The EU-ADR Web Platform: Delivering advanced pharmacovigilance tools

Oliveira JL[1], Lopes P[1], Nunes T[1], Campos D[1], Boyer S[2], Ahlberg Helgee E[2], Van Mullingen EM[3], Kors JA[3], Singh B[3], Furlong LI[4], Sanz F[4], Bauer-Mehren A[4], Carrascosa MC[4], Mestres J[4], Avillach P[5,6], Diallo G[6], Diaz C[7], Van der Lei J[3]

[1] DETI/IEETA, University of Aveiro, Portugal
[2] AstraZeneca, Molndal, Sweden
[3] Erasmus University Medical Center, Rotterdam, The Netherlands
[4] Research Programme on Biomedical Informatics (GRIB), IMIM Hospital del Mar Research Institute and Universitat Pompeu Fabra, Barcelona, Catalonia, Spain
[5] LERTIM, EA 3283, Faculté de Médecine, Université de Aix-Marseille, France
[6] LESIM-ISPED, Université de Bordeaux, France
[7] Synapse Research Management Partners, Barcelona, Spain

Corresponding Author

José Luis Oliveira
DETI/IEETA, University of Aveiro, 3810-193 Aveiro, Portugal
Tel: +351 234 370 500, fax: +351 234 370 545
Email: jlo@ua.pt

Keywords

Pharmacovigilance, ADR, adverse drug reactions, drug safety, *in silico* pharmacology.

Key Points

- Progress in pharmacovigilance demands new methods to further improve data exploration from traditional spontaneous reporting systems. Advanced tools are in

place to mine data from general practitioners research databases, establishing useful connections to other well-known resources.

- Web services for the analysis of drug-event associations were developed, requiring the implementation of service composition strategies to foster interoperability within the pharmacovigilance software ecosystem.
- A unique web-based workspace, the EU-ADR Web Platform, is introduced to deliver advanced pharmacovigilance software to everyone, empowering the research community with pioneering tools to identify, monitor and evaluate adverse drug reactions.

Acknowledgments

We wish to thank to all the members of the EU-ADR project. This work was supported by the European Commission (EU-ADR, ICT-215847), FCT (PTDC/EIA-CCO/100541/2008), and Instituto de Salud Carlos III FEDER (CP10/00524). The Research Programme on Biomedical Informatics (GRIB) is a node of the Spanish National Institute of Bioinformatics (INB).

Conflict of Interest

The authors have declared that no competing interests exist.

Summary

Purpose

Pharmacovigilance methods have advanced greatly during the last decades, making post-market drug assessment an essential drug evaluation component. This strategy uses spontaneous reporting systems and health information databases to collect expertise from huge amounts of real-world reports. The EU-ADR Web Platform was built to further facilitate accessing, monitoring and exploring these data, enabling an in-depth analysis of adverse drug reactions risks.

Methods

The EU-ADR Web Platform exploits the wealth of data collected within the EU-ADR project. Millions of electronic health records are mined for specific drug events, which are correlated with literature, protein and pathway data, resulting in a rich drug-event dataset. Next, service composition strategies are tailored to coordinate the execution of distributed web services performing data-mining and statistical analysis tasks. This permits obtaining a ranked drug-event list, removing spurious entries and highlighting relationships with high risk potential.

Results

The EU-ADR Web Platform is an open workspace for the integrated analysis of pharmacovigilance datasets. Using this software, researchers can access a variety of tools provided by distinct partners in a single centralized environment. Besides performing standalone drug-event assessments, they can also control the pipeline for an improved batch analysis of custom datasets. Drug-event pairs can be filtered, substantiated and statistically analyzed within the platform's innovative working environment.

Conclusions

A pioneering workspace for delivering advanced drug studies has been developed within the EU-ADR project consortium. This tool, targeted at the pharmacovigilance community, is available online at https://bioinformatics.ua.pt/euadr/.

Introduction

Contemporary prevention and treatment of diseases revolves around a dynamic medication market where pharmaceutical companies compete, aiming to investigate, develop and introduce new drugs in daily healthcare provision. Despite the expected therapeutic benefit of these innovations, drug safety is a major concern for worldwide policy stakeholders as several marketed drugs continue to pose serious risks to the wellbeing of many patients, having become in recent years one of the leading causes of mortality[1].

The traditional approach tackles this problem from a pre-market perspective, conditioning drug approval. Both the European Medicines Agency (EMA)[2] and the US Food and Drug Administration (FDA)[3] establish rigorous guidelines for new medicine approval, requiring intense testing and trials, which result in a long and complex lab-to-market development cycle[4]. Along with these guidelines, pharmaceutical companies must also define thorough risk management plans for post-market drug stages[5,6].

Consequently, the relevance of post-market pharmacovigilance in the health domain has been growing steadily over the last four decades[7,8]. Research in this area involves the exploration and assessment of signals, defined by the World Health Organization as undisclosed assertions on direct relationships between adverse events and a drug[9]. Clinicians use spontaneous reporting systems (SRS) to identify adverse drug reactions. These systems empower physicians with tools to report suspicions on certain drugs to a pharmacovigilance center. Latest advances take these tools even further, completing the drug loop by providing a reporting infrastructure to pharmacists and patients. Although many ADRs were detected through these systems, there are inherent limitations that hamper signal detection[10,11]. They depend entirely on the ability of a physician to recognize an adverse event as being related to the drug, and on his availability to report the case to the local spontaneous reporting database. The greatest limitations, therefore, are under-reporting and biases due to selective reporting. Investigations have shown that the percentage of ADRs being reported varies between 1 and 10%[12-14].

