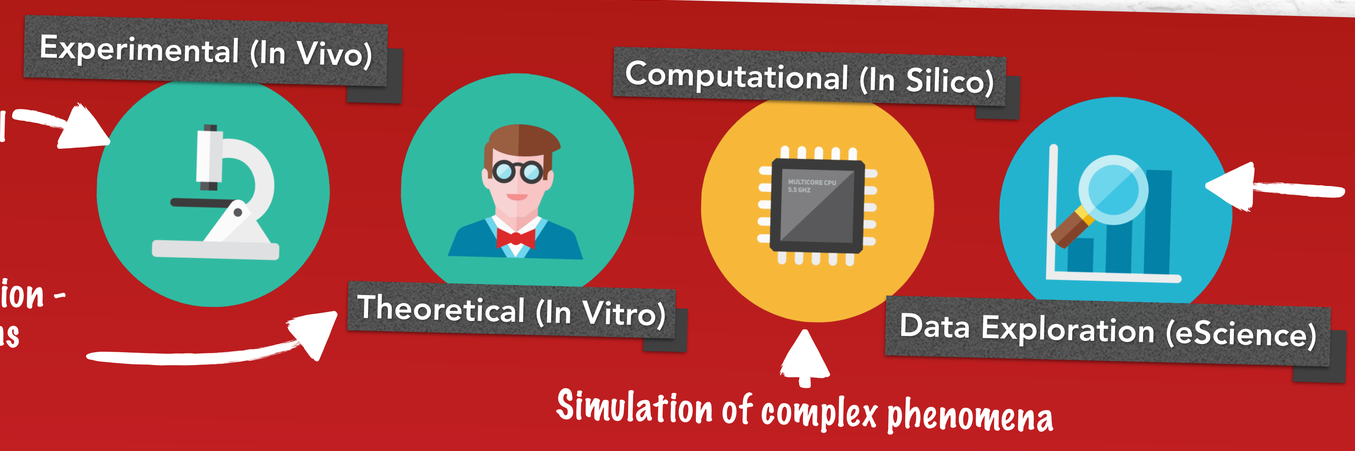
Gwénaël Dumont [a, b]
Omar Boucelma [b]
Isabelle Perseil [a]

BigPharma - Processing Large Pharmacovigilance Data Sets

Context



Scientific research nowadays involves exploration of data from various sources; the result of this aggregation of data with different formats and structures ultimately form massive data sets or big data.

The challenges of eScience are the standardization of data sets, infrastructures sharing and interoperability to build platforms accessible to biomedical and health researchers.

The goal of this study is to explore a new way to optimize the processing of the massive data sets in pharmacovigilance by proposing algorithms, languages and services.

Definitions

Pharmacovigilance: ``The science and activities relating to the **detection, assessment, understanding and prevention of adverse effects** or any other drug-related problem.''

Signal: ``Reported information on a **possible causal relationship between an adverse event and a drug**, the relationship being unknown or incompletely documented previously. Usually more than a single report is required to generate a signal, depending upon the seriousness of the event and the quality of the information.''

Challenges

Structures & **formats** of Data & Metadata

Interoperability across heterogeneous data (structured & unstructured)

Size & **number** of files

- Hierarchical **storage**
- Parallel & distributed **systems**

Massive data processing

- Connectivity & **bandwidth**
- Access

- **Analytics**
- **Statistics**

Data **anonymization**

Ethical & **legal** regulations

Methods

State of the art

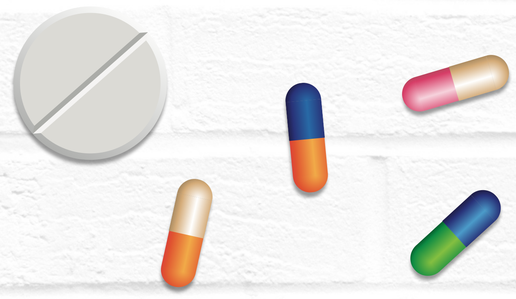
- Define **infrastructures** and **data store** (Hadoop, NoSQL, NewSQL, …)
- **Ingest, transform** and **normalize** (Storm, Yarn, ZooKeeper, …)
- **Analyze** results (Impala, Drill, Mahout, Spark, Flink, …)

