

# DATA Architecture

PxDATA  
Joël BIZINGRE  
CONIX

[joel.bizingre@conix.fr](mailto:joel.bizingre@conix.fr)

Le 09 07 2014

# Data architecture

- **1ère partie : Data World**
  - **Point de bascule : signaux autour des données**
  - **Tour d'horizon de thèmes Data : Open Data, Data Vizualisation, Big Data, Data policy, « AGL » Data, Data provider, Master Data, MetaData, Data Lake...**
  - **Du Data game à la Data war ... quel sera le monde d'après ?**
  - **Je suis Data Scientist, Data Manager... inventaire des métiers Data**
  
- **2ème partie : Data architecture**
  - **Nouveau paradigme ?**
  - **Intérêt d'un apport méthodologique ?**
    - **L'initiative PxDATA**

# Data architecture : point de bascule

- Il y a quelques années...



**Données**



**Données**



# Data architecture : point de bascule

- **Flot de données**



# Data architecture : point de bascule

- **Flot de données**

700 000 recherches Google

600 nouvelles vidéos Youtube

13 000 applications iPhone téléchargées

100 000 tweets

100 nouveaux comptes LinkedIn

1 500 posts sur les blogs

510 000 commentaires sur Facebook

1 700 publicités ciblées émises par Razorfish

810 visiteurs uniques sur Amazon.fr

16 700 transactions clients - Walmart

**En 60''**



**Données, données, données ... données**

# Explosion des sources et des données

- Flot de données



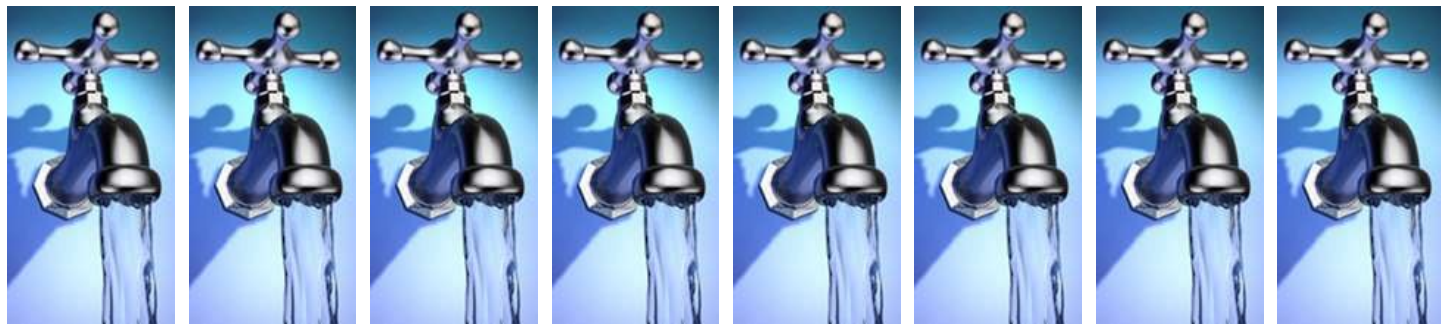
A screenshot of the Treato website homepage. The main heading reads "See what millions of patients are saying". Below it is a search bar with the placeholder text "Search medications" and an orange "Search" button. To the right is a video player showing a cartoon character holding a pill bottle next to a large play button. Below the search bar, it lists "Featured searches: Lipitor, Bupropion vs. Wellbutrin, Prozac". At the bottom, a blue oval highlights four statistics: 190,055,576 patients, 1,769,452,076 posts, 26,627 medications, and 13,850 conditions. The date "Juin 2014" is visible in the bottom right corner of the screenshot.

# Data architecture : point de bascule

- Flot de données



Données  
non  
structurées



Données, données, données ... données

# Data architecture : point de bascule

- **Capture des données**

- 2000 : < 10 capteurs pour une voiture standard
- 2013 : 76 capteurs dernière Ford Hybrid
- Moins de 1\$ le cout d'une puce GPS
- 1 Milliards de smartphones vendus en 2013
- > 10 capteurs embarqués par smartphone



<https://www.indiegogo.com/projects/beast-the-sensor-that-tracks-your-workout-performances-and-improvements>



# Data architecture : point de bascule

- Capture des données



GERIATRICS / MEDICINE

## EarlySense Chair Sensor Monitors Vitals Through The Butt

by EDITORS on Jul 7, 2014 · 1:45 pm



<http://www.medgadget.com/2014/07/earlsense-chair-sensor-monitors-vitals-through-the-butt.html>

# Data architecture : point de bascule

- **Capture des données**

## The smart bracelet that can read your emotions

7 July 2014 Last updated at 16:03 BST

<http://www.bbc.com/news/technology-28196132>



### XOX - The Emotional Technology Platform

Novel processing and ergonomic wearable technology sets the XOX platform apart offering the wearer a 360-degree seamless experience in emotive technology that is turn key and fully integrable.

The XOX platform can be used for stadium and tour, cinema and film, gaming, dating, hospitality, upper class travel, fashion shows, market research as well as boosting shopping and retail experiences. It also has a wider reach in oil safety, health and well being as well as child location and crowd control.

<http://www.xoxemotionaltech.com/>

# Data architecture : point de bascule

- **Capture des données**

Aménagement urbain : où les gens courent ?

Avant :

- Sondage
- Échantillon
- Observateur

Après :

- Capteurs connectés



<http://flowingdata.com/wp-content/uploads/2014/02/Paris.png>

# Data architecture : Data world

- Capture des données

PARISDATA MAIRIE DE PARIS

Les données L'API La licence La démarche Le forum Cartographe

Retour

181 enregistrements

Recherche textuelle

Rechercher...

Arrondissement

- 75015 (18)
- 75018 (18)
- 75011 (14)
- 75012 (13)
- 75005 (12)
- 75020 (11)

Plus

### Liste des cafés à un euro

Informations Tableau Carte Analyse Export API 0 commentaire

Nom	Adresse	Arrondissement	geo_latitude	Prix comptoir
Brasloja	16 rue Garneron	75 018	48.8864467, 2.3274352	1
Café beauveau	9 rue de Miromesnil	75 008	48.8717999, 2.3159991	1
Café Varenne	36 rue de Varenne	75 007	48.8541218, 2.3235382	1
Caprice café	12 avenue Jean Moulin	75 014	48.8259048, 2.3223627	1
Chez Miamophile	6 rue Mélingue	75 019	48.8748714, 2.3860558	1
Denfert café	58 boulevard Saint-Jacques	75 014	48.8341513, 2.3327984	1
Etienne	14 rue Turbigo	75 001	48.8638124, 2.3486623	1
L'antre d'eux	16 rue DE MEZIERES	75 006	48.8503225, 2.3303242	1
L'Assassin	99 rue Jean-Pierre Timbaud	75 011	48.8687437, 2.3799936	1
L'Entracte	place de l'opera	75 002	48.870144, 2.332444	1
L'Inévitable	22 rue Linné	75 005	48.8454475, 2.3547843	1
La Brasserie Gaieté	3 rue de la Gaieté	75 014	48.8396528, 2.3235536	1
La cantine de Zoé	136 rue du Faubourg poissonnière	75 010	48.8806562, 2.3498506	1
La cantoche de Paname	40 Boulevard Beaumarchais	75 011	48.8605653, 2.3673548	0
Le bal du pirate	60 rue des bergers	75 015	48.84255, 2.2808071	1
Le BB (Bouchon des Batignolles)	2 rue Lemercier	75 017	48.8853124, 2.3253739	1
Le Bloc	21 avenue Brochant	75 017	48.8894753, 2.3188696	1
Le Bosquet	46 avenue Bosquet	75 007	48.8559992, 2.3045876	1
Le Brigadier	12 rue Blanche	75 009	48.8775907, 2.3321008	1
Le Brio	216, rue Marcadet	75 018	48.8919384, 2.3333864	1

1 2 3 4 5 6 7 8 >>

# Data architecture : Data world

- Capture des données



Un café, 1 euro

Vous connaissez des cafés ou l'expresso coûte 1€ ou moins ?