Consequently, there is a high-demand for novel software tools capable of improving the post-marketing drug monitoring workflow[15]. By taking advantage of modern knowledge engineering technologies we are able to overcome the limitations associated with insufficient clinical trial data, complex monitoring statistics and closed general practice data silos. Text

and data-mining tools, combined with service composition strategies, pave the way for enhanced *in silico* signal identification and adverse drug reaction assessment[16].

ADR reporting and analysis

Hårmark and Grootheest research explains the underlying pharmacovigilance concerns with current drug evaluation approaches[17]. Whilst drug safety concerns are becoming more prominent, the lack of adequate software to correctly understand drug adverse reactions continues to challenge the pharmaceutical industry and research community[18,19].

The risk associated with any marketed drug triggers critical safety concerns, which, in their turn, leverage a constant revision and update of medical products' information. For these tasks, modern adverse drug reaction (ADR) monitoring becomes essential. Despite the complex set of drug trials, including the final randomized double blind evaluation, clinical trials data is in most scenarios insufficient to assess drug risk. Rare ADRs, ADRs identified in particular population cohorts or ADRs with long latency, require intensive post-marketing drug analysis.

At this level, spontaneous drug reporting systems (SRS) come to play. Pharmacovigilance centres task is to collect these reports, generating enough data to inform stakeholders of potential risks as soon as they appear in the system. Despite the invaluable data coming from SRS, their data alone is meaningless in most scenarios. Viewing SRS as independent entities makes it nigh impossible to establish direct relationships between the causes (a drug, or drug interaction) and consequences (a phenotype). Hence, to extract meaningful insights from these SRS records, we need to rely on advanced data mining techniques[20]. These provide distinct perspectives over acquired data and their connections to other information topics[21].

Another strategy is in place to complement spontaneous reporting systems. Intensive monitoring systems rely on prescription data, forcing drug prescribers to ask about any adverse reaction during the drug intake cycle. Once these data are collected, they are processed for signal evaluation. Unlike SRS, which is based on monitoring specific drugs over a controlled time period, intensive reporting relies on a non-interventional observational cohort. Hence, generated data is much nearer real-world scenarios than data obtained through SRS. Intensive reporting also renewed the interest in the importance of health information systems and general practice research databases.

The EU-ADR initiative

Despite the myriad of international developments in these fronts, most efforts approach this problem from a pre-market approach, focusing on conditioning drug approval and defining guidelines for risk management plans. Hence, modern projects such as EU-ADR[22] or RADAR[23], define a proactive strategy for post-marketing drug assessment. To overcome the 'reporting bias' and underreporting of physicians, the EU-ADR solution was based on automatically exploiting the data stored in large Electronic Health Record (EHR) databases. Modern regional and national health information systems tend to store miscellaneous information regarding patients' clinical history, including drug prescriptions, vaccinations, height, weight or laboratory test results, among others[24]. These wide collections of data are traditionally a good general representation of region demographics. Furthermore, collected data is already used for pharmacoepidemiology (56%), disease epidemiology (30%) and, to a lesser extent, pharmacoeconomics including drug usage monitoring[25-27]. From a pharmacovigilance perspective and in a European or worldwide scale, mining the amount and type of data collected in these databases is of tremendous importance for an improved post-marketing drug evaluation[28].

The foundation for this strategy is doing in-depth semantic data mining on the wealth of electronic health records to generate filtered data that can be easily substantiated through distributed computational tools[16,29]. The final output, a ranked signal list, provides a broad look over identified signals and their significance in health risk.

The EU-ADR Web Platform tackles these challenges, extending the availability of existing tools to every stakeholder, through a web-based pharmacovigilance suite. This system enables an insightful exploration of pharmacovigilance signals' evolution resulting in a comprehensive risk evaluation. This is possible through innovative features such as the creation of custom drug studies, the remote execution of signal substantiation workflows or the cross-analysis against millions of anonymous electronic health records[30]. The platform is publicly available online at http://bioinformatics.ua.pt/euadr.

Methods

The EU-ADR project exploits data from electronic healthcare records (EHR) and health information systems (HIS) of about 30 million European patients, channelling it through edge-of-breed distributed software and enhancing signal detection[31]. This large-scale drug safety monitoring relies in various mining, epidemiological, statistical and computing techniques to assess acquired data and generate a ranked signal list – Figure 1.

(Figure 1)

With EU-ADR's huge knowledge base in place, innovative methods to access and explore collected data are required, since many drug adverse reactions can be biologically explained if we are able to integrate current biomedical knowledge. We call this process signal substantiation[16]. This signal substantiation is performed through several distributed Taverna workflows[32,33].

The **Medline ADR** signal filtering workflow automates literature analysis tasks by assessing a list of publications regarding a specific signal. The algorithm adopts a semantics-based approach that processes Medline annotations looking for particular MeSH terms[34]. This workflow's output is a direct relationship between an adverse drug reaction and its descriptions in Medline, if present.

Second, the signal filtering co-occurrence process is divided into three similar workflows, each targeting a distinct resource. These evaluate the relationships between drugs and side effects that might have been reported previously in Medline literature (**Medline co-occurrence**) or drug databases such as DailyMed[35] (**DailyMed**) or DrugBank[36] (**DrugBank**). These workflows use statistical and text-mining techniques to evaluate drug names, ATC codes and event co-occurrences in the indexed resources.

At last, the **Signal Substantiation** workflow analyses the drugs, proteins and pathways interaction graphs. This involves searching for proteins targeted by the drug and associated with the clinical event, directly or through biological pathways. The algorithm generates drug-target and event-protein profiles that are searched for common sets of proteins, the intersecting portion of the graph.

