

SophiaConf 
2014 du 30 juin au 3 juillet

Open Source Web Sémantique Google
Web Informatique Cloud Sécurité
Internet of things DIY BI Big Data
No SQL Développement Frameworks

5^{ème} édition

Big data, No SQL, BI et sécurité du cloud

"Le datalab : clef de voûte du système décisionnel ?" par URAEUS Consult

Edmond Cissé

URÆUS Consult
Ingénierie & Sécurité des Données
www.uraeus-consult.com

edcisse@uraeus-consult.com

04 93 95 52 10

Historique du BI

Depuis 20 ans, le marché BI est stable et maintenu dans des zones de confort empêchant les innovations technologiques (internet et stockage massif) disponibles de près de 10 ans d'émerger et de faire évoluer les pratiques décisionnelles.

La pression des directions métiers est très forte

« L'IT a du mal à suivre le rythme des demandes de notre équipe en charge du traitement analytique qui est trop en avance pour la fonction technologique. »

« Je ne peux pas expliquer à mon DG que les prévisions ne sont optimisées car que nous n'avons pas suffisamment de flexibilité et de contrôle sur les données nécessaires »

« Je suis obligé de mettre en place une infrastructure dédiée autonome adaptée à nos besoins, avec son propre entrepôt . J'ai une certaine autonomie budgétaire ! »

« Les informations dont nous avons besoin sont en silos et réparties à travers différentes bases de données ! »

Les exigences métiers

- 👉 Nouvelles sources de données hors du périmètre référentiel de l'entreprise
- 👉 Analyse exploratoire de données non-structurées et non-modélisées
- 👉 BI dédié à usage ponctuel (POC) dans un délai raisonnable et à coût maîtrisé
- 👉

Oui à l'appel du BigData mais ...

- 👉 75% des décideurs français interrogés portent un intérêt au big data
- 👉 Projets conduits (6%) - en cours - budget inconnu (31%), pas de projet (38%)
- 👉 Telecom, Banque/Assurance, Distribution ..

Les principaux freins au l'adoption du BigData

- 👉 Difficulté à identifier des gains concrets ou du ROI pour vendre le projet en interne
- 👉 Manque de budget
- 👉 Difficulté à identifier l'impact potentiel sur les SID en place
- 👉 Manque de compétences et de ressources en interne
- 👉 Manque d'initiative ou de compréhension de la DG
- 👉 Pas de connaissance des solutions/offres
- 👉 Mauvaise qualité des données

Les principaux risques

- 👉 Frein à la performance analytique et décisionnelle de l'entreprise
- 👉 Promotion par les DG métiers d'outils "agiles" pour pallier aux déficiences du SID
- 👉 Contournement de l'IT : entrepôts et infrastructures BI parallèles
- 👉 Mise en péril de la chaîne décisionnelle de l'entreprise

Qu'est qu'un Data Lab ?

Un **DataLab** (ou laboratoire de données) est un **environnement exploratoire analytique** agile qui orchestre des compétences statistiques, métiers et informatiques **dans l'objectif de révéler le sens des données en expérimentant des algorithmes statistiques, de prévision et de modélisation prédictive**. Il intègre d'emblée ou à terme les technologies big data et fait usage d'outils BI novateurs.