\* Rappel

Ben de taguet\*

Adresse

Arrondissement

Prix du café au comptoir

Prix du café en salle

Lien de site internet du café

Votre prénom/nom

Submit

Power by Google Data



Aidez-nous à trouver tous les #cafés parisiens où on peut boire un expresso pour 1€ ou moins, nous en ferons un article [bit.ly /KMo1V0](http://bit.ly/KMo1V0)

↳ Répondre ↻ Retweeter ★ Favori ... Plus

RETWEETS 104 FAVORIS 18



fraxeme

09/07/2014

## Liste des cafés à un euro

Informations Tableau Carte Analyse Export API 0 commentaire

Nom	Adresse	Arrondissement	geo_latitude	Prix comptoir
Brasilloja	16 rue Ganneon	75 018	48.8864467, 2.3274352	1
Café beauveau	9 rue de Miromesnil	75 008	48.8717999, 2.3159991	1
Café Varenne	36 rue de Varenne	75 007	48.8541218, 2.3235382	1
Caprice café	12 avenue Jean Moulin	75 014	48.8259048, 2.3223627	1
Chez Miamophilie	6 rue Mélingue	75 019	48.8748714, 2.3860558	1
Danfert café	58 boulevard Saint-Jacques	75 014	48.8341513, 2.3337964	1
Etienne	14 rue Turbigo	75 001	48.8638124, 2.3486623	1
L'antre d'eux	16 rue DE MEZIERES	75 006	48.8503225, 2.3303242	1
L'Assassin	99 rue Jean-Pierre Timbaud	75 011	48.8687437, 2.3799936	1
L'Entracte	place de l'opera	75 002	48.870144, 2.332444	1
L'Inévitable	22 rue Linné	75 005	48.8454475, 2.3547943	1
La Brasserie Gaité	3 rue de la Gaité	75 014	48.8396528, 2.3235536	1
La cantine de Zoé	136 rue du Faubourg poissonnière	75 010	48.8806562, 2.3489586	1
La canteoche de Paname	40 Boulevard Beaumarchais	75 011	48.8605653, 2.3673548	0
Le bal du pirate	80 rue des bergers	75 015	48.84255, 2.2808071	1
	2 rue Lemercier	75 017	48.8853124, 2.3253739	1
	21 avenue Brochant	75 017	48.8894753, 2.3188696	1
	46 avenue Bosquet	75 007	48.8559992, 2.3045876	1
	12 rue Blanche	75 009	48.8775907, 2.3321008	1
	216, rue Marcadet	75 018	48.8919384, 2.3333864	1

Suivre

1 2 3 4 5 6 7 8 >>



Stratégie de collecte : ici appel aux internautes

# Data architecture : Data world

- Capture des données



Un café, 1 euro

La démarche Le forum Cartographie

Vous connaissez des cafés ou l'expresso coûte 1€ ou moins ?

Ben du taguet\*

Adresse

Arrondissement

Prix du café au comptoir

Prix du café en salle

Lien de site internet du café

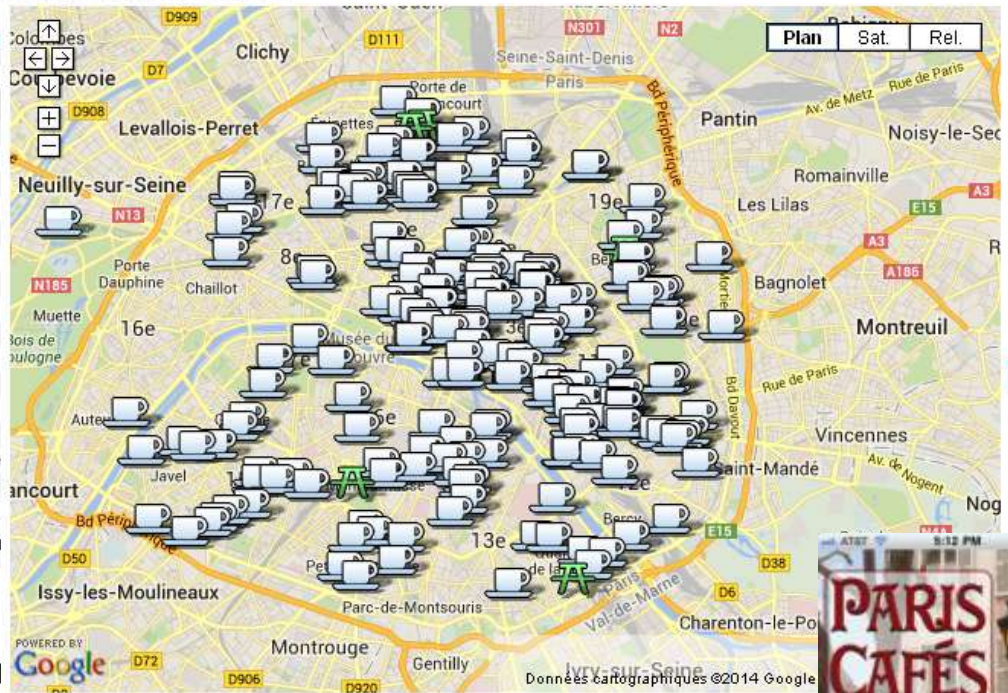
Votre prénom/nom

Submit

Powered by Google Data

## Liste des cafés à un euro

- Informations
- Nom
  - Brasilloja
  - Café beauveau
  - Café Varenne
  - Caprice café
  - Chez Miamophile
  - Danfert café
  - Etienne
  - L'antre d'eux
  - L'Assassin
  - L'Entracte
  - L'Inévitable
  - La Brasserie Gaité
  - La cantine de Zoé
  - La canteoche de Panam
  - Le bal du pirate



Aidez-nous à trouver tous les #cafés parisiens où on peut boire un expresso pour 1€ ou moins, nous en ferons un article [bit.ly/KMo1V0](http://bit.ly/KMo1V0)

↳ Répondre ↻ Retweeter ★ Favori ... Plus

RETWEETS 104 FAVORIS 18



09/07/2014



Service



# Data architecture : point de bascule

- **Numérique**
  - **2000 : 75% de l'information est encore analogique vs numérique**
  - **2013 : 2%...**

# Data architecture : point de bascule

- **Capacités technologiques : traitement et stockage**
  - **VISA : 1 mois pour traiter 65 Milliards de transactions -> 13'**
  - **Prix de stockage divisé par 2 tous les 2 ans**

## Humanity Passes 1 Zettabyte Mark in 2010

A zettabyte is 1,000,000,000,000,000,000 bytes (that's 21 zeroes for those counting), or one trillion gigabytes. That's enough data to fill 75 billion 16-gigabyte-sized iPads.