These five workflows accept a similar input, a drug-event pair, and produce a similar output, standardized XML. Workflow interactions are made possible by EU-ADR's XML

schema language[a]. The data flow from and to workflows is exchanged in XML described using a EU-ADR internal schema. This is a true interoperability enabler as data is shared in a format that is understood by all tools in the EU-ADR ecosystem.

EU-ADR workflows can be used independently in Taverna workbench[32]. However, they are fit for programmatic usage but neglect the general pharmacovigilance community with low computer expertise. Also, combining these workflows' results is essential for a better understanding of drug-event relationships. As such, to foster an easier usage and promote the aggregation of results, a centralized workflow execution tool is needed. This complements local workflow usage for individual analysis with remote workflow execution for processing large heterogeneous datasets. Moreover, when executed online in the EU-ADR Web Platform, workflow results are presented in a specialized interface, designed to highlight their relevant parts and facilitate evidence analysis. This unique interface contrasts with the raw text and XML data obtained directly from Taverna.

Evidence combination is a central part of signal substantiation. Whilst each workflow' result has value on its own, through combination of different results we can leverage knowledge from multiple sources and better assess the plausibility of a given drug-event relationship. Each EU-ADR workflow yields a binary score, representative of evidence for a given drug-event relationship being found or not. Then, using Dempster–Shafer theory[37] (DST), we combine evidence from several disparate sources and arrive at a degree of belief that takes into account all the available evidence – Figure 2.

(Figure 2)

In a heterogeneous ecosystem offering different means to evaluate signal plausibility, it is important to weigh the trustworthiness of one method against another. Hence, for greater flexibility in signal detection, we must customize the reliability of individual substantiation methods, both nominal workflow' scores and numerical values obtained from statistical analysis of EHR data. Since confidence in any given substantiation method is highly subjective, users should be able to tailor the evidence combination process for their needs, and save their settings privately on the system.

[a] http://bioinformatics.ua.pt/euadr/euadr_types.xsd

The EU-ADR Web Platform also tackles the data sharing and research reproducibility issues[38]. By storing data and workflows online, the EU-ADR Web Platform enables replicating research strategies to follow previous procedures, to confirm previous results or to test if there are novel substantiation outcomes. As the same data and services are used, researchers are assured that their results are unique and longstanding. We can create collaborative groups (Projects) that unlock read and write access to the same data environment. Additionally, existing datasets can be shared to any number of users.

In order to build a complex system and maintain focus on the core features that make it unique, implementation of commonly required functionality and boilerplate code should be delegated to third-party frameworks and libraries. The EU-ADR platform is a web-based collaborative workspace built over a solid foundation of open-source software components. Users interact with the platform client, a highly responsive Google Web Toolkit (GWT) application that runs inside their browser. Client-side components are downloaded only when needed, to allow for faster application loading and conservative bandwidth usage. Communication with the server is made using the command pattern through secure HTTP remote procedure calls (RPC).

Since web-based distributed systems are affected by connection quality and inherently prone to availability issues, the system client depends on remote resources only for data submission, data loading and signal substantiation. This means once a dataset and related evidence is loaded, connectivity loss doesn't hamper system usage. Moreover, all unexpected errors are reported and logged to the server whenever possible, effectively leading to continuous improvement of the system over time.

Results

Setup

The EU-ADR Web Platform is sustained by a distributed computerized system combining multiple components in a single software ecosystem. Figure 3 highlights the data flow from the user submissions to the multiple workflow interactions.

(Figure 3)

EU-ADR workflows play an active role in the EU-ADR Web Platform, as they are required for data analysis and signal evaluation. The challenging tasks of accessing and executing workflows required the development of a new workflow execution engine, enabling real-time web-based communication with the workflows.

Since Taverna is in charge of workflow execution, we need to feed the services with input data, manipulate intermediate results and extract the resulting output documents. The final data is then parsed by the Web Platform and presented to users on the client-side in a way that facilitates evidence analysis and assessment. A thin wrapper was developed in Java, launching parameterized calls to the Taverna command-line tool, which runs in its own process, controlled by system calls.

Workflow execution is a non-blocking asynchronous process. From a usability perspective, this results in a more interactive experience as the workspace can still be used during background workflow execution. Furthermore, EU-ADR's workflows involve services that are not physically or logically co-located, leveraging a truly distributed service execution.

The client application uses a myriad of advanced user interaction components to provide a unique perspective on the huge drug datasets and easy access to data exploration features. Investigation of any drug-event pair does not end after the primary relative risk assessment, as evidence can be combined to reach a final score, helping the separation between spurious signals and potential adverse drug reactions.

Feature Highlights

The EU-ADR Web Platform is built to support advanced pharmacovigilance studies. The invite-based registration system allows authorized researchers to join the Web Platform by giving them access to a personal closed workspace. Registered users are able to upload and analyse drug-event datasets, create targeted drug studies, collaborate with their research peers through the available sharing features and execute EU-ADR workflows locally or remotely.

EU-ADR Web Platform features are available in an online user portal, divided in **Datasets** and **Workflow** views. These sections provide an entry point for exploring drug-event data and accessing project workflows respectively.

The **Dataset** list view, shown in Figure 4, enables managing each user's datasets. Datasets are divided in two sections, **My Datasets**, listing the user personal datasets, and **Shared by others**, listing datasets shared with the user. Both sections include a dataset management action box, allowing the upload of new standardized datasets or the creation of drug-specific datasets, among others. Members of the EU-ADR project have access to an additional section, the **EU-ADR Project** collaborative workspace. This secure workspace facilitates cooperative study of EU-ADR datasets amid project members and assures the confidentiality of all project-private data.

