MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220176

Exploiting Social Media for Active Pharmacovigilance: The PVClinical Social Media Workspace

Christine Kakalou[a], Stella Dimitsaki[a], Vlasios K. Dimitriadis[a], Pantelis Natsiavas[a*]

[a] Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece

Abstract

The value of social media data for Adverse Drug Reaction (ADR) monitoring is actively investigated. While social media provide a vast amount of data, these data are hard to analyse due to their unstructured nature and lack of credibility. Despite these challenges, social media have been identified as a potentially useful data source, potentially able to "strengthen" the evidence for new ADRs. To this end, PVClinical project aims to build a platform facilitating the investigation of multiple heterogeneous data sources, including social media, to support pharmacovigilance (PV) processes, both in the clinical environment and beyond. In this study, we present the PVClinical Twitter workspace, also highlighting the rationale behind the main design choices, while also discussing the respective challenges.

Keywords:
Pharmacovigilance, Adverse Drug Reactions, Drug Safety

Introduction

Pharmacovigilance (PV) is defined as "the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem"[1] aiming to mitigate the impact of Adverse Drug Reactions (ADRs). As ADRs cause a significant clinical and economic burden [2], PV is a highly regulated domain [3] and also a domain of active research in terms of exploiting data and "intelligent" computational approaches [4][5][6][7]. To this end, Information and Communication Technologies (ICT) are vital for the collection, assessment and analysis of the respective data.

Typically, potential new or undocumented ADRs, called "*ADR signals*", are captured via Spontaneous Reporting Systems (SRS). SRSs are used to collect Individual Case Safety Reports (ICSRs), submitted by healthcare professionals or patients [8]. Beyond SRSs, new "emerging" data sources are investigated to expand the PV related data space. To this end, social media have been identified as one of the data sources which could (at least in principle) provide unsolicited, real time, user-generated information about suspected medication side-effects.

While obviously social media come with inherent limitations (e.g. data biases, credibility etc.), they could play a complementary role acting either as an early warning system for potentially new ADR signals ("proto-signals") [9][10] or as a validation source for candidate signals from other PV sources [11], or even providing a "profiling" of ADRs in terms of their relation with patients potential lifestyle. Based on the above, regulatory organizations have produced guidelines on how unsolicited reports of ADRs collected from social media such as Twitter should be handled for PV purposes [12].

Notably, the WEB-RADR and WEB-RADR 2 projects[1] explored the use of social media for drug safety monitoring scenarios and have published recommendations for the integration of social media for PV purposes [13] and guidance regarding related data collection good practices [14]. Furthermore, in the context of these projects Adverse Events (AE) recognition tasks were elaborated [15] and an evaluation/benchmarking dataset was published to facilitate the evaluation and the comparison of relevant algorithms and ICT tools [16].

The PVClinical project[2] aspires to build an ICT platform, i.e. a web application, aiming to support PV processes in clinical environment and beyond, exploiting Knowledge Engineering (KE) as its main technical paradigm [17]. Based on the identified User Goals, the PVClinical platform integrates various heterogeneous sources of information (FDA - FAERS, PubMed, Twitter and Electronic Health Record data) [18].

In this paper, we present PVClinical's Twitter workspace collecting, storing and facilitating the analysis of user-generated data to support the investigation of specific ADRs. We emphasize on the methodological choices during the design of the specific workspace, while also depicting the challenges faced during the integration of such a data source in a PV tool.

Methods

According to the main PVClinical platform's rationale, the end-user defines an "investigation" scenario, namely the Drug-Event Combination (DEC) corresponding to the specific ADR under investigation. Well defined and widely accepted terminologies available in RDF format, i.e. ATC for drugs and MedDRA for adverse reactions are used to define the DEC concepts, and the terminological search space is "expanded" via potential synonyms.