Les échanges de données via Internet devraient dépasser le Zettabyte en 2016

# X 2 tous les 2 ans

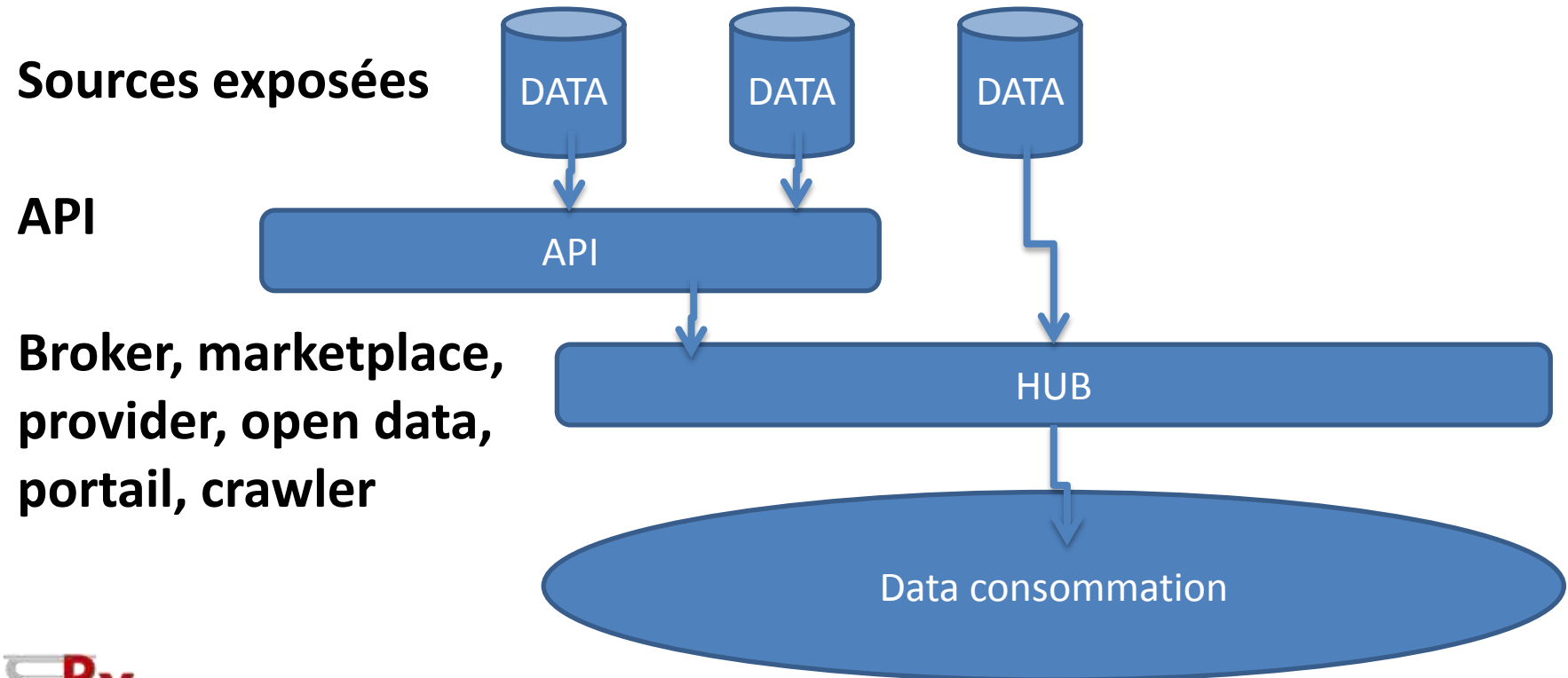
(RAM 4K, DD 64K  
pour la mission Apollo)





# Data architecture : point de bascule

- **Le web s'empare des données ... des milliers de ressources disponibles ... qui en vivent ... Marché de la donnée**
  - La donnée comme matière première
  - La donnée comme service



# Data architecture : point de bascule

- Ressources web

Integrate health data from anywhere, in minutes.

Human API gives you a single, simple connection to all your user's health data from any vendor or data source.

[Get started for free](#) [View data sources](#)

humanapi.co

Human API / Documentation

- ✓ Start here
- API Guide
- Main API Resources
- App API Resources
- Reference
- API Explorer

The Human resource is a summary object for a person's latest health measurements and summary information about the person.

Detailed description of the Human attributes

Property	Type	Description
userid	String	The Human API id.
gender	String	The gender of the person, if known.
bloodGlucose	BloodGlucose	The latest blood glucose measurement.
bloodPressure	BloodPressure	The latest blood pressure measurement.
bmi	Bmi	The latest body mass index measurement.
bodyFat	BodyFat	The latest body fat measurement.
height	Height	The latest height measurement.

GET	/v1/human/blood_glucose	Returns the most recent blood glucose reading
GET	/v1/human/blood_glucose/readings	Returns a list of blood glucose readings
GET	/v1/human/blood_glucose/readings/{id}	Returns a specific blood glucose reading
GET	/v1/human/blood_glucose/readings/daily/{date}	Returns blood glucose readings for a specific date
GET	/v1/human/blood_oxygen	Returns the most recent blood oxygen reading
GET	/v1/human/blood_oxygen/readings	Returns a list of blood oxygen readings
GET	/v1/human/blood_oxygen/readings/{id}	Returns a specific blood oxygen reading
GET	/v1/human/blood_oxygen/readings/daily/{date}	Returns blood oxygen readings for a specific date
GET	/v1/human/blood_pressure	Returns the most recent blood pressure reading



# Data architecture : point de bascule

- **Culture de la données (Data World)**
  - **Vivre dans un monde de données : nos ombres numériques**
    - Données de consommation
    - Données de communication
    - Données sociales
    - Données professionnelles
    - Données de santé
    - Nos données de déplacement et de localisation
    - Données de scolarité
    - Nos données de projet
    - Nos données de sentiment
    - Données d'activité physique, de jeu
    - ...

Et plus encore – courant extrême : Quantified self

# Data architecture : point de bascule

- **Culture de la données (Data World)**

*Qu'avez-vous fait le matin de Noël ?*

*Quelles étaient vos émotions à votre RDV avec votre banquier ? ...*



**Datafication** is a modern technological trend turning many aspects of our life into computerised data [\[1\]](#) and transforming this information into new forms of value. [\[2\]](#) Examples of datafication are how [Twitter](#) datafies stray thoughts or datafication of [HR](#) by [LinkedIn](#) and others

wikipedia

<http://cestpasmonidee.blogspot.fr/2013/09/bnz-analyse-votre-attitude-face-largent.html>  
09/07/2014

# Data architecture : point de bascule

- Flot de données
- « Mise en données » - capture
- Numérique
- Traitement et stockage
- Données comme ressources
- Culture de la données



# Data architecture : Data world

- **Nouvelles technologies**
  - **Nouvelles problématiques**
  - **Nouveaux métiers**
  - **Nouvelles offres de service**
  - **Nouveaux concepts**
  - **Nouvelles batailles**
  - **Nouveaux jeux**
  - **...**
- 
- **Mais aussi d'anciennes problématique, technologies, d'anciens concepts...**

# Data architecture : Data world

## Termes ... Data

- Data sourcing
- Open Data
- Data management
- Big Data
- Data as a service
- SI Data centric socle/puit de données
- Data analysis
- Référentiels de données
- Data quality
- Data security
- Echanges de données – Data exchange
- Data vizualisation
- Data storytelling
- Data policy
- Smart Data
- Data virtualization
- Master Data
- Data scraping
- Data gouvernance
- AGL Data
- Data Lake
- Data provider
- Data Hub
- Data grid
- Unstructured Data
- Data API
- Data crawler
- Data broker
- Data marketplace
- Data cleansing
- Data science
- Data warehouse
- Metadata
- Linked Data
- Data portal
- Enterprise Data Architecture
- All Data ...