(Figure 4)

Drug-event datasets can be imported to the system from plain-text files in CSV or TXT format or Microsoft Excel spreadsheets in XLS or XLSX format. Each imported file can include up to 5000 drug-event pairs in a standardized format, where the mandatory fields are the drug ATC code and the EU-ADR event acronym – Table 1. Apart from an optional "Name" field, treated as the drug name, each signal can contain any number of additional attributes, which are imported to the platform database and can later be visualized in the dataset view.

(Table 1)

Targeted datasets are focused on a single drug, statistically analysed against the 30 million anonymous EU-ADR records. The dataset signal list is automatically generated from

all signals in the database. That is, the drug is related to EU-ADR events covering 11 clinically relevant adverse reactions.

Double-clicking on a dataset loads its content in a new workspace tab. This view lists all dataset signals and their respective data in a single table. This listing is enriched when the substantiation process is triggered (**Substantiate** action button), filling in the results from each external workflow and from the evidence combination analysis.

The **Workflows** menu loads the five EU-ADR workflows. In this view, each workflow is described and a variety of actions are displayed. Workflows can be exported for local execution or substantiated remotely with custom relationships or using the example signals.

The combination of dataset management features with targeted drug-event analysis features delivers an innovative framework for filtering and substantiation. With the inclusion of direct sharing possibilities, the EU-ADR Web Platform enables the creation of a pharmacovigilance collaborative research environment.

Discussion

For an assessment of EU-ADR Web Platform's applicability to a real-world research workflow, a sample drug analysis scenario is presented here. A researcher interested in studying potential adverse reactions of patients treated with a given drug, *Drug_X* for the purpose of this discussion, begins its study by automatically generating a dataset focused on the targeted drug. The system then combines this drug with the 11 potential adverse events considered in EU-ADR, substantiates the resulting signals using the available workflows and combines all individual pieces of evidence into an aggregate score representing the predicted risk of each drug-event relationship – Figure 5. Signals classified as moderately or highly risky should be further investigated by analysing presented evidence and following hyperlinks to biomedical literature, as well as to external drug and biological data resources.

(Figure 5)

Pharmacovigilance research over the last decades has focused mainly on evaluating the best strategies to identify and measure specific adverse drug reactions in a post-marketing stage[39-41]. The EU-ADR initiative further expands this trend by introducing a complete framework for drug-event interaction analysis, from electronic health records data sources to a researcher-oriented web-based workspace.

To our knowledge, the EU-ADR Web Platform is the only current tool allowing researchers to exploit the wealth of data for a vast European-wide cohort. It enables independent drug analysis crossed against the millions of collected records. Furthermore, rather than being a single proof-of-concept application for the EU-ADR research project, this platform opens the door for broader adverse drug reaction assessments beyond the limited EU-ADR event scope.

Conclusion

The EU-ADR European project embraces innovative pharmacovigilance research methodologies through the creation of a web platform providing advanced drug data exploration and assessment features. Whereas in the past post-marketing drug assessment required intense validation tasks, the *in silico* pharmacology community is now endowed with the tools required to quickly analyse specific adverse drug reactions, further improving drug safety monitoring.

The EU-ADR Web Platform enables streamlined access to drug dataset analysis features, including the evaluation of results from EU-ADR workflows and the sharing of data amongst research partners, all within a highly responsive and unique web-based workspace, which is available at http://bioinformatics.ua.pt/euadr.

References

1. Giacomini KM, Krauss RM, Roden DM, Eichelbaum M, Hayden MR, Nakamura Y 2007. When good drugs go bad. Nature 446:975-977.
2. EMA EMA. 2010. Annual Report. ed.: European Medicines Agency.
3. Robb MA, Racoosin JA, Sherman RE, Gross TP, Ball R, Reichman ME, Midthun K, Woodcock J 2012. The US Food and Drug Administration's Sentinel Initiative: Expanding the horizons of medical product safety. Pharmacoepidemiology and Drug Safety 21:9-11.
4. Xu L, Anchordoquy T 2011. Drug delivery trends in clinical trials and translational medicine: Challenges and opportunities in the delivery of nucleic acid-based therapeutics. Journal of Pharmaceutical Sciences 100(1):38-52.
5. Nelson JC, Cook AJ, Yu O, Dominguez C, Zhao S, Greene SK, Fireman BH, Jacobsen SJ, Weintraub ES, Jackson LA 2012. Challenges in the design and analysis of sequentially monitored postmarket safety surveillance evaluations using electronic observational health care data. Pharmacoepidemiology and Drug Safety 21:62-71.
6. Staffa JA, Dal Pan GJ 2012. Regulatory Innovation in Postmarketing Risk Assessment and Management. Clin Pharmacol Ther 91(3):555-557.
7. Shibata A, Hauben M. Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on, 5-8 July 2011 2011, pp 1-7.
8. McClure DL 2009. Improving Drug Safety: Active Surveillance Systems Should be Paramount. Pharmaceutical Medicine 23(3):127-130.
9. Stahl M, Edwards IR, Bowring G, Kiuru A, Lindquist M 2003. Assessing the Impact of Drug Safety Signals from the WHO Database Presented inSIGNAL': Results from a Questionnaire of National Pharmacovigilance Centres. Drug safety 26(10):721-727.
10. Meyboom RHB, Lindquist M, Egberts ACG, Edwards IR 2002. Signal Selection and Follow-Up in Pharmacovigilance. Drug safety 25(6):459-465.
11. Lazarou J, Pomeranz BH, Corey PN 1998. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. Jama The Journal Of The American Medical Association 279(15):1200-1205.
12. Alvarez-Requejo A, Carvajal A, Begaud B, Moride Y, Vega T, Arias LHM 1998. Under-reporting of adverse drug reactions Estimate based on a spontaneous reporting scheme and a sentinel system. European Journal of Clinical Pharmacology 54(6):483-488.
13. Grootheest V 1999. Attitudinal survey of voluntary reporting of adverse drug reactions. British Journal of Clinical Pharmacology 48(4):623-627.
14. De Bruin M, Van Puijenbroek E, Egberts A, Hoes A, Leufkens H 2002. Non‐sedating antihistamine drugs and cardiac arrhythmias–biased risk estimates from spontaneous reporting systems? British Journal of Clinical Pharmacology 53(4):370-374.
15. Wadman M 2007. Experts call for active surveillance of drug safety. Nature 446(7134):358-359.
16. Bauer-Mehren A, van Mullingen EM, Avillach P, Carrascosa MdC, Garcia-Serna R, Piñero J, Singh B, Lopes P, Oliveira JL, Diallo G, Ahlberg Helgee E, Boyer S, Mestres J, Sanz F, Kors JA, Furlong LI 2012. Automatic Filtering and Substantiation of Drug Safety Signals. PLoS Comput Biol 8(4):e1002457.
17. Härmark L, Van Grootheest AC 2008. Pharmacovigilance: methods, recent developments and future perspectives. European Journal of Clinical Pharmacology 64(8):743-752.

