

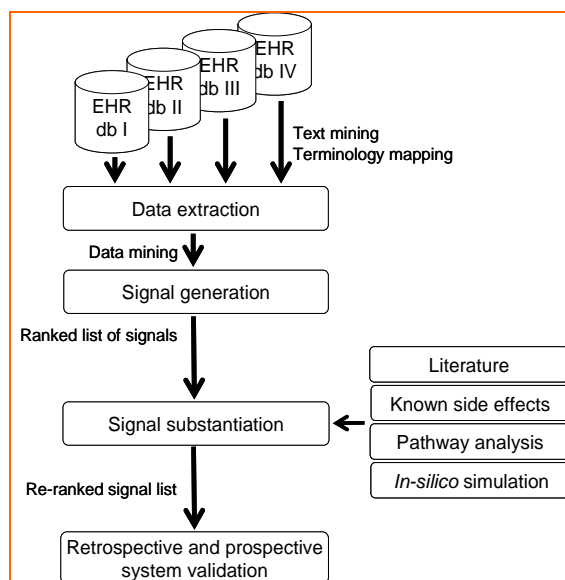
**Exploring and Understanding Adverse Drug Reactions by Integrative Mining of Clinical Records and Biomedical Knowledge**  
 (EU-ADR- Grant Agreement number 215847)

Before launching a new drug to the market, it is tested on a few thousands of people, but adverse drug reactions (ADRs) may not be detected until many more patients have used the drug. Once the drug is on the market, clinicians are responsible for recognizing and reporting suspected side effects, which are collected in so-called spontaneous reporting systems. However, a number of recent, highly publicized drug safety issues showed that adverse effects of drugs may be detected too late, when millions of patients have already been exposed.

EU-ADR intends to develop an alternative approach for the detection of ADR signals. Rather than relying on the physician’s capability and willingness to recognize and report suspected ADRs, the resulting system will systematically calculate the occurrence of disease (potentially ADRs) during specific drug use based on data available in electronic patient records. In this project, electronic health records (EHRs) of over 30 million patients from several European countries will be available. In an environment where rapid signal detection is feasible, rapid signal assessment is equally important. To rapidly assess signals, a number of resources will be used to substantiate the signals: causal reasoning based on information in the EHRs, semantic mining of the biomedical literature, and computational analysis of biological and chemical information (drugs, targets, anti-targets, SNPs, pathways, etc.).

**The EU-ADR system: a focus on data extraction.**

The overall objective of EU-ADR is the design, development and validation of a computerized system that exploits data from electronic healthcare records and biomedical databases for the early detection of adverse drug reactions. The EU-ADR system will generate signals using data and text mining, epidemiological and other computational techniques, and subsequently substantiate these signals in the light of current knowledge of biological mechanisms and *in silico* prediction capabilities. The system should be able to detect signals better and faster than spontaneous reporting systems and should allow for identification of subpopulations at higher risk for ADRs.



In this context, a common framework for data extraction and aggregation has been established. Because sharing of patient level data is often not possible due to privacy concerns and governance regulation, aggregation needs to be performed before the data is collected and processed at a central location. A software tool called Jerboa has been developed that can be run by an EHR database in its local

environment. Jerboa takes a number of input files containing data on drug exposure, occurrence of adverse events, and patient information. The output file produced by Jerboa consists of the number of events and the exposure time, stratified according to ATC code (or combination of ATC codes), age category, and gender. The output file of each database is encrypted and uploaded for subsequent processing. Data processing and aggregation parameters of Jerboa are specified in a script file. This makes it easy to test different parameter settings and facilitates uniform data aggregation across the databases. Jerboa is written in Java, and thus it is highly portable. So far, the different EHR database owners have tested Jerboa in their local environments with satisfactory results.

Moreover, Jerboa is being used or will be used by other European Projects (VAESCO, SOS, ARITMO, ARPEC) and by other international initiatives (OMOP).