataset structure and quality was the most complex of the monitored countries and it is a great example of what can be achieved by open data. Secondly, we were able to make valuable market insights, even though the complexity of the data from the Czech Republic provided a powerful example of what can be achieved regarding competitive business intelligence. However, then we added the comparison possibility between the states with less structured content to demonstrate the minimum analysis context. The requested class of medicine was antidepressants, according to research question two. We have used Tableau (2019) to create an interactive visualization which can be shared and analyzed (Datig and Whiting 2018).

By focusing on the Czech Republic, we can gain very detailed insights. To begin with, we wanted to analyze the antidepressant consumption trend among Czech citizens (KIQ1). We used an open dataset covering the time period from 1991 to 2018. Figure 1 demonstrates the increase in antidepressant consumption during this period.
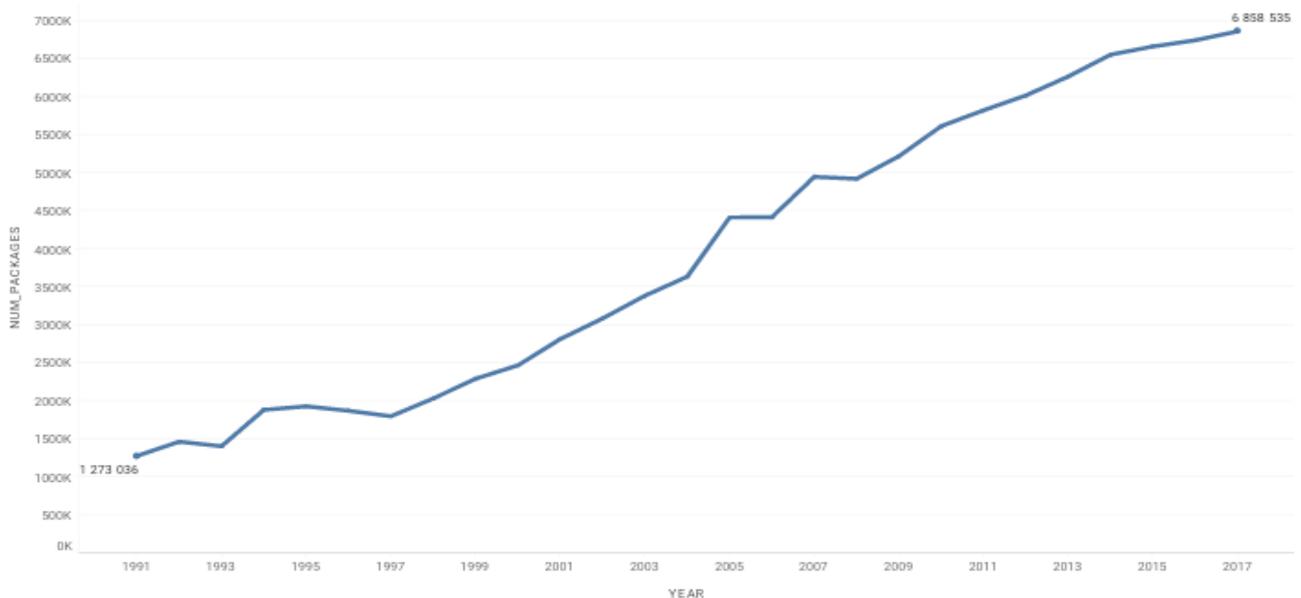

Figure 1 Czech antidepressant consumption 1991-2018.

Table 3 Czech antidepressant medicine leader market insights through open data 2009-2018 with prescribing information.