18. Garcia-Serna R, Mestres J 2010. Anticipating drug side effects by comparative pharmacology. Expert Opin Drug Metab Toxicol 6(10):1253-1263.
19. Mestres J, Seifert SA, Oprea TI 2011. Linking pharmacology to clinical reports: cyclobenzaprine and its possible association with serotonin syndrome. Clinical pharmacology and therapeutics 90(5):662-665.
20. Trifirò G, Pariente A, Coloma PM, Kors JA, Polimeni G, Miremont-Salamé G, Catania MA, Salvo F, David A, Moore N, Caputi AP, Sturkenboom M, Molokhia M, Hippisley-Cox J, Acedo CD, Van Der Lei J, Fourrier-Reglat A 2009. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? Pharmacoepidemiology and Drug Safety 18(12):1176-1184.
21. Ruttenberg A, Clark T, Bug W, Samwald M, Bodenreider O, Chen H, Doherty D, Forsberg K, Gao Y, Kashyap V, Kinoshita J, Luciano J, Marshall MS, Ogbuji C, Rees J, Stephens S, Wong GT, Wu E, Zaccagnini D, Hongsermeier T, Neumann E, Herman I, Cheung K-HH 2007. Advancing translational research with the Semantic Web. BMC Bioinformatics 8(Suppl 3):S2-S2.
22. Trifiro G, Fourrier-Reglat A, Sturkenboom MCJM, Diaz Acedo C, Van Der Lei J, Díaz Acedo C 2009. The EU-ADR project: preliminary results and perspective. Studies In Health Technology And Informatics 148:43-49.
23. Trontell AE 2005. The RADAR Project and the FDA. JAMA: The Journal of the American Medical Association 294(10):1206.
24. Coloma PM, Trifirò G, Schuemie MJ, Gini R, Herings R, Hippisley-Cox J, Mazzaglia G, Picelli G, Corrao G, Pedersen L, van der Lei J, Sturkenboom M, on behalf of the EUADRc 2012. Electronic healthcare databases for active drug safety surveillance: is there enough leverage? Pharmacoepidemiology and Drug Safety:n/a-n/a.
25. Parkinson J, Davis S, Staa Tv. 2006. The General Practice Research Database: Now and the Future. Pharmacovigilance, ed.: John Wiley & Sons, Ltd. p 341-348.
26. Wood L, Martinez C 2004. The General Practice Research Database: Role in Pharmacovigilance. Drug safety 27(12):871-881.
27. Stark RG, John J, Leidl R 2011. Health care use and costs of adverse drug events emerging from outpatient treatment in Germany: A modelling approach. BMC Health Services Research 11:9-9.
28. Chan KA, Hauben M 2005. Signal detection in pharmacovigilance: empirical evaluation of data mining tools. Pharmacoepidemiology and Drug Safety 14(9):597-599.
29. Avillach P, Mougin F, Joubert M, Thiessard F, Pariente A, Dufour JCCJ-C, Trifirò G, Polimeni G, Catania MA, Giaquinto C, Mazzaglia G, Baio G, Herings R, Gini R, Hippisley-Cox J, Molokhia M, Pedersen L, Fourrier-Réglat A, Sturkenboom M, Fieschi M, Fornari C, Trifiro G, Fourrier-Reglat A 2009. A semantic approach for the homogeneous identification of events in eight patient databases: a contribution to the European eu-ADR project. Medical Informatics 160(Pt 2):1085-1089.
30. Trifiro G, Patadia V, Schuemie MJ, Coloma PM, Gini R, Herings R, Hippisley-Cox J, Mazzaglia G, Giaquinto C, Scotti L, Pedersen L, Avillach P, Sturkenboom MC, van der Lei J, The Eu-Adr G 2011. EU-ADR Healthcare Database Network vs. Spontaneous Reporting System Database: Preliminary Comparison of Signal Detection. Studies in health technology and informatics 166:25-30.
31. Trifirò G, Patadia V, Schuemie MJ, Coloma PM, Gini R, Herings R, Hippisley-Cox J, Mazzaglia G, Giaquinto C, Scotti L, Pedersen L, Avillach P, Sturkenboom MCJM, van der Lei J, Trifiro G, The Eu-Adr G 2011. EU-ADR healthcare database network vs. spontaneous reporting system database: preliminary comparison of signal detection. Studies In Health Technology And Informatics 166:25-30.

