Big Data Analytics in Healthcare: Overcoming Legal Obstacles

jos.dumortier@timelex.eu

ISSE 2016 – Issy-Les-Moulineaux
Introduction

AEGLE is a European Union “Horizon 2020” Innovation Action

Interdisciplinary team with partners from Belgium, France, Greece, Italy, Netherlands, Portugal, Sweden & UK

Started in March 2015 (42 months)
Objective

Reference big-data architecture

Covering a large part of the health spectrum

- Malignant chronic diseases
- Non-malignant chronic diseases
- Acute care
AEGLE Principles

Alignment with the data value chain

Inventory of available data
Syntax structure and semantics
Integrated data analysis
Analytic results to decision makers
Inform based decisions

Access control rules
Common representation of data

collect prepare organize integrate analyze visualize decide

DII ICU CLL

Providers
Hospitals
Specialists
Patients
Researchers

Users
Stakeholders

Product Oriented
Research Oriented

itechlaw Madrid - November 2016
www.aegle-uhealth.eu
info@aegle-uhealth.eu
User-oriented Challenges

- Understand the exploitation perspective of real-life big *bioclinical* data for diverse use cases

- Provide a framework to accommodate diverse big *bioclinical* data management and analytics requirements according to the *data type* and the *application domain*

- Enable scientific question answering by exploiting big *bioclinical* data in a way not possible until now
Technical Challenges

- Address the requirements posed by different types of big *bioclinical* data:
  - Biomolecular and clinical data (the CLL use case)
  - Real-time streaming and clinical events data (the ICU use case)
  - Large observational healthcare databases (the DII use case)

- Implement the necessary mechanisms for comprehensive data management and analytics, fully-compliant with privacy, legal and ethical norms (e.g. anonymisation, policies, etc.)

- Provide efficient response time (when needed) via acceleration technologies

- Offer a scalable & sustainable IT solution
Business Challenges

- Create sustainable business models taking into account the needs of the three scenarios to which AEGLE is applied
- Accurately identify the exploitable items
- Define a strategy for the long-term viability of the platform
- Create an ecosystem of stakeholders
AEGLE Use Case Scenarios
Intensive Care Unit (ICU)

Challenges

- Mechanical ventilation & patient-ventilator interactions
- Personalization of Patient Care & Early Identification of Deterioration

Questions to be answered

- Ineffective Efforts (IE) characteristics
- Recognition, Incidence, Significance, Prediction of IE
- Tools to guide and monitor nutrition
- Tools to identify early deteriorating trends
Chronic Lymphocytic Leukemia (CLL)

Challenges

• Clinically and biologically heterogeneous
• Optimal care and treatment decisions depend on the integration of tumor- and host-derived variables

Questions to be answered

• Identification of novel prognostic markers
• Prediction model for Monoclonal B Lymphocytosis evolution
• Prediction model for CLL natural course
Type 2 Diabetes (T2DM)

Challenges

- Long term condition
- Increasing in prevalent
- Increasing in morbidity and mortality
- Improving disease management

Questions to be answered

- Define why some cases do better than others – prognostic indicators
- Improve patient outcomes – methods of intervention
- Define accurate cohort and feasibility for doing clinical trial
- Identify potential points for intervention and types of intervention available
Where we stand now?

- User-centered design approach for developing a big *bioclinical* data analytics platform

- First release of the AEGLE system architecture

- Rapid prototyping for proof-of-concept illustration and user engagement: data, analytics and data management mechanisms already in place

- Beginning of first validation phase

- Initial legal and ethical assessment

- First steps taken on the business landscape for AEGLE
AEGLE Big Data Framework Software Stack

- **WebHDFS**: HDFS REST API
- **RM RT**: YARN REST API
- **LIVY**: SPARK REST API

**SQOOP**
- data transfer to HDFS
- **HIVE**: hadoop sql api
- **PIG**: scripting workflow mgnt
- **Pydoop**: Python hadoop api

**HADOOP MAPREDUCE**
- distributed engine for batch jobs processing

**YARN**
- virtual cluster resource manager

**HDFS2**
- virtual cluster distributed file system

**SPARK**
- distributed engine for fast in-memory processing

**SPARK SQL**
- sql api for spark

**MLlib**
- machine learning library

**R MAN**
- YARN REST API

**VM Node**
Legal Challenges

- Identify and incorporate all regulatory issues underlying the realization of the project aims

- Contribute in the definition of a common regulatory framework for big bioclinical data management and analytics at the European level
Two legal perspectives

1. Short term: legal compliance of research & innovation activities in the scope of the AEGLE action (all phases of the data value chain)

2. Long term: legal framework for the AEGLE start-up (and for other European “big data” initiatives in the health sector)
1. Short term

- Focus on compliance of AEGLE RIA with current data protection law (“avoid doing something illegal”)

- Core provision (Directive 95/46, article 6.1(b)): “Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards”

- AEGLE: further processing of clinical data for scientific purposes

- Main question to examine: which “appropriate safeguards” of (which) Member States have to be taken into account? (example: need of approval of DPA if re-use is not based on patient consent)
2. Longer term: EU legal framework for big data processing in the health sector

- Context: AEGLE start-up

- Initial legal framework: General Data Protection Regulation

- Core provision: art. 6.1(b) juncto art. 83

- Objective: analyse the (current/developing) legal situation in the 28 Member States + recommend possible EU initiatives
Jos Dumortier

time.lex - Information & Technology Law
Sablon Tower – rue Joseph Stevens 7
B-1000 Brussels
(t) +32 (0)2 893 20 95
www.timelex.eu / jos.dumortier@timelex.eu

SSRN: https://ssrn.com/author=566730

Co-funded by the Horizon
2020 Framework Programme
of the European Union under
Grant Agreement n° 644906.

Partners
EXUS AE (Coordinator), ICCS, KINGSTON, CERTH, Maxeler
Technologies Limited, UPPSALA UNIVERSITET, UNISR,
Time.Lex, EUR, CHS, LOBA, PAGNI, GNUBILA FRANCE