The Twitter workspace per se, collects data using a home developed software module [19], already utilized to analyze data for other public health scenarios [20]. Upon the collected tweets, Natural Language Processing (NLP) and interactive visualization techniques are employed. The collected tweets, hashtags and key influencers are presented in a longitudinal fashion, using analytics based on RShiny framework[3]

In order to effectively address the challenges of using social media for PV monitoring, the outcomes of the IMI WEB-

[1] https://web-radr.eu/web-radr2/
[2] https://pvclinical-project.eu/
[3] https://shiny.rstudio.com/

RADR[4] project have been used as guidelines [13]. More specifically, the recommendations employed regarding the use of social media use as part of ADR signal investigation processes can be summarized as following:

R1. Social media should not be used as a source of ICSRs.

R2. Facebook and Twitter are currently not worthwhile to employ for the purpose of broad-ranging statistical signal detection at the expense of other PV activities.

R3. Future research should explore the value of social media as a source of information for additional cases in signal refinement/evaluation of ADRs that may significantly affect a patient's quality of life.

R4. If social media is considered for use in PV, it is recommended that a prior assessment of the absolute and relative number of available posts related to the drug and/or event of interest in different online sources is made in relation to its intended use.

R5. Further research should be carried out to determine whether there is value in social media data for niche areas of PV.

R6. Consider using a predictive algorithm to identify and eliminate from the search query any medicinal product names with high levels of ambiguity to optimize time efficiency and, where applicable, cost effectiveness.

Results

In order to demonstrate the respective implementation, we present snapshots of data collection regarding a vaccine which has recently gained publicity due to COVID-19 pandemic and the discovery of very rare incidences of unusual blood clots [21].

Real-time analytics can be presented in an ad-hoc function once the collection process has started and the user can stop the collection process at any time. The main overview of the collected tweets is presented in an aggregated timeline along with relevant metrics such as the total number of tweets, the number of retweets, the duration of the collection process and the daily rate of tweets collected, through an interactive timeline visualization (see Figure 1). Moreover, the user has the option of exploring the actual raw data in a "drill-down" fashion and searching among the entire dataset for specific terms using regular expressions (Figure 2).

Moreover, hashtags have been identified as a communication token emerging as a "special" communication artifact heavily used in the social media context, worth being analyzed independently. More specifically, in Twitter, hashtags are widely used as an indicator of a specific discussion topic and can be used to track discussion trends over time. To this end, the tracking of popular hashtags could enhance the understanding of the trending topics related with a specific ADR; Using an interactive aggregated streamgraph, the user can inspect and separately highlight the top hashtags in real time as seen in Figure 3.

Finally, in order to identify any accounts that post heavily about a specific DEC, the user can consult a Treemap visualization of the most active accounts for each scenario's collected tweets.

Discussion

The potential benefits and the challenges of the use of social media for drug safety purposes are actively investigated. Their value is mostly related to the vast amount of available data and their time-evolving nature enabling the identification of potential longitudinal trends. However, still there are not yet convincing results that these data could be used to identify ADR signals. Therefore, social media are considered a data source which could act in a complementary fashion along other data sources, e.g. SRSs.