32. Ludascher B, Altintas I, Berkley C, Higgings D, Jaeger E, Jones M, Lee EA, Tao J, Zhao Y 2006. Taverna: Scientific Workflow Management and the Kepler System. Research Articles, Concurrency and Computation: Practice & Experience 18(10):1039 - 1065.
33. Gil Y, Deelman E, Ellisman M, Fahringer T, Fox G, Gannon D, Goble C, Livny M, Moreau L, Myers J 2007. Examining the Challenges of Scientific Workflows. Computer 40(12):24-32.
34. Avillach P, Joubert M, Thiessard F, Trifiro G, Dufour JC, Pariente A, Mougin F, Polimeni G, Catania MA, Giaquinto C, Mazzaglia G, Fornari C, Herings R, Gini R, Hippisley-Cox J, Molokhia M, Pedersen L, Fourrier-Reglat A, Sturkenboom M, Fieschi M 2010. Design and evaluation of a semantic approach for the homogeneous identification of events in eight patient databases: a contribution to the European EU-ADR project. Studies in health technology and informatics 160(Pt 2):1085-1089.
35. de Leon J 2011. Highlights of Drug Package Inserts and the Website DailyMed: The Need for Further Improvement in Package Inserts to Help Busy Prescribers. J Clin Psychopharmacol.
36. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M 2008. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Research 36(suppl 1):D901-D906.
37. Zadeh LA 1986. A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. AI magazine 7(2):85.
38. Dudley JT, Butte AJ 2010. In silico research in the era of cloud computing. Nat Biotech 28(11):1181-1185.
39. Papay J, Yuen N, Powell G, Mockenhaupt M, Bogenrieder T 2012. Spontaneous adverse event reports of Stevens–Johnson syndrome/toxic epidermal necrolysis: detecting associations with medications. Pharmacoepidemiology and Drug Safety 21(3):289-296.
40. Sommet A, Durrieu G, Lapeyre-Mestre M, Montastruc J-L, The Association of French PharmacoVigilance C 2012. A comparative study of adverse drug reactions during two heat waves that occurred in France in 2003 and 2006. Pharmacoepidemiology and Drug Safety 21(3):285-288.
41. Vilar S, Harpaz R, Chase HS, Costanzi S, Rabadan R, Friedman C 2011. Facilitating adverse drug event detection in pharmacovigilance databases using molecular structure similarity: application to rhabdomyolysis. Journal of the American Medical Informatics Association.

Figures

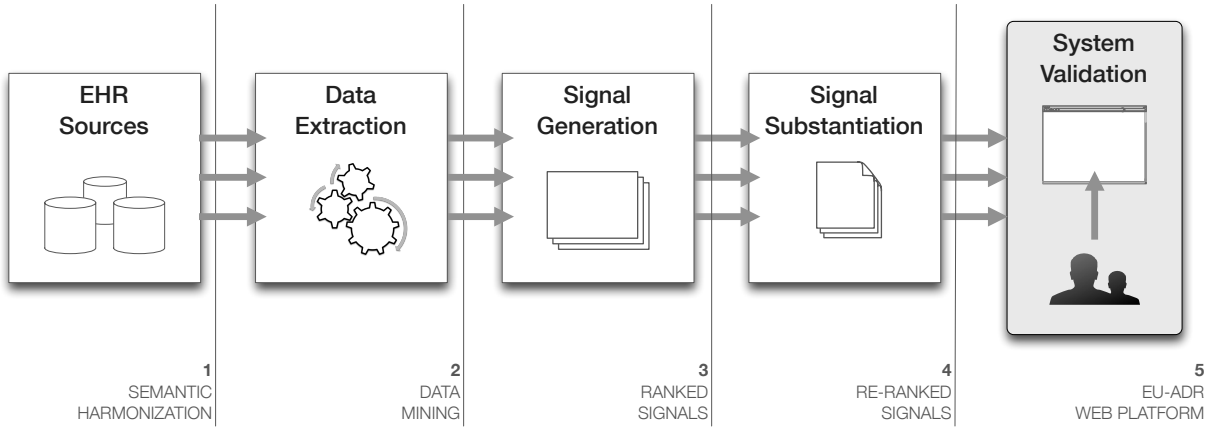

Figure 1

EU-ADR data flow. 1) Data from electronic health record (EHR) resources is semantically harmonized for data extraction. 2) Extracted data is mined for drug-event pairs and other relationships. 3) The signal generation process takes mined data and forms the first ranked signal dataset. 4) The signal substantiation process re-ranks the signal list, based on evidences from biomedical databases and literature, *in silico* simulations and pathway analyses. 5) The EU-ADR Web Platform enables completing the retrospective and prospective system validation.