Producer	Medicine Name	Total packages	Percent packages	Treatment (from prescribing information)
H. Lundbeck	Cipralex	5 891 747	23.29	Depression and anxiety disorders (panic disorder with or without agoraphobia, social anxiety disorder, generalized anxiety disorder and obsessive-compulsive disorder)
Zentiva	Citalec	4 976 308	22.19	Depression and anxiety disorders
Pfizer	Zoloft	3 881 554	15.34	Depression with or without anxiety, panic disorder and obsessive-compulsive disorder and the treatment of post-traumatic stress disorder
Krka	Asentra	3 843 125	15.19	Depression and prevention of depression (adult), social anxiety disorder (in adults), post-traumatic stress disorder (in adults), panic disorder (in adults), obsessive compulsive disorder (in adults and children and adolescents aged 6-17)
Angelini	Trittico	3 763 352	14.88	Anti-anxiety, tension, restlessness, sleep disturbance, and sexual function

In the context of this trend, our further point of interest was to uncover the most prescribed antidepressants (KIQ2, KIQ3) and their market share in the country. We have narrowed the time period, as demonstrated in Table 3, to get the most accurate market data. This step was necessary due to significant market changes, e.g. Prothiaden was de facto the most prescribed antidepressant medicine until the year 2005, and then its popularity fell rapidly.

As we can uncover the main medicine representatives in the Czech Republic, we can connect these with the types of the mental disorder as shown in Table 3. This is used to predict mental health trends in a particular area.

However, thanks to open data, we can monitor the whole market share of antidepressants (KIQ4) and compare whether the producer of the main medicine representatives is similar to the whole antidepressant market share (Table 4).

Table 4 Czech antidepressant market share 2009-2018.

Producer	Total antidepressant packages	Percent of all packages sold
Zentiva	12 193 957	19,46
Krka	9 838 079	15,70
H. Lundbeck	7 680 389	12,25
Angelini	4 006 332	6,39
Pfizer	3 974 158	6,34

5.2 Google search data results

To analyze the context between open human medicine data and information seeking behavior we have used Google Trends (Nuti et al. 2014) including Google Search data (the context description is given above). The aim of the analysis was to confirm a correlation between Google Search data and market information about specific antidepressant consumption. Google Trends (available on trends.google.com) providing Google Search data in an available form and as confirmed by (Nuti et al. 2014) and (Nuti et al. 2014) is used by the health sector.

The identified set of related keywords (general antidepressant terms and specific names of the medicine) was gradually inserted into Google Trends, all data was downloaded into a CSV file and aggregated. It is important to emphasize that data was downloaded at the regional level for the period of analysis. Consolidated CSV files were used as a basis for the following analysis (Figure 2).