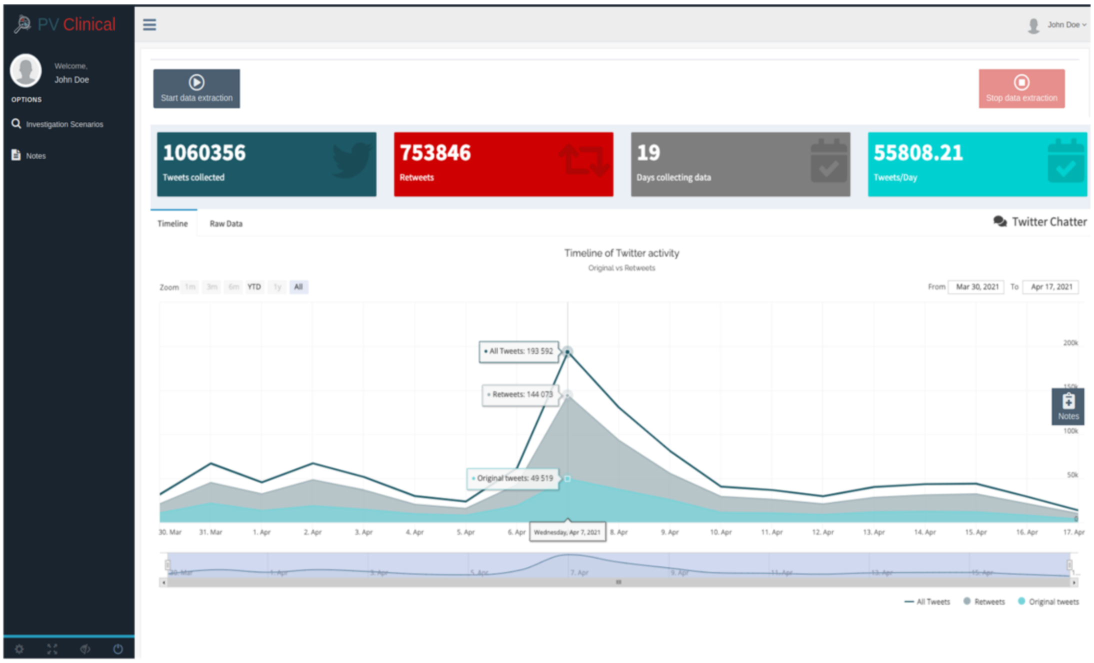

Figure 1 – Interactive timeline visualisation of collected Twitter data.

[4] https://web-radr.eu/web-radr2/

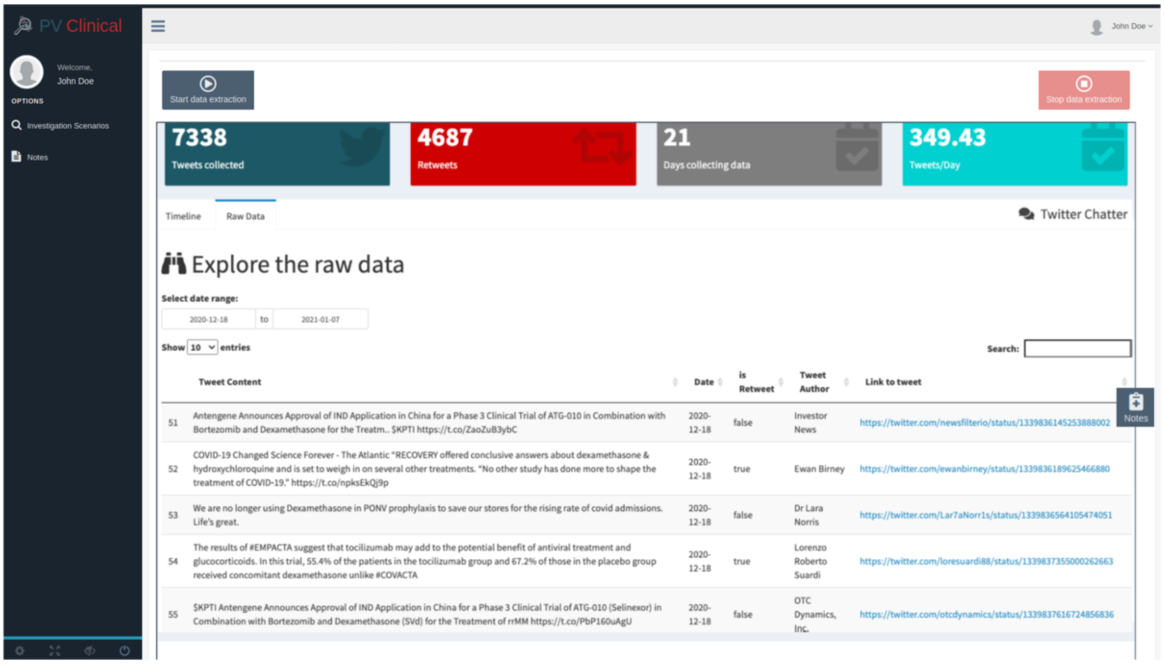

Figure 2 – Raw data exploration and advanced search using regular expressions within the Twitter workspace.