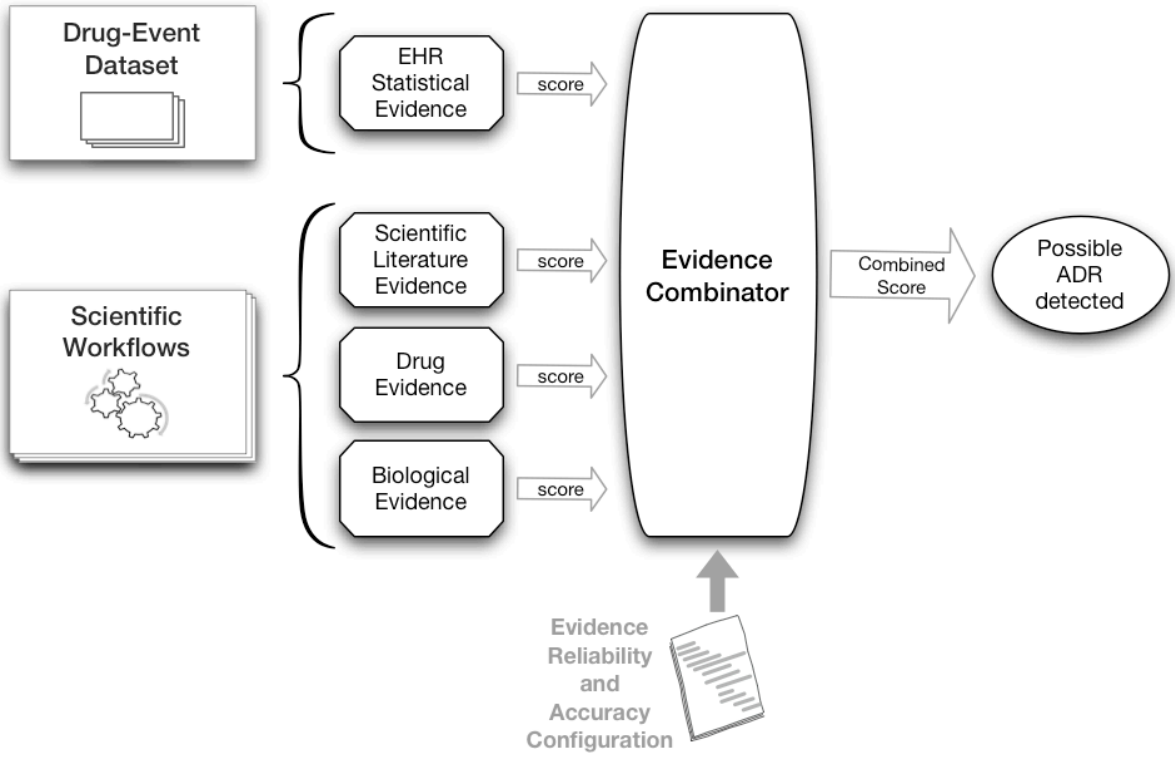

Figure 2

Evidence combination process. Various evidence scores from multiple sources are combined into a single score using configurable reliability and accuracy parameters for each evidence source. The Dempster-Shafer theory is used to arrive at a degree of belief that takes into account all the available evidence and facilitate detection of possible adverse drug reactions.

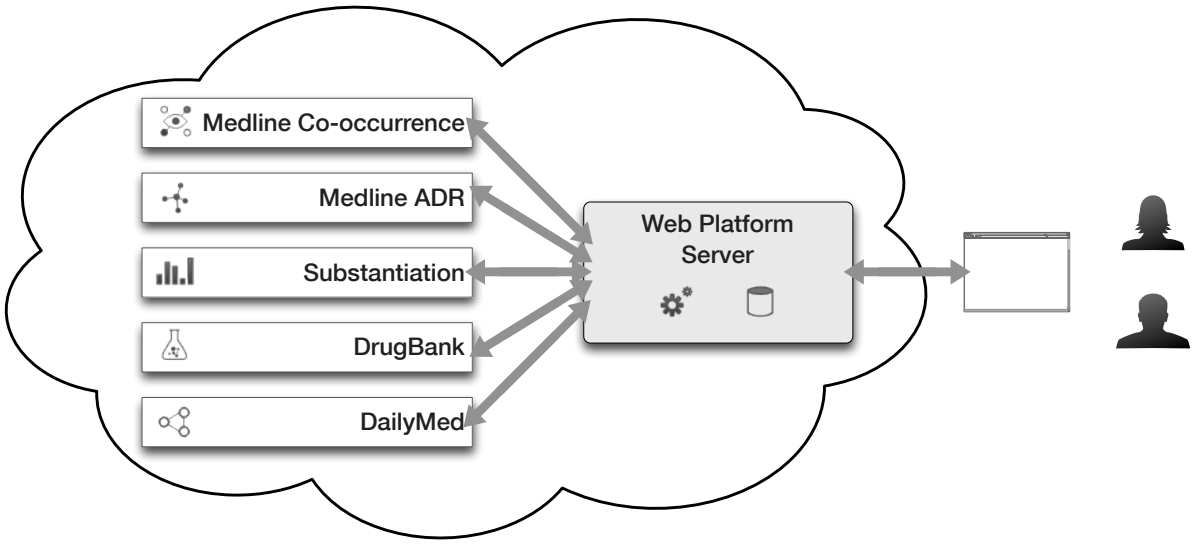

Figure 3

Simplified EU-ADR Web Platform setup. From right to left: users interact with the EU-ADR Web Platform using any modern web browser; data is exchanged with the application controllers in the Web Platform server, pushing data to and pulling data from the internal database; the ranked signal list is obtained through the communication with external distributed services, stored in the internal database and published to users.