# Data architecture : Data world

- Publier des données (Producteurs) : Open Data, DaaS
- Exemple :

The screenshot shows the Antidot website interface. At the top left is the logo 'antidot' and at the top right is the text 'Monuments historiques'. Below this is a search bar with the word 'Recherche' and a magnifying glass icon. On the left side, there is a sidebar with filters under the heading 'Affiner'. The filters include: 'Photos du monument?' with 'Non' (249) and 'Oui' (246) options; 'Région / Département / Ville' with 'Île-De-France' (495) and 'Paris' (495) selected; 'Type de monument' with 'Monument' (495) selected; and several other categories like 'Période historique', 'Gares les plus proches', 'Stations de métro', 'Type de propriété', and 'Recherche géographique' with input fields for latitude, longitude, and distance. The main content area is titled 'Résultats de la recherche' and shows '495 résultats (54ms)'. It includes sorting options ('Trier par Pertinence') and display options ('Afficher 10 résultats par page'). The first result is 'Pavillon de l'Arsenal, actuellement bibliothèque de l'Arsenal', featuring a medieval manuscript illumination and a description: 'Type : Monument', 'Adresse : 1, 3 rue Sully', 'Localité : Paris 04 (Paris)', and 'Liens externes : Base Mérimée - Wikimedia Commons'. The second result is 'Domaine national du Palais-Royal (ancien Palais Cardinal), comprenant Conseil d'Etat, Conseil Constitutionnel, Ministère de la Culture, théâtre de la Comédie Française', featuring a photograph of the building and a description: 'Type : Monument', 'Adresse : place du Palais-Royal', 'Localité : Paris 01 (Paris)'. At the bottom left is the 'Paxeme' logo, and at the bottom right is the 'CONIX' logo.



# Data architecture : Data world

- Open Data
  - Exemple :

Notre application « **Monuments historiques** » a été réalisée en exploitant 7 sources de données ouvertes :

1. la liste des Immeubles protégés au titre des Monuments Historiques disponible sur [data.gouv.fr](http://data.gouv.fr). Cette source de données décrit **43.720 monuments** dans un **fichier CSV**.
2. la liste des **gares de voyageurs du Réseau Ferré National** avec leurs coordonnées telle que fournie par [data.gouv.fr](http://data.gouv.fr). Cette source de données décrit **3.065 gares** dans un **fichier XLS**. Elle est exploitée pour situer les monuments à proximité d'une gare.
3. la liste des stations du métro parisien avec leurs coordonnées, fournie par [OpenStreetMap](http://OpenStreetMap). Cette source de données décrit **301 stations** et elle est exploitée pour situer les monuments à proximité d'une station de métro.
4. les données du code officiel géographique (COG) de **l'INSEE**. Cette source de données décrit **22 régions, 99 départements, plus de 4.000 cantons et chefs lieux** dans un **graphe RDF**.
5. Les photos des **monuments historiques de Wikipedia** proposée par Wikimedia Commons. Cette source de données, notamment alimentée par le concours **Wiki loves monuments**, apporte **122.828 photos** pour **12.586 monuments** historiques désignés par leur code PA : il s'agit d'un code délivré de façon unique pour chaque monument et présent dans la liste citée en 1.
6. La description des **monuments historiques de Wikipedia** fournie par **DBpedia**. Cette source de données **en RDF** décrit **3,64 millions d'objets**, dont **413.000 lieux**. Cette source est accessible directement à partir des informations de Wikimedia Commons
7. Les informations de géolocalisation de Yahoo! via [Yahoo! PlaceFinder](http://Yahoo! PlaceFinder). Cette source permet de géolocaliser à partir de leur adresse les monuments non géolocalisés dans Wikimedia Commons ou DBpedia



## Liste des Immeubles protégés au titre des Monuments Historiques

Ce jeu de données présente un service public certifié  
Publié le 7 novembre 2013 par Ministère de la Culture et de la Communication

Liste des immeubles protégés au titre des Monuments Historiques (toute France en 2012 - 44 236 monuments)  
Liberté des champs : REF (référence de référence dans la base Mérimée) ; ETUD (type d'étude : reconnaissance immeubles SH, avec éventuellement la mention sabb Xae) ; LOCA OU REG + DEPT + COM (localisation de l'immeuble) ; INSEE (code INSEE) ; TICO (appellation courante de l'immeuble) ; ADRES (adresse, n° de la voie, type de voie, nom de la voie) ; STAT (statut propriété : type de propriétaire par catégorie - pas de nom) ; AFFE (affectataire : utile pour les propriétés de l'Etat, mentionne le nom du ministre affectataire) ; PPRAD (procédure sur la protection : article 1er de l'arrêté de protection + cadastre + nature et date de la protection) ; DPRO (date de protection) ; AUTR (architecture ou maître d'oeuvre) ; SCLT (période de construction).

### Ressources

CSV Liste des Immeubles protégés au titre des Monuments Historiques - Champagne-Ardenne  
Liste des immeubles protégés au titre des Monuments Historiques au 31.12.2012 pour le département Champagne-Ardenne (toute France) - 44 236



<http://blog.antidot.net/tag/sncf/>

# Data architecture : Data world

- Open Data : Exemple

**RENNES METROPOLE EN ACCES LIBRE**

18143 jeux de données  
participants au concours  
8539 appels aux API

ACCUEIL | **LES DONNÉES** | VOS APPLICATIONS | VOS IDÉES | NOTRE DÉMARCHE | LE CONCOURS | ESPACE DÉVELOPPEUR

Accueil / Les Données / Catalogue

## CATALOGUE

**DONNEE** (nom, féminin) : représentation conventionnelle d'une information sous une forme permettant d'en faire un traitement automatique et qui sert de base à un raisonnement ou une recherche

**Rechercher par mots clés**

**Rechercher par type de flux**

<input type="checkbox"/> TIFF (CC48)	<input type="checkbox"/> TXT	<input type="checkbox"/>
<input type="checkbox"/> JPG (CC48)	<input type="checkbox"/> DXF/DWG (WGS84)	<input type="checkbox"/>
<input type="checkbox"/> DXF/DWG (L93)	<input type="checkbox"/> DXF/DWG (CC48)	<input type="checkbox"/>
<input type="checkbox"/> SHP (WGS84)	<input type="checkbox"/> SHP (L93)	<input type="checkbox"/>
<input type="checkbox"/> SHP (CC48)	<input type="checkbox"/> GTFS	<input type="checkbox"/>
<input type="checkbox"/> XLS	<input type="checkbox"/> API	<input type="checkbox"/>
<input type="checkbox"/> ICS	<input type="checkbox"/> XML	<input type="checkbox"/>
<input type="checkbox"/> CSV	<input type="checkbox"/> KML	<input type="checkbox"/>

**Rechercher par catégorie**

- Accessibilité
- Citoyenneté
- Culture
- Culture : Agenda
- Culture : Annuaire
- Culture : Statistiques
- Données budgétaires
- Environnement
- Equipements
- Logement
- Référentiel géographique
- Sports et loisirs
- Stationnement
- Transports