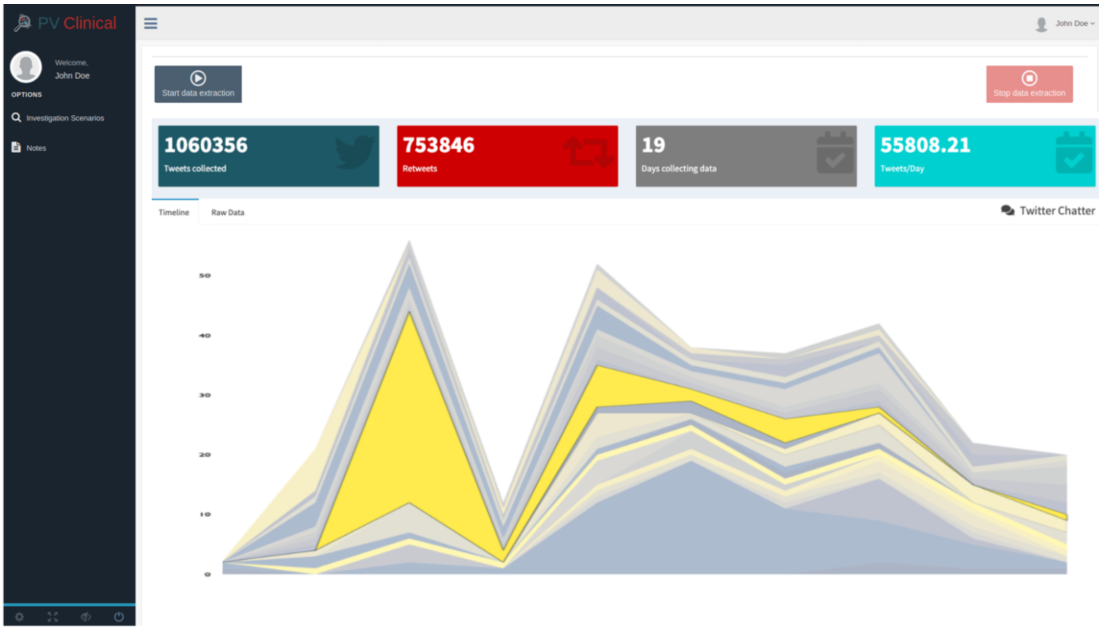

Figure 3 – Hashtags Streamgraph in the Twitter Workspace.

WEB-RADR recommendations for the use of social media in PV, guided the development of the presented module of the PVClinical platform workspace. Analyzing the compliance of the presented module with the respective recommendations, the following remarks could be made:

R1. While ICSRs are actively used in the PVClinical platform as part of another workspace (out of this paper's scope), tweets are clearly separated and managed in a different manner, using different statistic metrics and analytical tools, clearly highlighting that they are not and should not be considered of the same value like ICSRs.

R2. Indeed, no disproportionality analysis metrics are used, and no signal detection algorithms are applied on the Twitter data. The collected data are presented complimentary to other data (e.g. SRS data) which are part of other workspaces (out of the current paper's scope) and therefore they should be interpreted accordingly.

R3, R5. The presented workspace aims to facilitate the exploration of the specific data source in the context of PV usage scenarios, via descriptive analytics. Moreover, beyond the quantitative statistics, there is also the ability to drill down on each tweet and explore its special characteristics (time, date, if it is a retweet or not, user etc.). This process could support the qualitative evaluation of the collected data, beyond signal detection.

R4. As part of the PVClinical platform, other online sources (i.e. SRS data provided via OpenFDA platform) are used to provide insights regarding the DEC under investigation.