Les réponses d'un DataLab

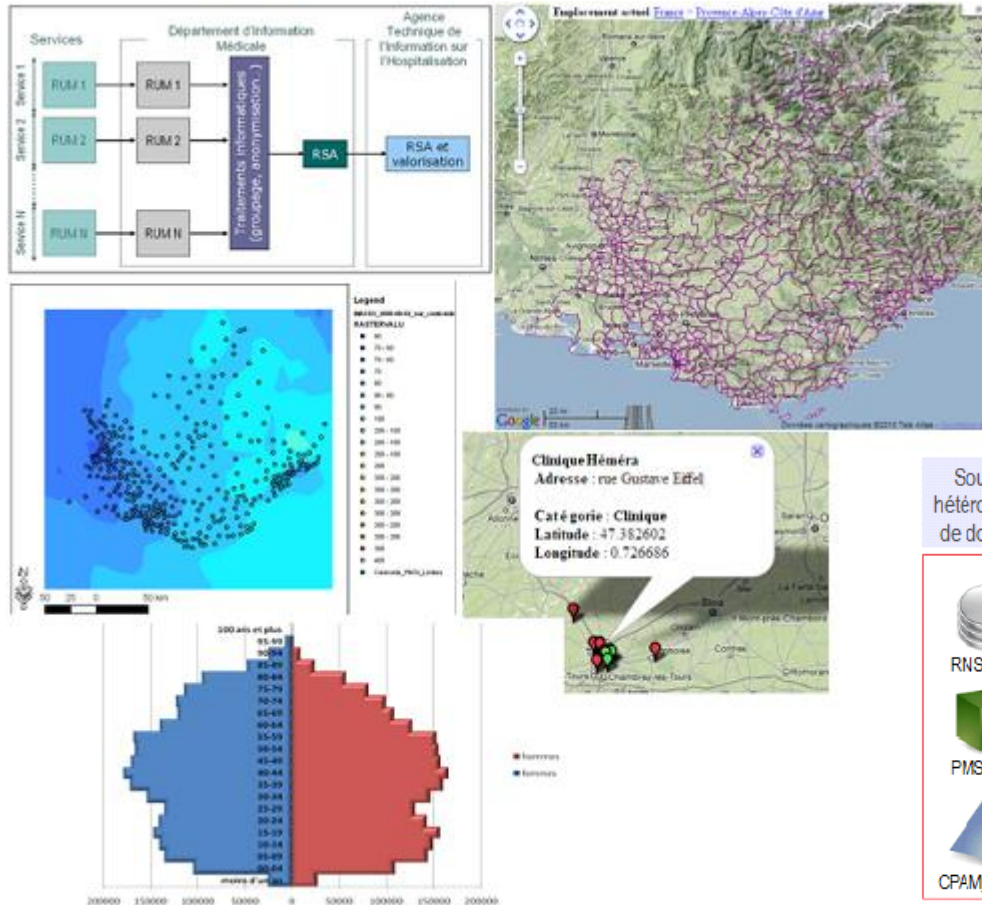
- ☞ Il permet de découvrir, de collecter, de qualifier, de valider des données externes
- ☞ Il est dédié même si à usage ponctuel pour analyses et études à des fins de POC
- ☞ Il est ouvert, accessible pour toute expérimentation tant métier que technique
- ☞ Il est construit en tenant compte des règles de gouvernance incontournables
- ☞ Son architecture doit être très flexible et hautement modulaire

Illustration d'un Data Lab

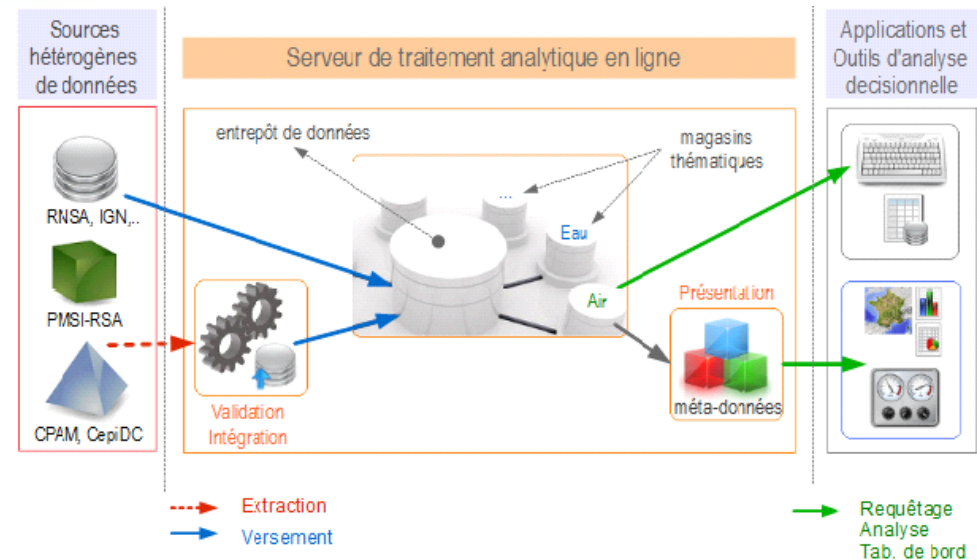
- 👉 Année : **2009**
- 👉 Contexte : mission Air&Santé diligentée par une collectivité et CHU
- 👉 Objectif : Recherche de lien (inférence) entre la densité de séjours hospitaliers et les caractéristiques environnementales choisies sur une zone géographique.
- 👉 IRIS - LabSTIC de l'UNS
 - Médecin - Responsable scientifique
 - Ingénieur biostatisticienne, gestion des données médicaux-économiques
 - Ingénieur géomaticien, gestion des données géographiques
 - Architecte système décisionnel et analytique
- 👉 Méthode : Croisement de données d'origines médico-économiques d'activité, environnementales, géographiques, structurelles, démographiques.