**Rechercher par éditeur**

- Arts vivants en Ile-et-Vilaine
- Association Trans Musicales
- Bibliothèques de Rennes
- Infocolale/Ouest-France
- INSEE
- Keolis Rennes
- Rennes Métropole
- Service SIG de Saint-Jacques de la Lande
- Service SIG Rennes Métropole
- Ville de Rennes

Remarque : tous les marchés de la ville de Paris comportent maintenant une clause Open Data

# Data architecture : Data world

- Open Data

Partie prenante  
Data journalisme

Lecture  
(Croisement)  
Interprétation  
*sens*

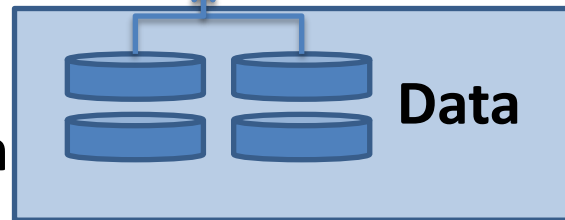
Organisme  
Data provider

Lecture  
(Croisement)  
Exploitation  
*service*

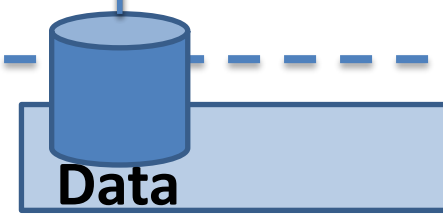
Open Data

Data policy

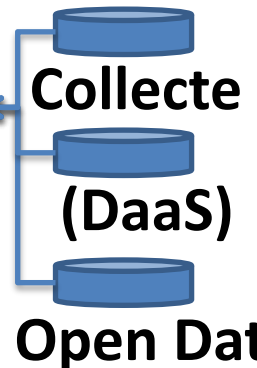
Systeme  
d'Information



Organisme

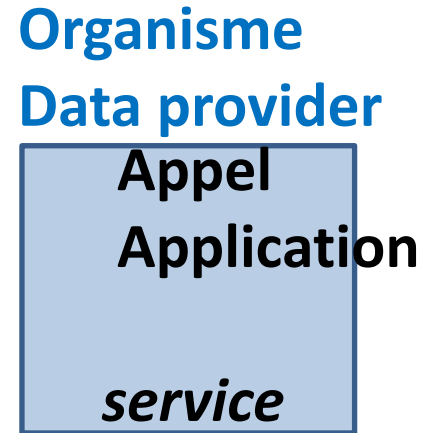
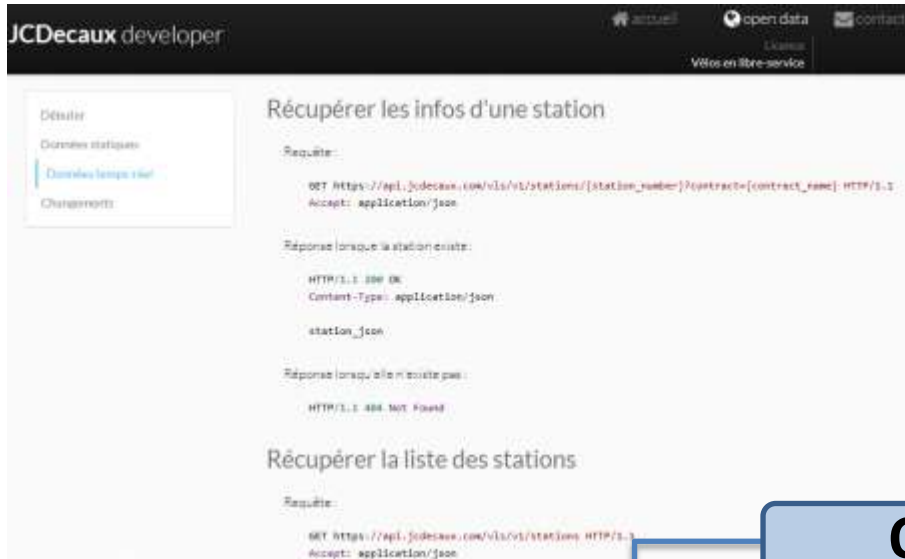


Data collecteur  
Data provider



# Data architecture : Data world

- Open Data ... Open API

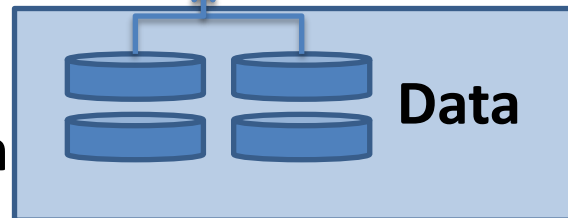


Open API

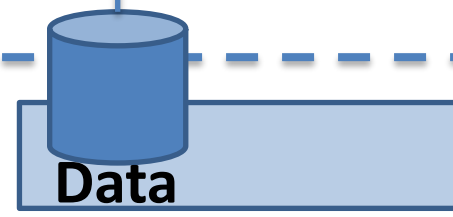
Open Data

Data policy

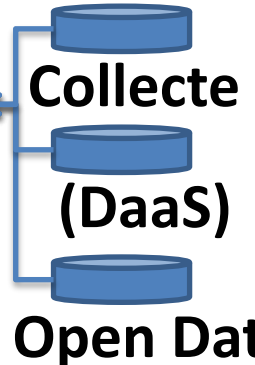
Systeme d'Information



Organisme



Data collecteur Data provider



Open Data

# Data architecture : Data world

- Publier des données (Producteurs) : Open Data

**Droits d'usage ?**

**Où sont les moyens ?**

**Open washing**

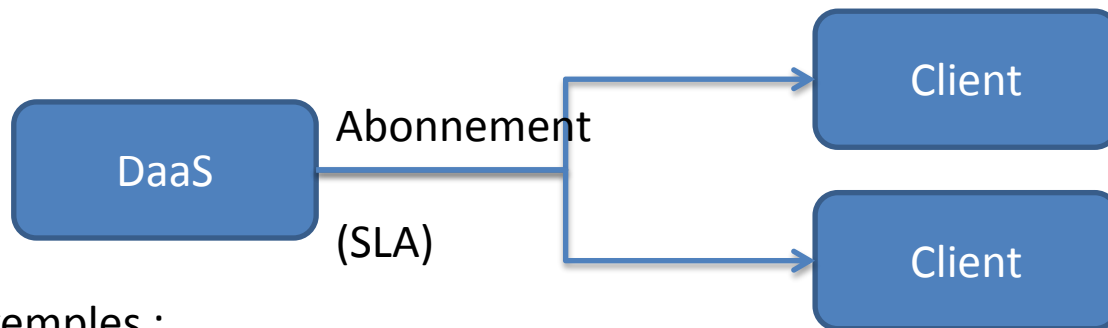
**Après l'enthousiasme la prudence**

**Détournements ... craintes ... risques**

**Cycle de fraîcheur de l'information ?**

# Data architecture : Data world

- **Data as a service (Data provider)**
  - **Commercialisation de données au travers du Cloud**
  - **Paiement à la consommation des données (en fonction de la quantité, de l'usage, de la nature des données, du mode de consommation – tps réel ...)**



