Hospital information systems

Heterogeneity Unicity

Extensive patient data collections

Free-text reports



Medical imaging

Drug prescriptions



Medical observations



Administrative data



Functionnal investigations

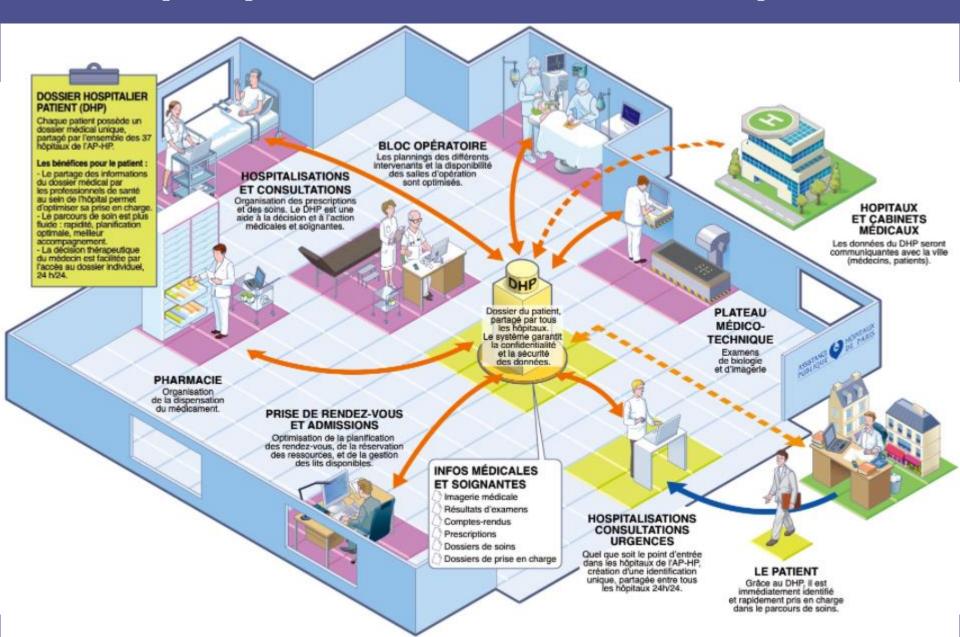


Laboratory results

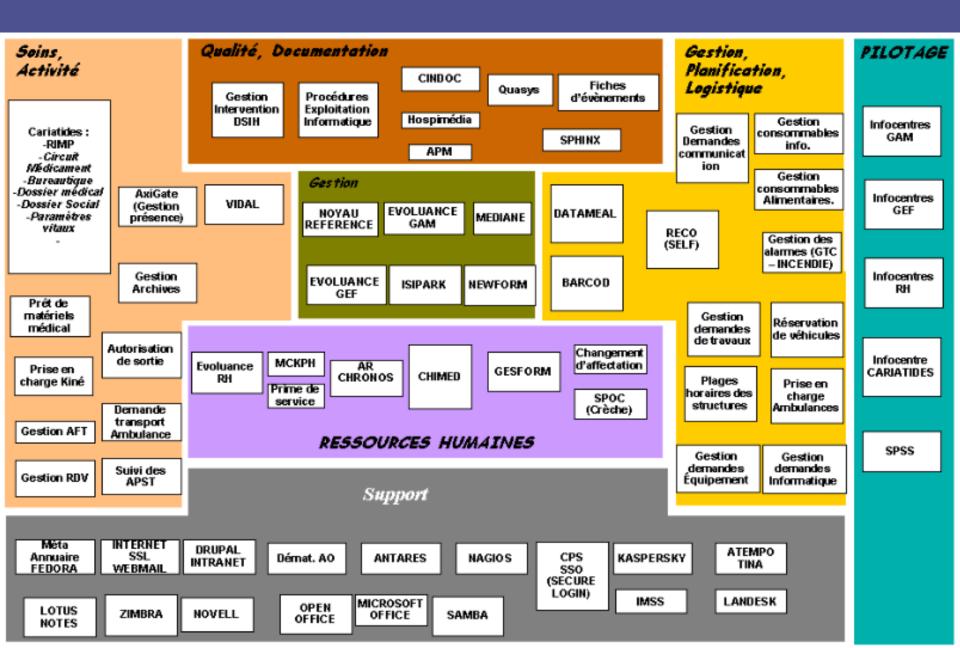




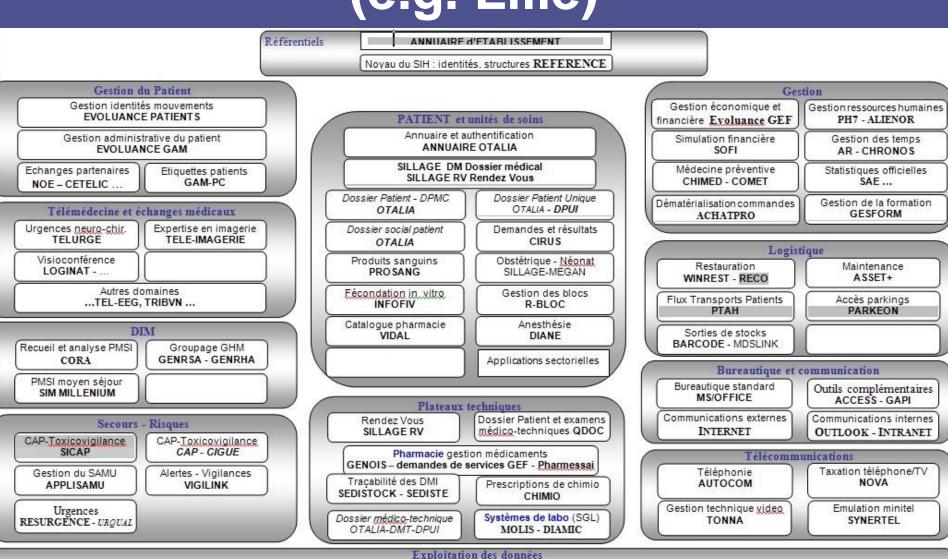
Multiple processes around the patient



Multiple processes around the care



Multiple processes around the care (e.g. Lille)