Data Lab : Clef de voûte du système décisionnel ?

Illustration d'un Data Lab



- 👉 Sources et natures variées
- 👉 Brutes, semi-structurées, à formats variables et volumineuses,
- 👉 Couvertures, échelles spatio-temporelles différentes



Data Lab : Clef de voûte du système décisionnel ?

The screenshot shows a web browser window with the URL `sig-air-sante./pentaho/AirSante`. The interface includes a menu bar (Fichier, Édition, Affichage, Historique, Marque-pages, Outils ?), a search bar with Google, and a sidebar with filters for 'Année' (2007) and 'Département' (06). The main map displays a geographical area with a color-coded overlay representing data. A pop-up window titled 'Analyse Territoire vs Département' provides detailed statistics for the selected territory (PMSI 83120) in 2007.

Année 2007 **Département** 06

Niveau de pollution
Polluant: RADON
Seuil min: NaN max: NaN

Densité de séjours
Affections: TUBE DIGESTIF
Seuil min: 0 max: 256
Opacité: 0%

Cliquez sur le territoire **13830**, pour visualiser les données ...

Année d'étude (2007)
Territoire PMSI - Démographie (83120)
Commune(s) : PLAN-DE-LA-TOUR, SAINTE-MAXIME
Superficie : 121 km²
Population : 16 439 (7 877 hommes - 8 562 femmes)
Densité territoriale : 135 hab./km²

Polluant (RADON)
Aucune donnée de pollution disponible ...

Fonction (TUBE DIGESTIF)
Densité de séjours pour 10000 habitants : 28.0
Nb. de séjours : 47 (42.6 % hommes - 57.4 % femmes)
Durée moyenne de séjour : 8.2 jour(s)
Moyenne d'âge : 64.8 ans.

Analyse Territoire vs Département

Fonction : TUBE DIGESTIF
Territoire PMSI : 83120
Année : 2007

Densité* de séjours du territoire : 28.0
Densité* moyenne du département : 164.6
Densité* maximale du département : 256.0
* densité moyenne annuelle pour 10000 habitants

- < à la moyenne départementale
- = à la moyenne départementale +/-5%
- > à la moyenne départementale

Data Lab : Clef de voûte du système décisionnel ?