Enrichissement de données  
Qualité des données  
Analyse

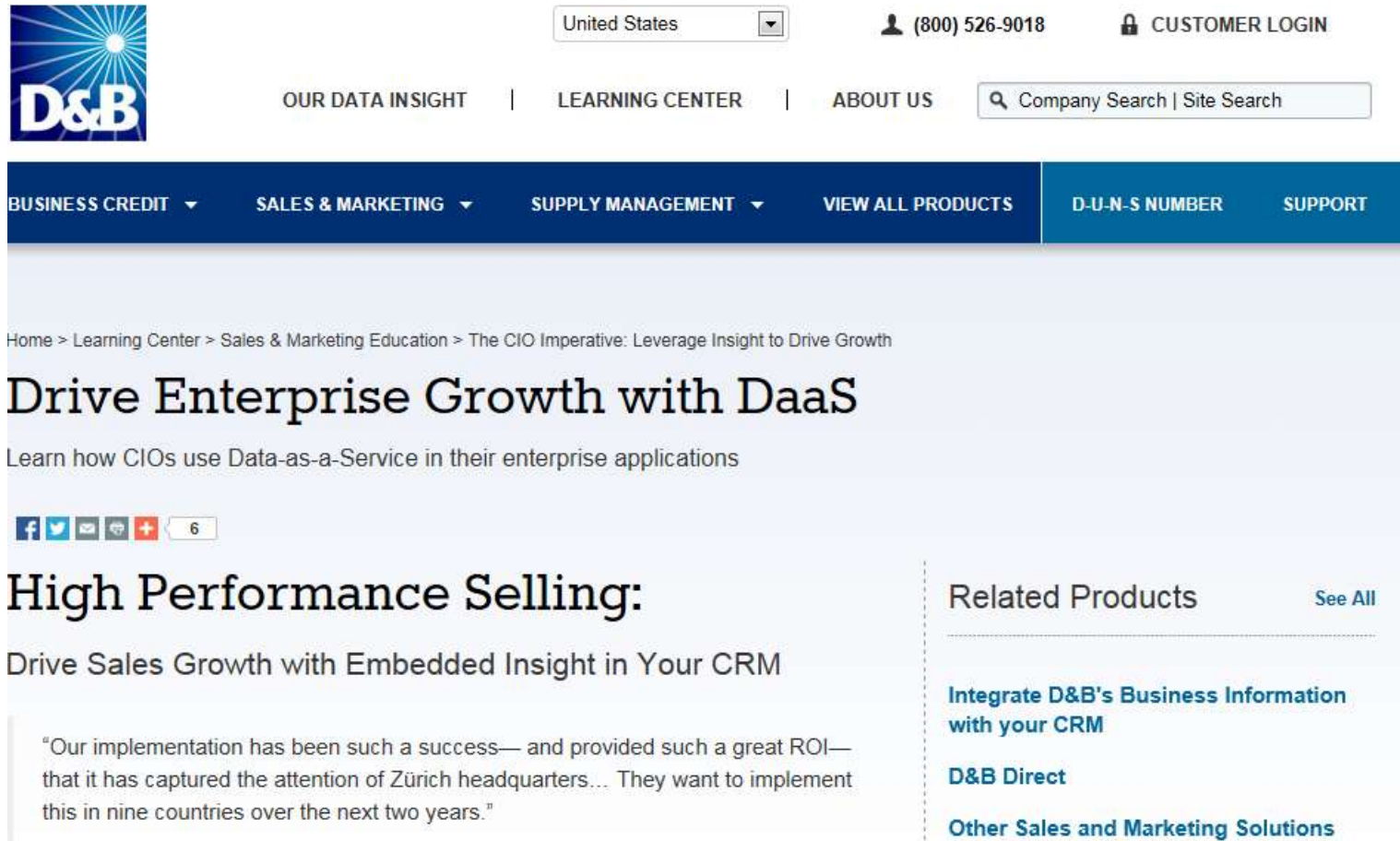
Exemples :

-Adresses

-Données sur les entreprises

# Data architecture : Data world

- Data as a service (Data provider)



The screenshot shows the Dun & Bradstreet (D&B) website. At the top left is the D&B logo. To its right is a dropdown menu for "United States". Further right are a phone icon with "(800) 526-9018" and a "CUSTOMER LOGIN" link. Below these are navigation links: "OUR DATA INSIGHT", "LEARNING CENTER", and "ABOUT US". A search bar contains "Company Search | Site Search". A dark blue navigation bar below has links for "BUSINESS CREDIT", "SALES & MARKETING", "SUPPLY MANAGEMENT", "VIEW ALL PRODUCTS", "D-U-N-S NUMBER", and "SUPPORT". The main content area has a breadcrumb trail: "Home > Learning Center > Sales & Marketing Education > The CIO Imperative: Leverage Insight to Drive Growth". The main heading is "Drive Enterprise Growth with DaaS". Below it is the sub-heading "Learn how CIOs use Data-as-a-Service in their enterprise applications". There are social media sharing icons and a "6" in a box. The main text reads "High Performance Selling: Drive Sales Growth with Embedded Insight in Your CRM". A quote follows: "Our implementation has been such a success—and provided such a great ROI—that it has captured the attention of Zürich headquarters... They want to implement this in nine countries over the next two years." To the right, a "Related Products" section includes "Integrate D&B's Business Information with your CRM", "D&B Direct", and "Other Sales and Marketing Solutions".

<http://www.dnb.com/lc/sales-marketing-education/daas-data-as-a-service-enterprise-use-case.html>

Dun & Bradstreet

# Data architecture : Data world

- **Marché de la données (Datamédiaire) : Data Hub**
  - **Data sourcing**
  - **Data scrapping**
  - **Data provider**
  - **Data marketplace**
  - **Data crawler**
  - **Data broker**
  - **...**



# Data architecture : Data world

- Data provider

**Government**  
Procurement Contracts  
Legislation and Voting Records  
Property and Real Asset Ownership  
Government Employee Salaries

**Public Records**  
Corporate Registrations  
FCC Licenses  
Granted Patents  
Liquor Licenses  
Professional Licenses

**Politics**  
Campaign Contributions  
Lobbying Records  
Political Organizations  
Political Relationships

**Economics**  
Bureau of Labor Statistics  
Global Economic Indicators  
National Labor Market Activity  
Sales Tax Receipts

**Financial**  
FDIC Institutions  
IRS Form 5500 Filings  
SEC Filings  
UCC Liens

**Commerce**  
Commodity Production and Trade  
Farm Subsidies  
Import Bills of Lading  
Trade Finance

**Energy**  
Electricity Transactions  
Oil and Gas Production by Well  
Pipelines and Storage Capacities  
Power Plant Locations and Sizes  
State Energy Consumption

**Healthcare**  
Adverse Event Reporting System  
Clinics and Look-a-likes Registry  
Healthcare Providers  
Medicare Reimbursements

**Real Estate**  
American Housing Survey  
Building Permits  
Foreclosures  
Home Mortgage Origination  
Property Tax Assessments

**Demographics**  
American Community Survey  
National Health Interview Survey  
National Longitudinal Survey  
U.S. Census

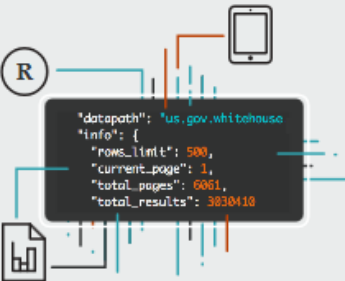
**Transportation**  
Air Carrier Financials  
Aviation Incidents  
FAA Flight Tracking  
Mass Transit Authorities  
Merchant Vessel Registration

**Environment**  
Daily Weather Data  
Hourly Pollutants  
Toxic Release Inventory

<http://enigma.io>

# Data architecture : Data world

- Data provider




```
'datapath': 'us.gov.whitehouse'
'info': {
  'poms_limit': 500,
  'current_page': 1,
  'total_pages': 6961,
  'total_results': 3830418
}
```

### PUBLIC DATA API


Enigma's API provides an infrastructure for developers to power applications and data-rich services, at any scale. Maintain real-time applications with direct access to billions of records.