The conversion per capita was used for the analysis to ensure comparable results between countries with different population sizes. The overall analysis done in number of searches per capita shows an increase in searches since 2011 and confirms the increase in consumption based on the analysis above. As the analysis shows, the relationship between the number of searches and the number of searches per capita is not affected only by size of the country, but also by other factors. Norway, Estonia, Switzerland, Netherlands and Austria are among the countries with the largest number of searches per capita. Both number of searches
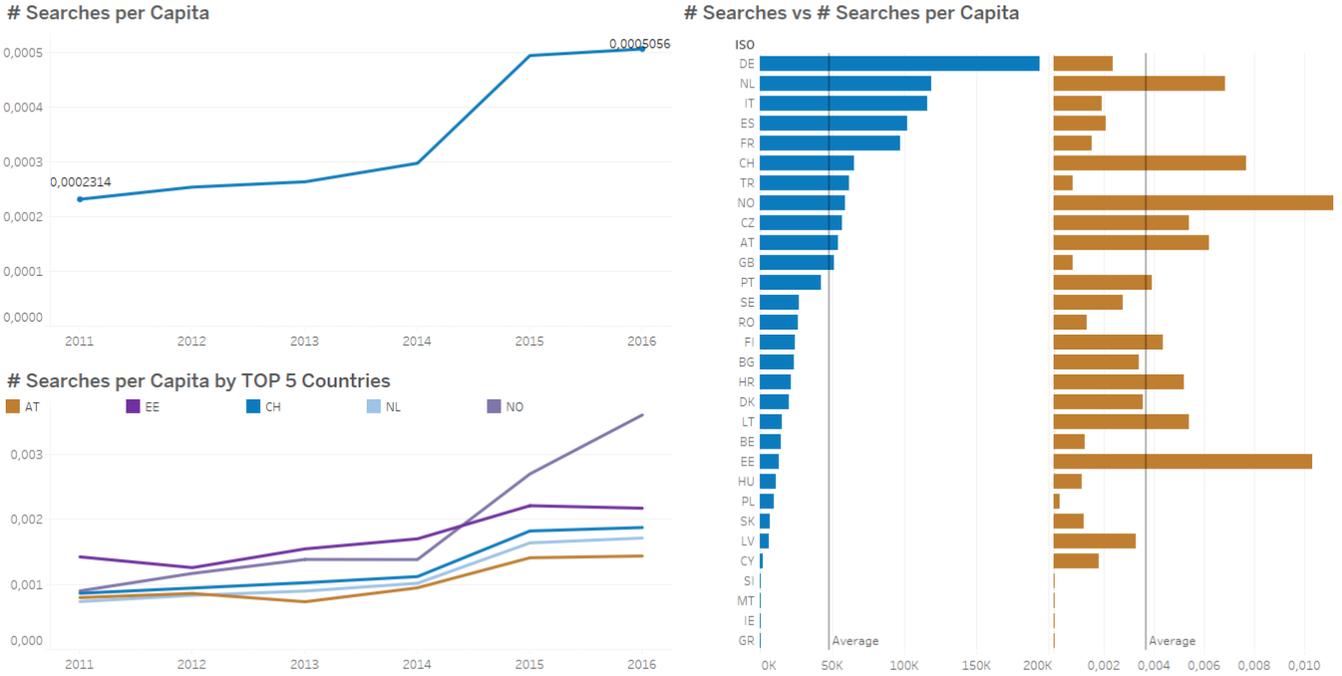
Search per Capita Analysis

Figure 2 Search per capita analysis.

and number of searches per capita are above average in the Czech Republic.

It is important to analyze the correlation between Czech market trends and Czech searching trends. We have chosen the two most prescribed medicines in the Czech Republic, Cipralex and Citalec, and compared their market performance with Google search performance (Figure 3 and Figure 4).

We demonstrate with these analyses that there is a significant similarity between market data and Google data. To sum up, if the

state control office does not provide open human medicine data with territorial dimension, we can use Google data (KIQ5) to narrow our market analysis (Table 5 and Table 6).

6. CONCLUSIONS

Open human health data can be considered to be crucial information entities for competitive environment analysis and for showing particular health trends across a large geographic area. There are two conditions that
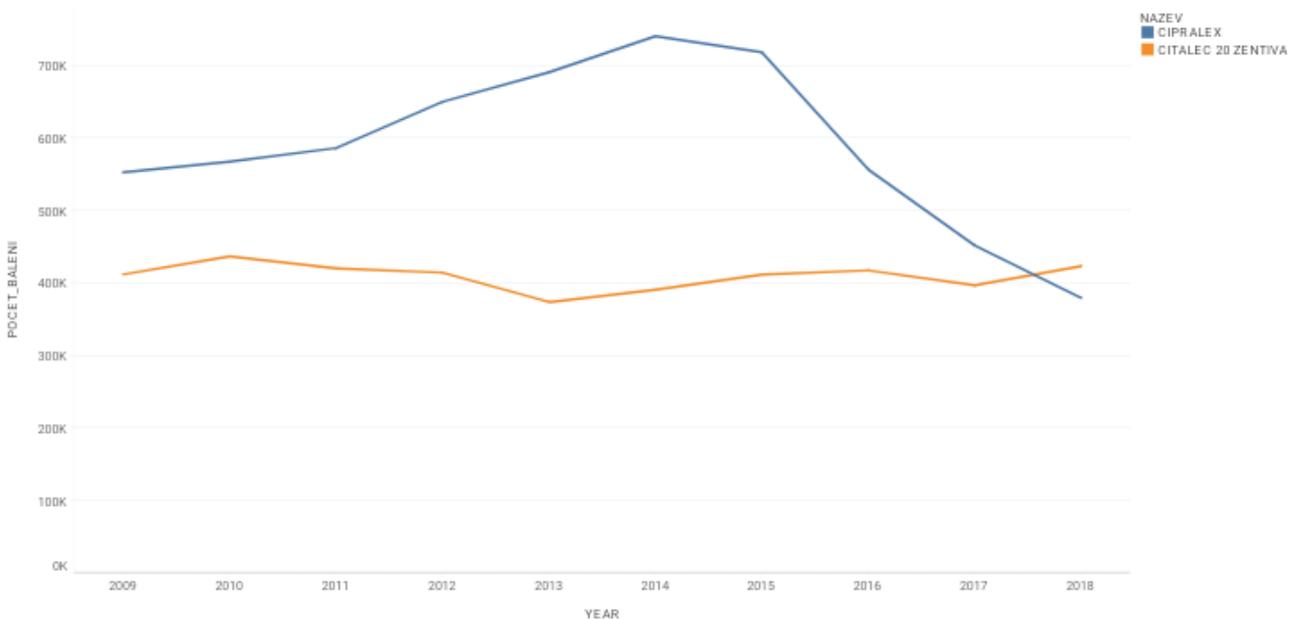