R6. Using "predictive" algorithms on social media data was considered out of the presented tool scope. Given that the presented tool is going to be pilot tested in both clinical and non-clinical settings, the investigation of such "predictive" algorithms are actually part of the "research" line of the project and not the presented tool.

In terms of future work, "intelligent" algorithms are actively investigated not only in terms of predictive algorithms but also in terms of "fake news" prevention. Furthermore, a pilot testing of the application is planned, also focusing on usability aspects.

Conclusion

The use of social media for PV purposes comes with many challenges but also with wide potential. PVClinical platform provides a gateway to integrate Twitter data along other data sources following a recommendations-based approach, aiming to support PV activities in both clinical and non-clinical context.

Acknowledgements

This research has been cofinanced by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE - INNOVATE (project code: T1EDK-03789).

The authors would also like to highlight the major contribution of the recently deceased Dr. Vassilis Koutkias, Coordinator/Principal Investigator of PVClinical project.

References

[1] W.C.C. for I.D.M. World Health Organization, The importance of pharmacovigilance., World Health Organization, 2002. http://apps.who.int/medicinedocs/en/d/Js4893e/ (accessed December 12, 2018).

[2] J. Sultana, P. Cutroneo, and G. Trifirò, Clinical and economic burden of adverse drug reactions., *J. Pharmacol. Pharmacother.* **4** (2013) S73-7. doi:10.4103/0976-500X.120957.

[3] J.J. Borg, A. Tanti, D. Kouvelas, C. Lungu, M. Pirozynski, A. Serracino-Inglott, and G. Aislaitner, European Union pharmacovigilance capabilities: Potential for the new legislation, *Ther. Adv. Drug Saf.* **6** (2015) 120–140. doi:10.1177/2042098615591802.

[4] S.N. Pal, C. Duncombe, D. Falzon, and S. Olsson, WHO strategy for collecting safety data in public health programmes: complementing spontaneous reporting systems., *Drug Saf.* **36** (2013) 75–81. doi:10.1007/s40264-012-0014-6.

[5] A. Bate, and S.F. Hobbiger, Artificial Intelligence, Real-World Automation and the Safety of Medicines, *Drug Saf.* (2020) 1–8. doi:10.1007/s40264-020-01001-7.

[6] P. Natsiavas, A. Malousi, C. Bousquet, M.-C. Jaulent, and V. Koutkias, Computational Advances in Drug Safety: Systematic and Mapping Review of Knowledge Engineering Based Approaches, *Front. Pharmacol.* **10** (2019) 415. doi:10.3389/fphar.2019.00415.

[7] A. Bate, R.F. Reynolds, and P. Caubel, The hope, hype and reality of Big Data for pharmacovigilance., *Ther. Adv. Drug Saf.* **9** (2018) 5–11. doi:10.1177/2042098617736422.

[8] P. Inácio, A. Cavaco, and M. Airaksinen, The value of patient reporting to the pharmacovigilance system: a systematic review., *Br. J. Clin. Pharmacol.* **83** (2017) 227–246. doi:10.1111/bcp.13098.

[9] I. Convertino, S. Ferraro, C. Blandizzi, and M. Tuccori, The usefulness of listening social media for pharmacovigilance purposes: a systematic review, *Expert Opin. Drug Saf.* **17** (2018) 1081–1093. doi:10.1080/14740338.2018.1531847.

[10] O. Caster, J. Dietrich, M.-L. Kürzinger, M. Lerch, S. Maskell, G.N. Norén, S. Tcherny-Lessenot, B. Vroman, A. Wisniewski, and J. van Stekelenborg, Assessment of the Utility of Social Media for Broad-Ranging Statistical Signal Detection in Pharmacovigilance: Results from the WEB-RADR Project, *Drug Saf.* (2018) 1–15. doi:10.1007/s40264-018-0699-2.