---




#### Plug Directly Into Enigma

RESTful APIs provide access to the full range of Enigma's datasets and analytics.




#### Build Applications

Power internal models and external products with a minimized backend through on-the-fly data calls.




#### Augment your data

Gain a richer understanding of your data by placing it in context with relevant public datasets.




#### Data as a Service

Let Enigma do the heavy lifting: no more scraping, cleansing or dealing with unreliable sources.



#### Built for Performance

Enterprise-class reliability and scalability, from weekend hack-day to global deployment.




#### Analytics Engine

Analyze massive datasets on demand.

<http://enigma.io>

# Data architecture : Data world


- **Data provider**



## DATA SERVICES


Leverage Enigma's expertise to extend your organization's information awareness and decision-making capabilities.

---




### Entity Resolution

Identifying companies and individuals across an array of datasets unlocks connections and trends.




### Geocoding

Geocoding surfaces insight through physical proximity and geographical associations.




### Custom Data Acquisition

Whether it's going courthouse to courthouse, or lobbying agencies on your behalf, we will get you the data you need.




### Behind the Firewall

Leverage the value of public data while maintaining the highest standards of data security with a behind the firewall deployment.



### Time Series and Join Analysis

Large scale dataset joins enable cross-vertical analysis while time-series tools reduce the complexity of multi-period tables.











### Augmented Research Tools

Never miss a critical piece of information. Uncover entities and connections in unstructured text.

<http://enigma.io>

# Data architecture : Data world

- Data provider

PRICING PLANS				
	 <b>Public</b> Free	 <b>Professional</b> \$195 <sup>00</sup> /month	 <b>Developer</b> \$395 <sup>00</sup> /month	 <b>Enterprise</b> Flexible pricing
Public Data Search	✓	✓	✓	✓
Quick Analytics	✓	✓	✓	✓
1-Click Data Export	Limited	✓	✓	✓
Online Support	Email	Live chat	Dedicated support	White glove service
API Access	10k calls/month	20k calls/month	50k calls/month	Custom API solutions
Custom Data Acquisition	×	×	×	✓
				
	For non-commercial & open source use	For commercial use	For developers & data scientists	For teams, organizations, & commercial apps <a href="http://enigma.io">http://enigma.io</a>

# Data architecture : Data world

- **Data provider**

Source Data Publica



## EXPLORER LES DONNÉES

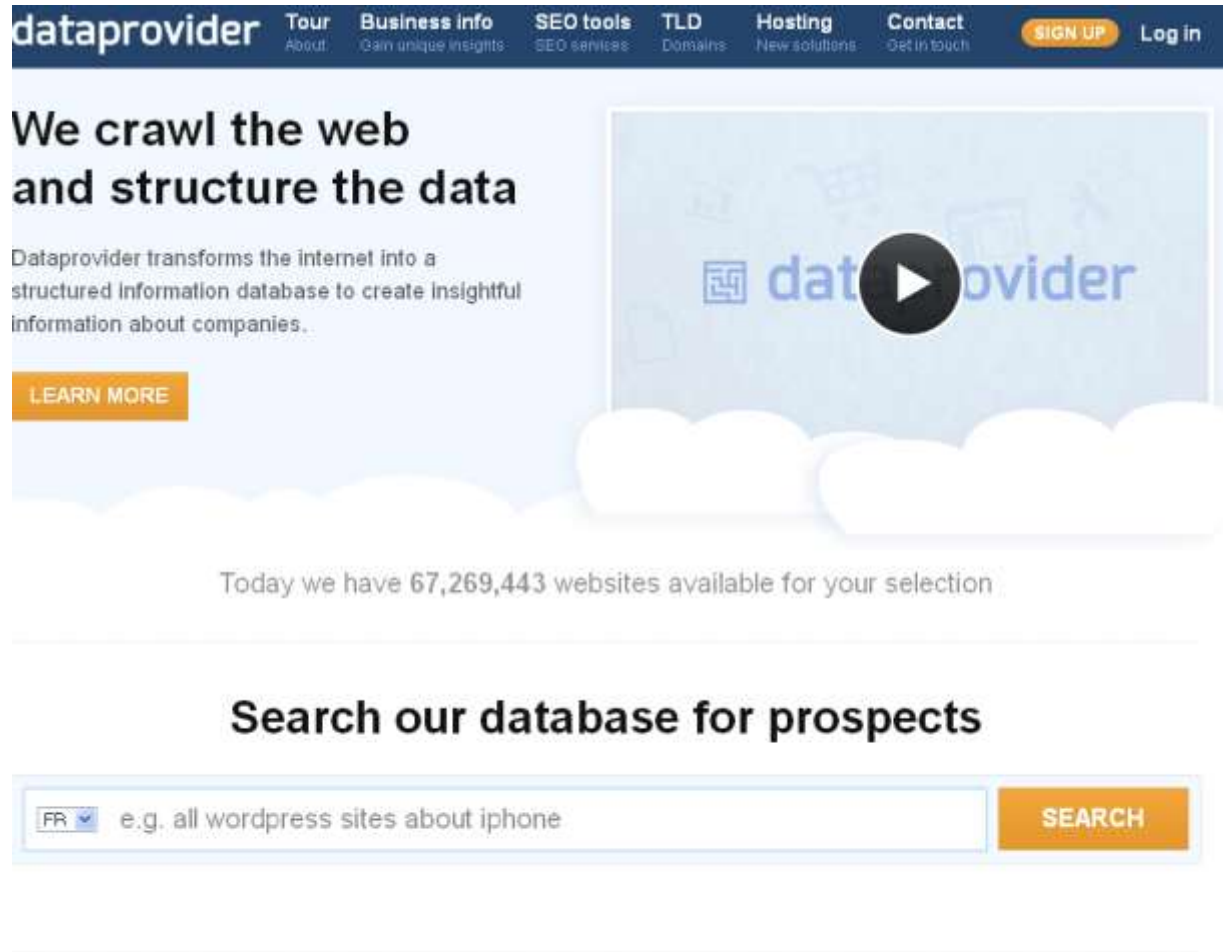
Economie & Finances (6341)	➤	Travail & Salaires (1913)	➤	Environnement & Energie (1680)	➤	Transport & Logistique (1261)	➤
Education & Formation (1215)	➤	Population & Démographie (1153)	➤	Agriculture & Pêche (1051)	➤	Santé & Nutrition (1033)	➤
Sociétés & Conditions de vie (926)	➤	Collectivités & Elections (781)	➤	Commerce & Services (778)	➤	Industrie & Production (616)	➤
Immobilier & Construction (479)	➤	Sciences & Technologies (423)	➤	Autres (419)	➤	Géographie & Espace (380)	➤
Sécurité & Justice (336)	➤	Sports & Loisirs (335)	➤	Informatique & Communications (318)	➤	Tourisme & Voyages (292)	➤
Arts & Culture (291)	➤	Banque & Assurance (130)	➤	Climat & Météorologie (50)	➤	Presse & Médias (34)	➤

Trouver des données publiques



# Data architecture : Data world

- Data provider / crawler



The screenshot shows the homepage of Dataprovider. At the top is a dark blue navigation bar with the following links: Tour (About), Business info (Gain unique insights), SEO tools (SEO services), TLD (Domains), Hosting (New solutions), and Contact (Get in touch). There are also buttons for SIGN UP and Log in. The main content area features a large heading: "We crawl the web and structure the data". Below this is a sub-heading: "Dataprovider transforms the internet into a structured information database to create insightful information about companies." and a "LEARN MORE" button. To the right is a video player with a play button and the text "dataprovider". Below the video player, it states: "Today we have 67,269,443 websites available for your selection". Further down is a section titled "Search our database for prospects" with a search input field containing "FR" and "e.g. all wordpress sites about iphone", and a "SEARCH" button.