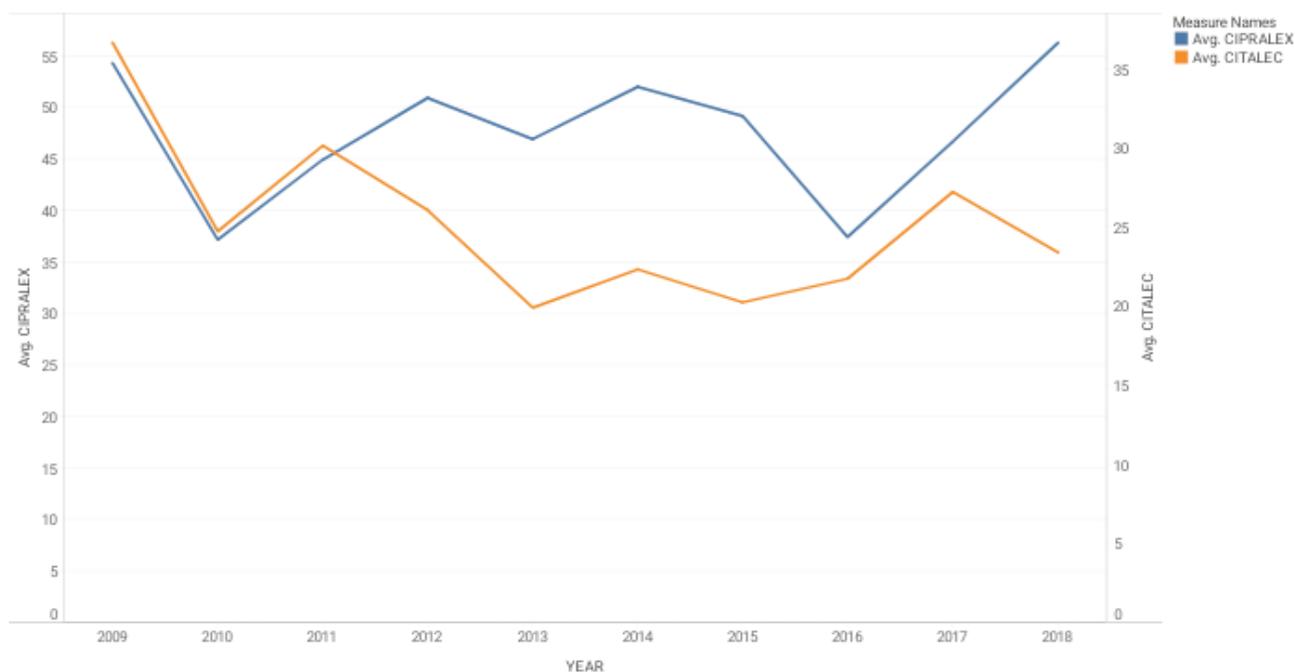*Figure 3* Cipralex and Citalec market comparison (no. of packages sold) 2009-2018.