Programming Model

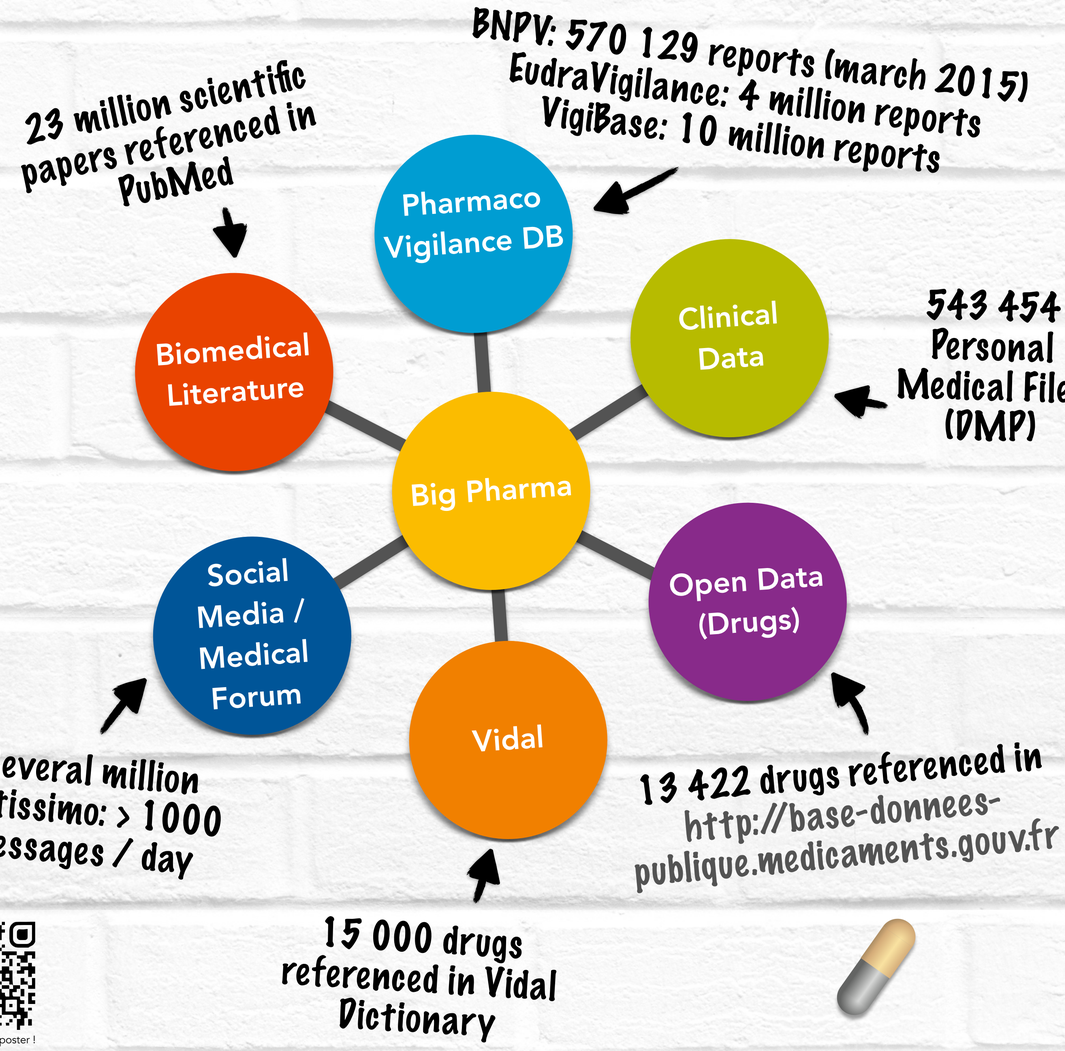
- **Dynamically decompose, distribute** and **optimize** processing functions
- Test **various infrastructures** (Grid -with IRods- vs academic Cloud)
- **Improve the overall quality** of the case studies

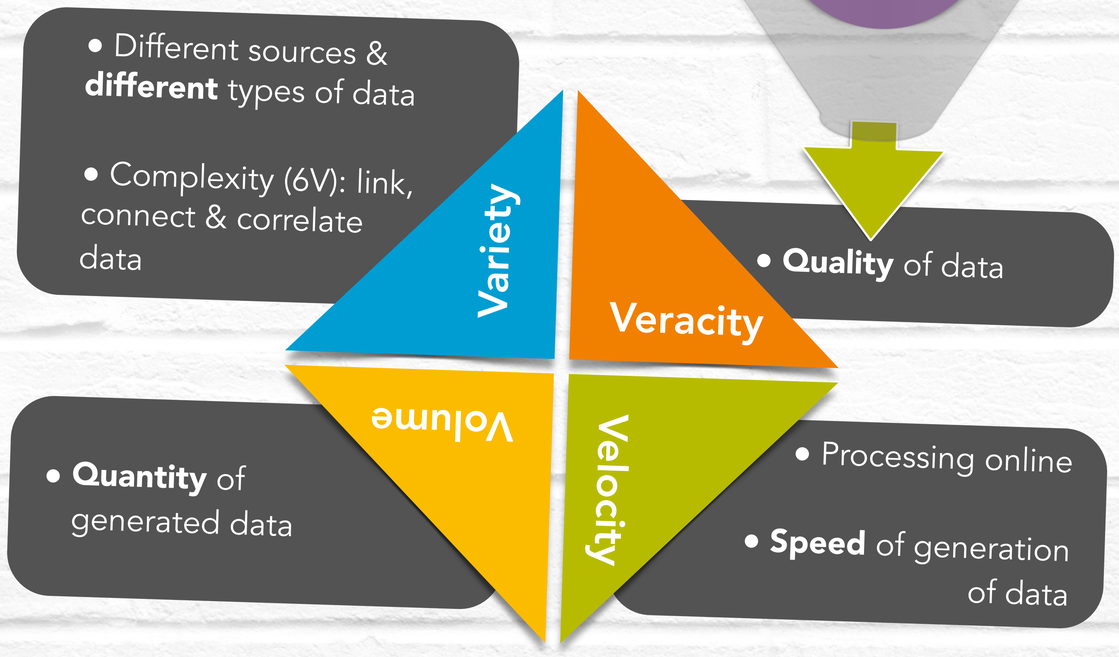
Services

- **Develop services** on top of this prototype
- Make them available to biomedical and health researchers in a **friendly environment**



MedDRA
WHO-ART
Bayesian methods
Luxid Navigator BioPharma Edition





Download this poster !

Inserm · Aix*Marseille université · cnrs · UNIVERSITÉ DE TOULON · ARTS ET MÉTIERS ParisTech

[a] Computational Science Coordination at Inserm, IT Department, INSERM, 13 rue Watt, F-75013 Paris, France
[b] LSIS, Aix-Marseille University, CNRS, Av. Escadrille Normandie-Niemen, F-13397 Marseille Cedex 20, France