[11] A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, T. Upadhaya, and G. Gonzalez, Utilizing social media data for pharmacovigilance: A review, *J. Biomed. Inform.* **54** (2015) 202–212. doi:10.1016/j.jbi.2015.02.004.

[12] E. Medicines Agency, Guideline on good pharmacovigilance practices (GVP) - Module VI – Collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2), 2017.

[13] J. van Stekelenborg, J. Ellenius, S. Maskell, T. Bergvall, O. Caster, N. Dasgupta, J. Dietrich, S. Gama, D. Lewis, V. Newbould, S. Brosch, C.E. Pierce, G. Powell, A. Ptaszyńska-Neophytou, A.F.Z. Wiśniewski, P. Tregunno, G.N. Norén, and M. Pirmohamed, Recommendations for the Use of Social Media in Pharmacovigilance: Lessons from IMI WEB-RADR, *Drug Saf.* **42** (2019) 1393–1407. doi:10.1007/s40264-019-00858-7.

[14] S. Brosch, A.-M. de Ferran, V. Newbould, D. Farkas, M. Lengsavath, and P. Tregunno, Establishing a Framework for the Use of Social Media in Pharmacovigilance in Europe, *Drug Saf.* (2019) 1–10. doi:10.1007/s40264-019-00811-8.

[15] L.M. Gattepaille, S. Hedfors Vidlin, T. Bergvall, C.E. Pierce, and J. Ellenius, Prospective Evaluation of Adverse Event Recognition Systems in Twitter: Results from the Web-RADR Project, *Drug Saf.* (2020). doi:10.1007/s40264-020-00942-3.

[16] J. Dietrich, L.M. Gattepaille, B.A. Grum, L. Jiri, M. Lerch, D. Sartori, and A. Wisniewski, Adverse Events in Twitter-Development of a Benchmark Reference Dataset: Results from IMI WEB-RADR, *Drug Saf.* **43** (2020) 467–478. doi:10.1007/s40264-020-00912-9.

[17] P. Natsiavas, G. Gavriilidis, Z. Linardaki, G. Kolangi, E. Gkaliagkousi, C. Zamboulis, and M.-C. Jaulent, Supporting Active Pharmacovigilance via IT Tools in

the Clinical Setting and Beyond: Regulatory and Management Aspects, in: Stud. Health Technol. Inform., IOS Press, 2020. doi:10.3233/SHTI200565.

[18] P. Natsiavas, M.-C. Jaulent, and V. Koutkias, A Knowledge-Based Platform for Assessing Potential Adverse Drug Reactions at the Point of Care: User Requirements and Design., in: Stud. Health Technol. Inform., 2019: pp. 1007–1011. doi:10.3233/SHTI190376.

[19] P. Natsiavas, N. Maglaveras, and V. Koutkias, A Public Health Surveillance Platform Exploiting Free-Text Sources via Natural Language Processing and Linked Data: Application in Adverse Drug Reaction Signal Detection Using PubMed and Twitter, in: D. Riaño, R. Lenz, and M. Reichert (Eds.), Knowl. Represent. Heal. Care HEC 2016 Int. Jt. Work. KR4HC/ProHealth 2016, Munich, Ger. Sept. 2, 2016, Revis. Sel. Pap., Springer, Cham, Cham, 2017: pp. 51–67. doi:10.1007/978-3-319-55014-5_4.

[20] C. Kakalou, J.V. Lazarus, and V. Koutkias, Mining social media for perceptions and trends on HIV pre-exposure prophylaxis, 2019. doi:10.3233/SHTI190366.

[21] AstraZeneca’s COVID-19 vaccine: EMA finds possible link to very rare cases of unusual blood clots with low blood platelets | European Medicines Agency, (n.d.).

Address for correspondence

Corresponding Author: Pantelis Natsiavas, Institute of Applied Bio-sciences, Centre for Research and Technology, Thermi, Thessaloniki, Greece; E-mail: pnatsiavas@certh.gr