Figure 4 Cipralex and Citalec Google Trend comparison 2009-2018.

could lead to this possible usage. Firstly, the data must follow a consistent structure with clearly defined variables. Secondly, the datasets have to include the classification codes to uncover specific medicines or active ingredients. We faced significant obstacles with data synthesis during our three-month collection period across the EU member states mainly caused by poor information services with significant differences regarding open human medicine data structure and quality. Finally, we were able to demonstrate open data CI analysis with a focus on the Czech antidepressant market. Moreover, the Czech datasets provided us with the possibility of showing insights into specific medicine and company market performance thanks to rich data quality including the ATC classification, name of the producers, consumption and pricing data, as well as reliable retrospectivity. We have used the ATC classification to filter out the antidepressant class during the time period from 1991 to 2018 for the consumption trend, and then from 2009 to 2018 to provide actual market insights. Afterwards, we were able to identify key antidepressant market players and the main antidepressants used in the Czech Republic, together with consumption data (daily doses, number of packages, etc.) and finally to uncover total antidepressant consumption. Our first research question is therefore confirmed: open data and its contextual analysis bring intelligence for a specific country.

Table 5 Google Search data territorial analysis with the keyword Cipralex 2009-2018.

Czech region	*Cipralex* keyword interest
Pardubický region	100
Region Vysočina	80
Prague	78
Ústecký region	71
Moravskoslezský region	69
Středočeský region	58
Jihomoravský region	58
Zlínský region	55
Olomoucký region	54
Královéhradecký region	50
Jihočeský region	45
Plzeňský region	43

Table 6 Google Search data territorial analysis with the keyword Citalec 2009-2018.

Czech Region	*Citalec* keyword interest
Moravskoslezský region	100
Středočeský region	55
Jihomoravský region	54
Prague	51
Zlínský region	33

Our second research question focused on geographical aspects of open medicine data. Firstly, medicine data are strictly anonymous. None of the institutions provided datasets with geolocation. We have decided to use information-seeking behavior data and show possible geographical context to antidepressant consumption. We have used the Google Search data together with related keywords that consisted of 'antidepressant' as the general term and the specific name of the medicine. The analysis has shown that Google Search data correlates with market trends uncovered by open data analysis, but the territorial insights were not significant in this case due to the small Google Search data sample and provided only a general regional overview. However, this method would be effective in the analysis of a larger western country (e.g. United States, United Kingdom) where it is possible to work with a more significant and detailed search data sample, e.g. on a city by city level. For this reason, we consider the second research question to be confirmed.

Our future work is directed towards finding possible intelligence links between the open human medicine market data and innovation processes with the perspective of using patent data.

7. DISCUSSION

Our aim in this paper was to provide possibilities for working with open data as a tool for hard-to-get intelligence insights within the pharmaceutical sector. Not only did our results provide significant and relevant market context, but they also confirmed that open human medicine data can serve as a trend analysis information commodity for a wide range of public entities, e.g. governmental bodies for decision-making processes aimed at increasing the level of public health. Our case with antidepressants has demonstrated the trend analysis possibility within a reliable time frame. Thanks to the ATC classification system we are able to determine specific health problems within the whole population of a specific country. More importantly, we can compare countries afterwards regarding their health condition.

Our collection phase regarding the data structure and quality led us to the conclusions that open human medicine data initiatives should be considered more seriously across the EU. We have designed optimal data fields as a common base. This could determine another

quality level across the whole of Europe and use open data in the most reliable way: to strengthen public health.

Still, open human medicine data in our designed structure plays an important role for intelligence studies. Based on our results, we could get insightful market information in any selected geographical area on pharmaceutical companies, medicine brands and the active ingredients of drugs together with their therapeutic, chemical and pharmacological properties.

