French opportunities of AI in Health

Emmanuel Bacry

Senior Researcher CNRS
Univ. Paris-Dauphine, PSL

Chief Scientific Officer
French Health Data Hub

Head of Data/Health projects
Ecole Polytechnique

emmanuel.bacry@polytechnique.fr
http://www.cmap.polytechnique.fr/~bacry

Health data in France

- Huge hospital databases (APHP, ldots)
- Cohort databases
- . . .
- **SNDS : one of the largest administrative healthcare databases in the world!**

    - A national database containing all healthcare reimbursements
    - More than 65 millions persons
    - $\simeq$ 400 Tb

A huge potential societal impact on health

$\longrightarrow$ **pharmaco-vigilance** : Identify drugs currently on the market that cause harmful side effects

But also a huge potential societal impact on the economy
(Public health budget in France $\simeq$ 200 billion/year)

$\longrightarrow$ **Care Path Optimization** for a given pathology

- Research partnership between Polytechnique and CNAM 2015-2017, 2018-2020
- 2017: team of 5 developers, 1 doctor, 2 postdocs, 3 researchers
- **Developed an automatic "screening" algorithm $\longrightarrow$ first successes in December 2017**!
- Care path (visualization, . . .)
- Fraud

$\Longrightarrow$ **Restructuring the infrastructure at CNAM !**

- **Actual SNDS infrastructure : not adapted to distributed computing needed for large scale statistical analysis**

- **Migration of infrastructure : proof of concept put in place by the partnership (sur $\simeq$ 50Tb)**
    - **Data Infrastructure**
      Proprietary SQL database (800 tables) $\longrightarrow$ large array centered on the person (open source architecture)
    - **Computer infrastructure**
      "vertical" proprietary architecture $\longrightarrow$ "horizontal" open-source architecture (distributed data)
    - **Software infrastructure for statistical analysis**
      Proprietary softwares $\longrightarrow$ Open-source libraries (Scala, Spark, C++, Python, ...)

- Cohort : **seniors**
- Number of Medications : >**400)**
- Adverse effect : **fall**

Few numbers (HDFS cluster 20 nodes)

- 12 million persons
- 4 years monitoring
- $\simeq$ 8Tb per year
- Flattening the database $\simeq$ 2-3 days
- Data processing (featuring) $\simeq$ 1 hour
- Inference duration $\simeq$ 1 hour

The Health Data Hub : A national project

- 2018 : Pitched by Member of Parliament (and Fields Medal) Cédric Villani in his report on AI
- Started in January 2019
- Budget : 80m € for the first 4 years

Main ambition : *Put at the service of the greatest number, health data financed by the national solidarity in the respect of the ethics and fundamental rights of our fellow citizens.*

More precisely

1. Consolidate massive and heteregeneous health data assets on a modern infrastructure
   - Massive data : SNDS, Public hospital data, private companies data, . . .
   - Heterogeneous data : health records, images, genomic, biological data, . . .

2. Sharing data according to a unified governance and guaranteed by law
   - Current evolution of the legislative framework

More precisely (continued)

3. Creation of a trusted third party to support access ($+$ shared ressources) for valuing data in a secure environment
- Strong connections with all the actors fo the actual (public/private) ecosystem
- Become one of the main actor responsible for the development of this ecosystem

4. Identification of promising projects in terms of public health and promotion of the industrial sector
- Become one of the best health data platforms in the world

5. Become a referenced platform for matching skills and interests

and ... a (essential) counter power to the GAFAM/BATX.

The first selected projects (10 out of . . . 189)

- **DeepSark** : Almost **complete database on sarcoma** in France
- **NS-Park** : Cohort of **20k Parkinson patients** (clinical, genomic, ...)
- **Rexetris** : Renal transplant patients (Cristal ABM database)
- **Deep.Piste** : **250k** annotated **mamography**
- **Hydro** : **Pacemaker** data of more than **8000 patients**

The first selected projects (10 out of ... 189)

- **Oscour** : Exhaustive databases of **emergency services** in France
- **Pimpon** : Vidal database
- **Ordei** : ANSM Pharmacovigilance database
- **Parcours IDM en IDF** : SAMU database in IdF
- **ARAC** : Full data from the Malakoff Mederic (French health mutual company) (several million people)

Some future important actions of the Health Data Hub

- Mobilization of a network of researchers (doctors, biologists, economists, computer scientists, mathematicians, . . .)
  - Some "hot" topics essential to the health data hub : privacy, interpretability, . . .
  - Meetings : INRIA 4/07/19, CNRS 21/06/19, INSERM 1/07/19, CEA, . . .

- Mobilization of health data producers

- Mobilization of the industrial ecosystem (including startups !)

- Build a repository (+ FAQ + forum + ...) for algorithms, softwares, database knowledge . . .

- (European) project calls related to health

- Develop links with other foreign national health projects

WE HIRE !

`www.health-data-hub.fr`
`jobs@health-data-hub.fr`

- We hire world-class project managers + data scientists/engineers to work on unique and diverse data
- We need your creativity and enthusiasm to
  - Set up the platform
  - Bring your expertise to very various projects
- We offer
  - agility/flexibility of a start-up
  - visibility of a government flagship project