Medico-technique

INFO SERVICE QDOC

Données gestion économique

GEF-EPRD-dépenses

INFOSERVICE

GRH - PH7

Comptabilité analytique

SEXTANT

INFO SERVICE PATIENT

GAM - CORA - DIANE

INFOSERVICE

racabilité activité

INFOSERVICE

BIOLOGIE

données cliniques

INTERROSIM

Patients' data in the hospital information systems (e.g. CHU Lille)

Medical activity

- 220,000 inpatient stays / year
- 1,000,000 outpatient visits / year
- 5,500 birth / year
- 70,000 surgeries / year
- 100,000 CT or MRI scan / year

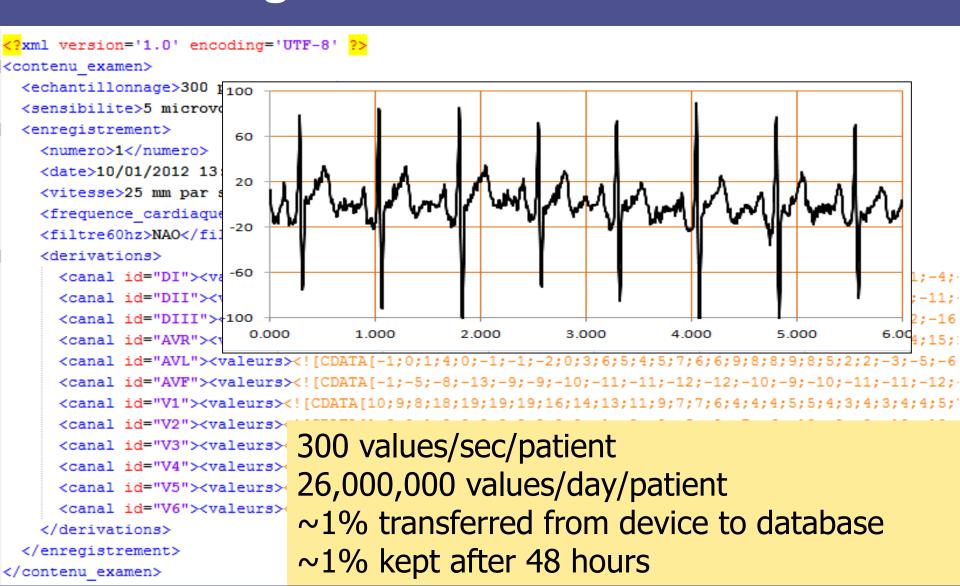
Volumetry

- Hundreds of pieces of software
- ~ 100,000 tables and relationships
- ~ 1,200,000 admissions
- ~ 30,000,000 laboratory results / year
- ~ 50 Tb/year for medical imaging
- ~ 50 Tb/year for omics data

Patients' data in the hospital information systems (e.g. CHU Lille)

- Heterogeneous data types
 - Mainly coded with terminologies
 - Some quantitative data
 - Some signal data, at high speed and high volume
 - Medical images, very large volume
 - Omics data, very large volume

Example of electrocardiographic monitoring data in intensive care units



Hospital information system

- Mainly relational databases
- Transactional use: highly normalized, many tables and relationships
- A lot of metadata aimed at tracing updates and describing the information flow (e.g. not only drugs and doses, but prescription, delivery, administrations, corrections, actions, etc.)
- Very efficient for the care of 1 patient select * from xxx where patient_id=xxx
- Not efficient at all for decisional analysis

Issues

- Datawarehouse conception
 - Denormalize and simplify the data schema
 - Filter the data of interest (e.g. sacrifice traceability metadata)
 - How to make structured and unstructured data (texts, images, omics) cohabit?
- Connection difficulties to feed a warehouse:
 - Do not slow down the production servers during exports
 - To export data that are often serialized

Our project of platform for science researchers

Findings...

- Machine learning Researchers:
 - Have difficulties to analyze health data
 - Sometimes analyze medical images
 - Rarely analyze structured data
- Barriers:
 - Difficulty to access health data: legal rules, data privacy
 - Need to understand medical issues
 - Need to understand health data

Our project of health data platform

- Simulated & realistic data available for researchers:
 - From open data, interpolated to be expressed on an individual basis (no more groups of patients)
 - From existing datasets:
 - Hospital datasets (encoded values, lab results, drugs, clinical textes, etc.)
 - Physiological signals
 - Nationwide databases (health insurance claims, etc.)
 - Downloadable by researchers
- Trainings: available data, basics of medicine (anatomy, physiology, pathology, therapy)
- Packages for data analysis (new packages, or at least easy-to-use wrappers)
- Services for data analysis: feature extraction...