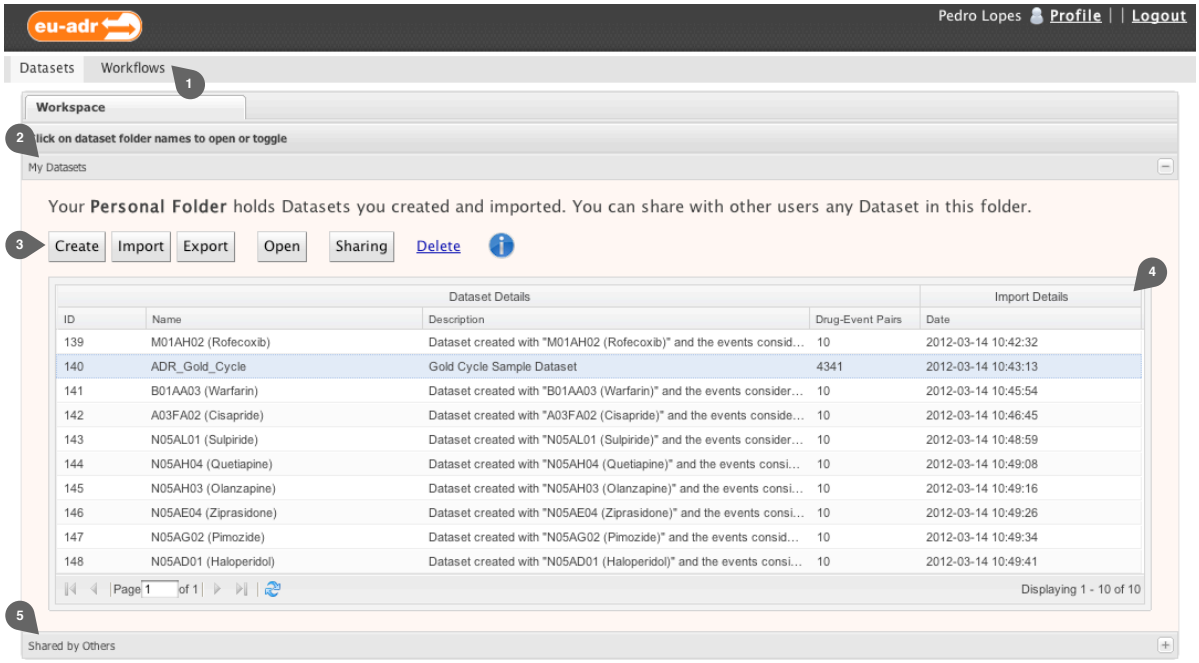

Figure 4

EU-ADR Web Platform Dataset list view. 1) Access to workflows list view. 2) “My Datasets” section listing user’s dataset list. 3) Dataset action buttons, from left to right: create new targeted datasets, import dataset from local file, export online dataset to local file, open selected dataset, share selected dataset, delete selected dataset, help. 4) Dataset list table detailing dataset name, description, number of drug-event pairs and creation date. 5) Access to “Shared by Others” section, listing datasets shared with the user.

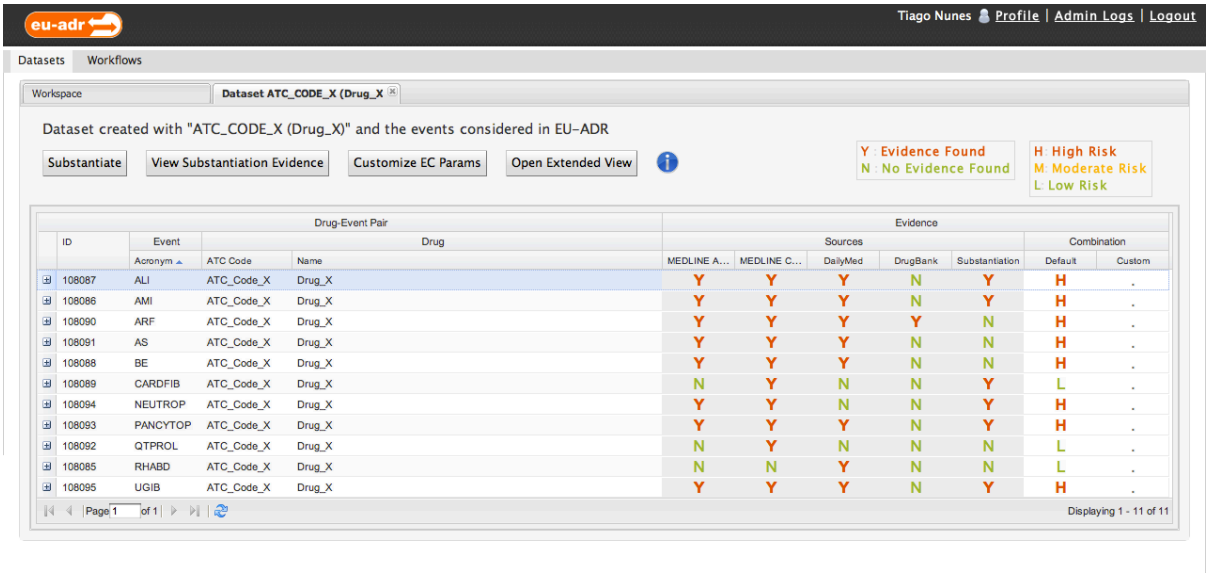

Figure 5

EU-ADR Web Platform results for an undisclosed drug (*Drug_X*) exploration scenario containing the signal list that results from distributed workflow outputs and evidence combination statistical analysis. Workflow results are labelled with Y in case sufficient evidence is found to support a potential drug-event relationship, or N otherwise. Evidence combination yields a score of H, M or L, indicating High, Moderate or Low risk, respectively, of a drug-event relationship being in fact an ADR signal.

Tables

ATC	Name	EventType	Exposure	Events	RR(MH)	...
A01AA01	Sodium fluoride	ALI	5.302.087	4	3,18	...
A01AA02	Sodium fluoride	AMI	4.897.540	6	2,39	...
...	...	...	...	...	...	...

Mandatory

Recommended

Custom (optional)

Table 1

EU-ADR standard dataset format. **EventType** and **ATC** fields are the mandatory attributes that make up a potential ADR signal. The recommended **Name** field represents the drug name. If omitted, names of drugs are looked up in an internal drug database. Any additional signal attribute is imported as is and later presented in the application's dataset visualization interface.