Illustration d'un Data Lab

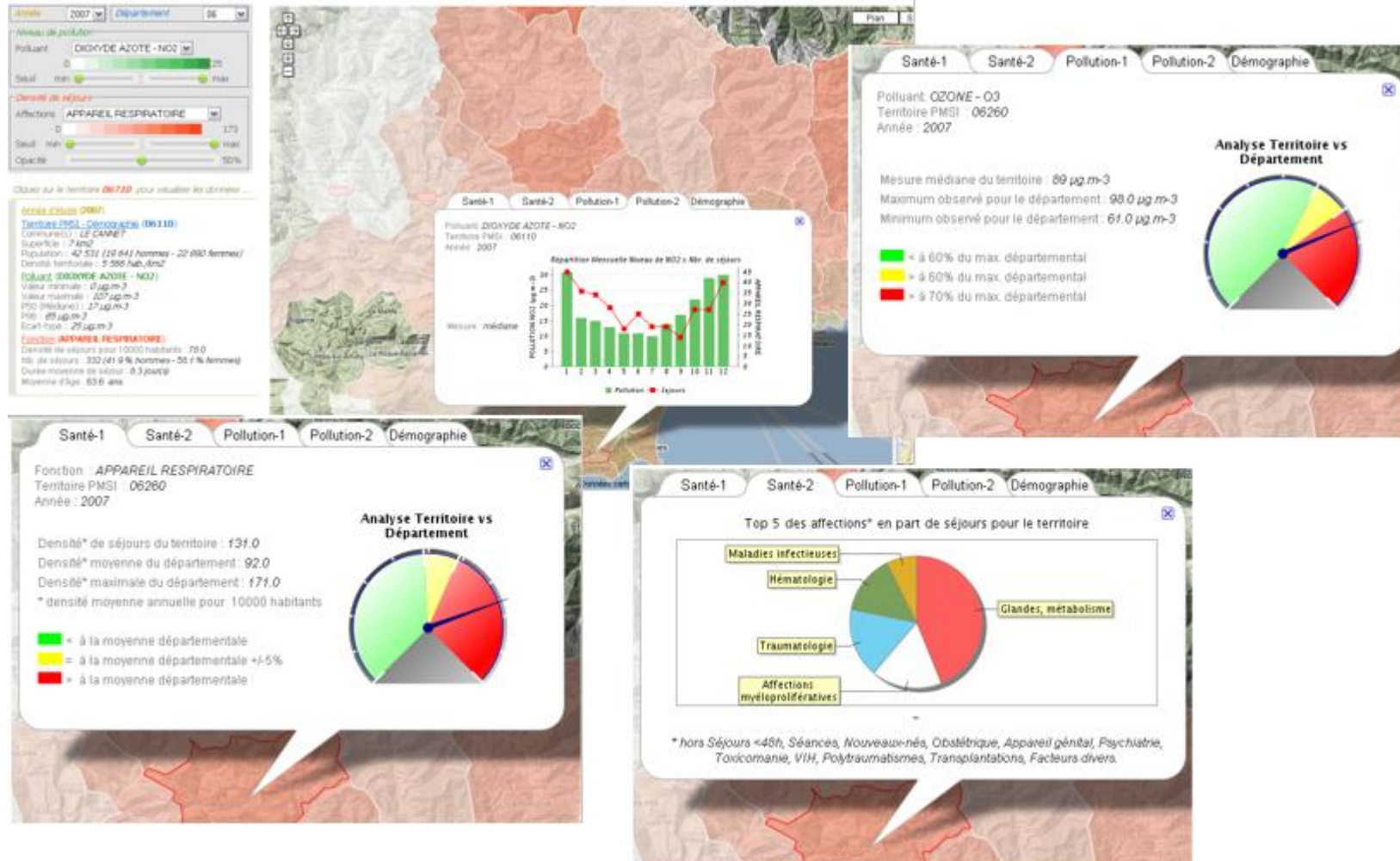


Illustration d'un Data Lab

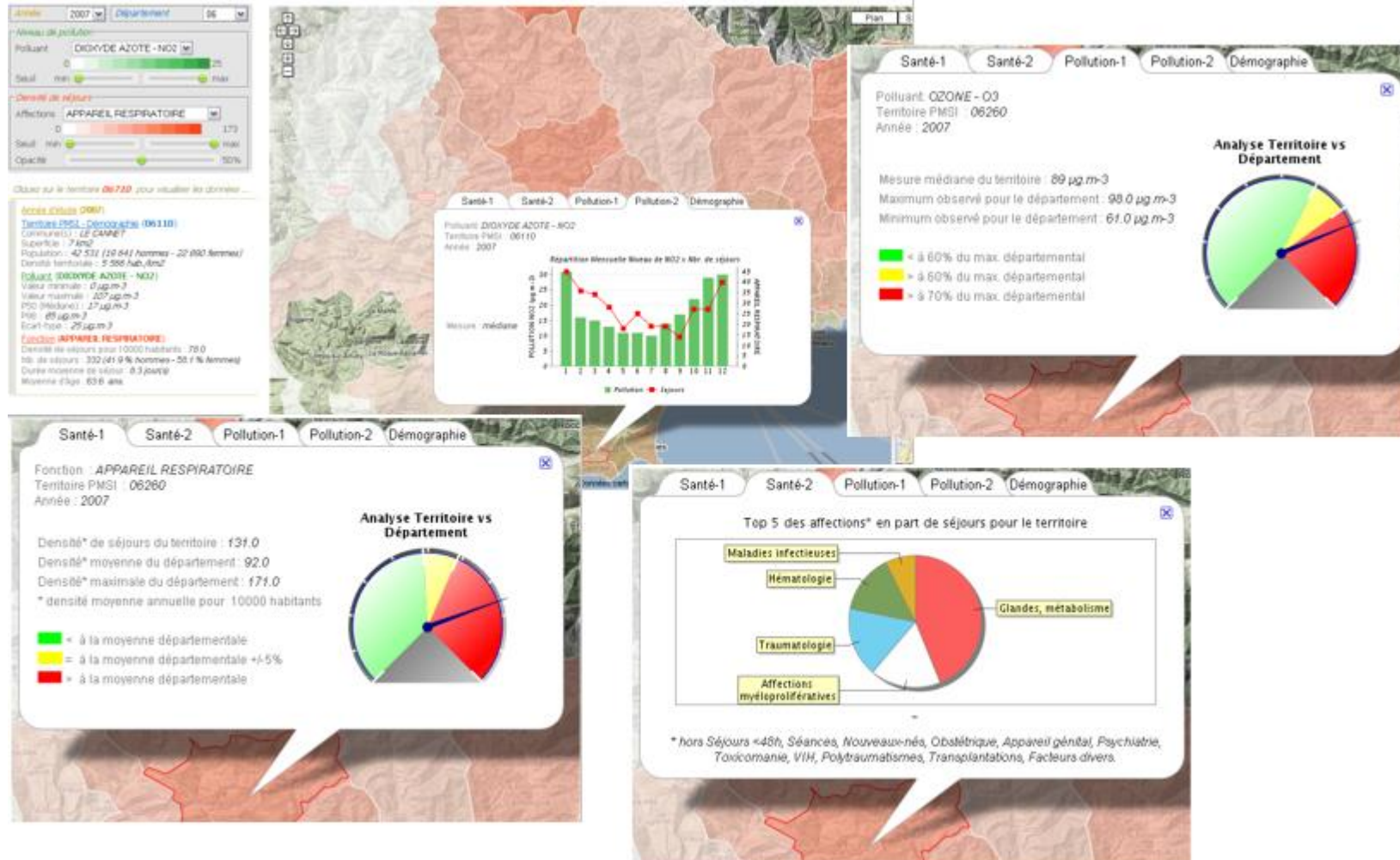


Illustration d'un Data Lab



La plateforme a permis :

- de valider le choix et la pertinence des différentes sources de données
- de collecter, nettoyer et générer de nouvelles données plus adaptées
- d'observer des possibilités de corrélation (en se gardant de toute analyse)

La plateforme a pu être adaptée pour des travaux épidémiologiques des cancers en moins de 6 mois par un stagiaire en Ingénierie de la Santé



Améliorations possibles

- Gain en temps de traitements
- Possibilité d'une observation proche temps-réel grâce à un accès aux données issues des réseaux sociaux et des données environnementales
- Passage de GoogleMaps à OpenStreetMap





Edmond Cissé

*URÆUS Consult
Ingénierie & Sécurité des Données
www.uraeus-consult.com*

edcisse@uraeus-consult.com

04 93 95 52 10

uræus
CONSULT



MERCI !