ACKNOWLEDGEMENTS

8. REFERENCES

Akhgar, B., Bayerl, P. S., & Sampson, F. (Eds.). (2016). *Open Source Intelligence Investigation.* Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-47671-1

Bernard, R., Bowsher, G., Milner, C., Boyle, P., Patel, P., & Sullivan, R. (2018). Intelligence and global health: assessing the role of open source and social media intelligence analysis in infectious disease outbreaks. *Journal of Public Health*, *26*(5), 509–514.

Brownstein, J. S., Freifeld, C. C., Reis, B. Y., & Mandl, K. D. (2008). Surveillance Sans Frontières: Internet-Based Emerging Infectious Disease Intelligence and the HealthMap Project. *PLoS Medicine*, *5*(7), e151.

Cantor, M. N., Chandras, R., & Pulgarin, C. (2018). FACETS: using open data to measure community social determinants of health. *Journal of the American Medical Informatics Association : JAMIA*, *25*(4), 419–422.

Cerny, J., Potancok, M., & Molnar, Z. (2018). Open human medicine data in European Union. In Oskrdal V., Doucek P., & Chroust G. (Eds.), *IDIMT-2018: strategic modeling in management, economy and society: 26th Interdisciplinary Information Management Talks, Sept. 5 -7, 2018, Kuntná Hora, Czech*

Republic (p. 509). Kutná Hora: University of Economics, Prague, Czech Republic.

Cook, S., Conrad, C., Fowlkes, A. L., & Mohebbi, M. H. (2011). Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic. *PLoS ONE*, *6*(8), e23610.

Datig, I., & Whiting, P. (2018). Telling your library story: tableau public for data visualization. *Library Hi Tech News*, *35*(4), 6–8.

European Commison. (2018). Building a European data economy.

European Commison. (2019). About | Open Data Portal. Retrieved from https://data.europa.eu/euodp/en/about

Farber, G. K. (2017). Can data repositories help find effective treatments for complex diseases? *Progress in Neurobiology*, *152*, 200–212.

Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press. Retrieved from https://books.google.cz/books?id=1xiHLgEAC AAJ

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144.

Grèzes, V. (2015). The definition of competitive intelligence needs through a synthesis model. *Journal of Intelligence Studies in Business*, 5(1), 40–56.

Herring, J. P. (1999). Key intelligence topics: A process to identify and define intelligence needs. *Competitive Intelligence Review*, *10*(2), 4–14.

Hu, J., Perer, A., & Wang, F. (2016). Data Driven Analytics for Personalized Healthcare BT - Healthcare Information Management Systems: Cases, Strategies, and Solutions. In C. A. Weaver, M. J. Ball, G. R. Kim, & J. M. Kiel (Eds.), *Healthcare Information Management Systems* (pp. 529–554). Cham: Springer International Publishing.

Hughes, S. F. (2017). A new model for identifying emerging technologies. *Journal of Intelligence Studies in Business*, 7(1), 79–86.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268.

Kostkova, P., Brewer, H., de Lusignan, S., Fottrell, E., Goldacre, B., Hart, G., … Tooke, J. (2016). Who Owns the Data? Open Data for Healthcare. *Frontiers in Public Health*, *4*, 7.

Nuti, S. V, Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The Use of Google Trends in Health Care Research: A Systematic Review. *PLOS ONE*, *9*(10), e109583.

Oubrich, M. (2011). Competitive intelligence and knowledge creation - outward insights from an empirical survey. *Journal of Intelligence Studies in Business*, 1(1), 97–106.

Perer, A., & Gotz, D. (2013). Data-driven Exploration of Care Plans for Patients. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 439–444). New York, NY, USA: ACM.

Tableau. (2019). Tableau: Business Intelligence and Analytics Software. Retrieved from https://www.tableau.com

U.S. National Library of Medicine. (2019). MeSH Browser. Retrieved from https://meshb.nlm.nih.gov/search

WHO. (2018). ATC/DDD Index 2018. Retrieved May 29, 2018, from https://www.whocc.no/atc_ddd_index/