Or build a data set in a few steps!

<https://www.dataprovider.com/en/>

# Data architecture : Data world

- « AGL » Data

- IT Human Middleware

# Data architecture

- **Data IT Human Middleware**

## AUTOMATE DATA STEWARDSHIP

Moving from excel and word to an integrated data stewardship platform will increase your efficiency and allow you to monitor your progress and ROI.  
Make use of the Collibra out of the box operating model and data governance workflows.



### MANAGER

Define the Governance Operating Model.  
Monitor KPIs to drive progress.



### BUSINESS STEWARD

Define, document, collaborate.  
Review, approve, publish.



### IT STEWARD

Implement & automate workflows.  
Resolve data issues.

Or **view all capabilities** →

<https://www.collibra.com/solution/capabilities/>



# Data architecture : Data world

- « AGL » Data

Extraire  
Filtrer  
Transformer  
Contrôler  
Tester  
Fusionner  
Enrichir  
Superviser  
Versionner  
Tracer  
Injecter  
Publier

Organisme  
Data provider

Lecture  
Exploitation  
Application  
*service*

AGL Data

Open Data

Data policy

Systeme  
d'Information

Data

Organisme

Data

Data collecteur  
Data provider

Collecte  
(DaaS)

Open Data

# Data architecture : Data world

- AGL Data



## Préparer les données

- Collecte multi-sources
- Extraction multi-formats
- Gestion automatique des types
- Normalisation et enrichissement des données



## Réutiliser les données

- API complète et sécurisée
- Publication dans des formats standards
- Datavisualisations interactives
- Extensions web et open source

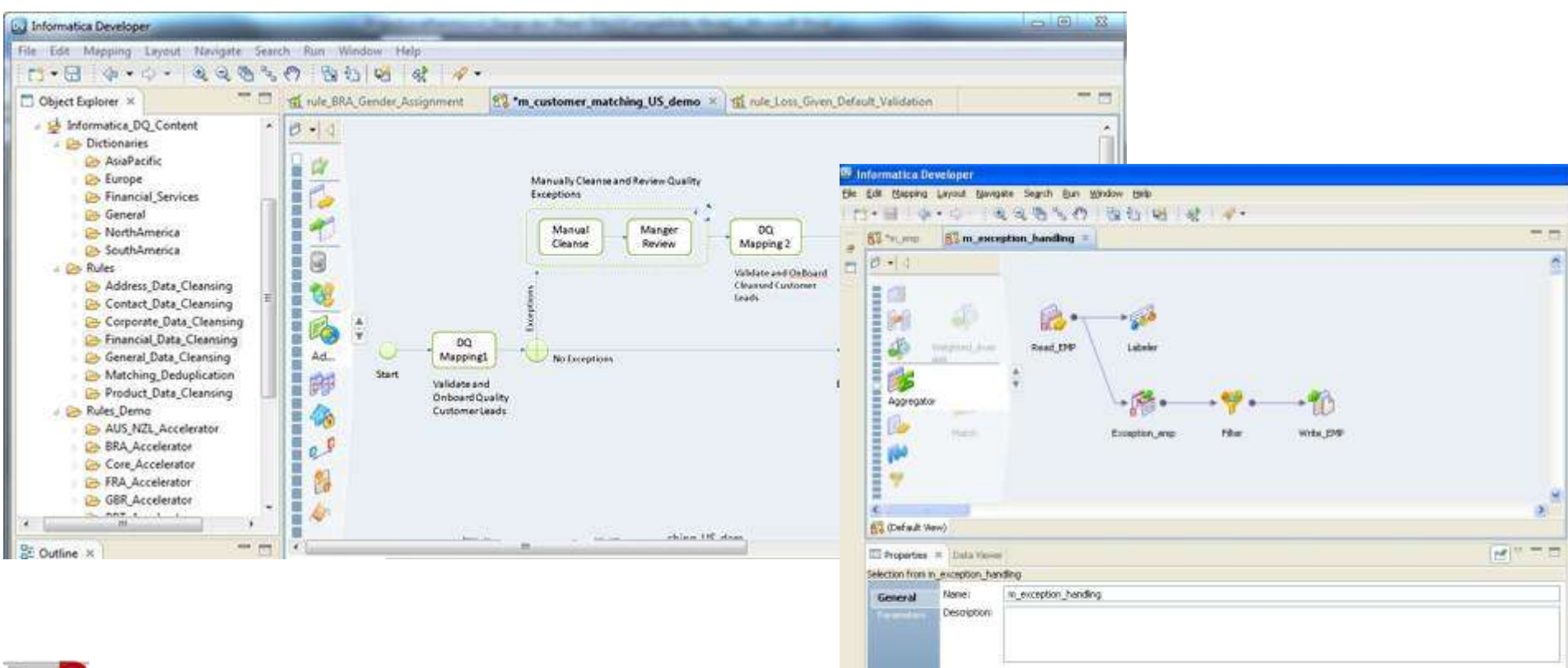


## Exploiter les services

- Technologies big data sur le cloud
- Monitoring des API
- Reporting de consommations
- Pas de limite de volumes + gestion temps réel

# Data architecture : Data world

- Data quality ... AGL Data : analyse des données, profilage, éditeur de règles de contrôle, traçabilité des données, dédoublonnage, data cleansing, mise en place de chaîne de traitements...



# Data architecture : Data world

- « AGL Data » ... publication



The screenshot shows the CKAN website interface. At the top, the CKAN logo and navigation menu are visible. The main content area features a browser window displaying the dataset page for 'UK: Adur District Council Spending'. The page title is 'UK: Adur Data' and the subtitle is 'Data and Resources'. The text describes CKAN as a powerful data management system that makes data accessible by providing tools to streamline publishing, sharing, finding, and using data. Below the text are three buttons: 'Take Tour', 'Get CKAN', and 'Live Demo'.

## Feature Overview

-  Complete catalog system with easy to use web interface and a powerful API
  -  Strong integration with third-party CMS's like Drupal and WordPress
  -  Data visualization and analytics
  -  Workflow support lets departments or groups manage their own data publishing
  -  Fine-grained access control
  -  Integrated data storage and full data API
  -  Federated structure: easily set up new instances with common search
- [Read More](#)

## Support and Hosted Solutions

CKAN is open source and can be downloaded and used for free. Users can also get hosting and support from a range of suppliers.

A full-time professional development team at the Open Knowledge Foundation maintains CKAN and can provide full support and hosting with SLAs.

[Read More](#)

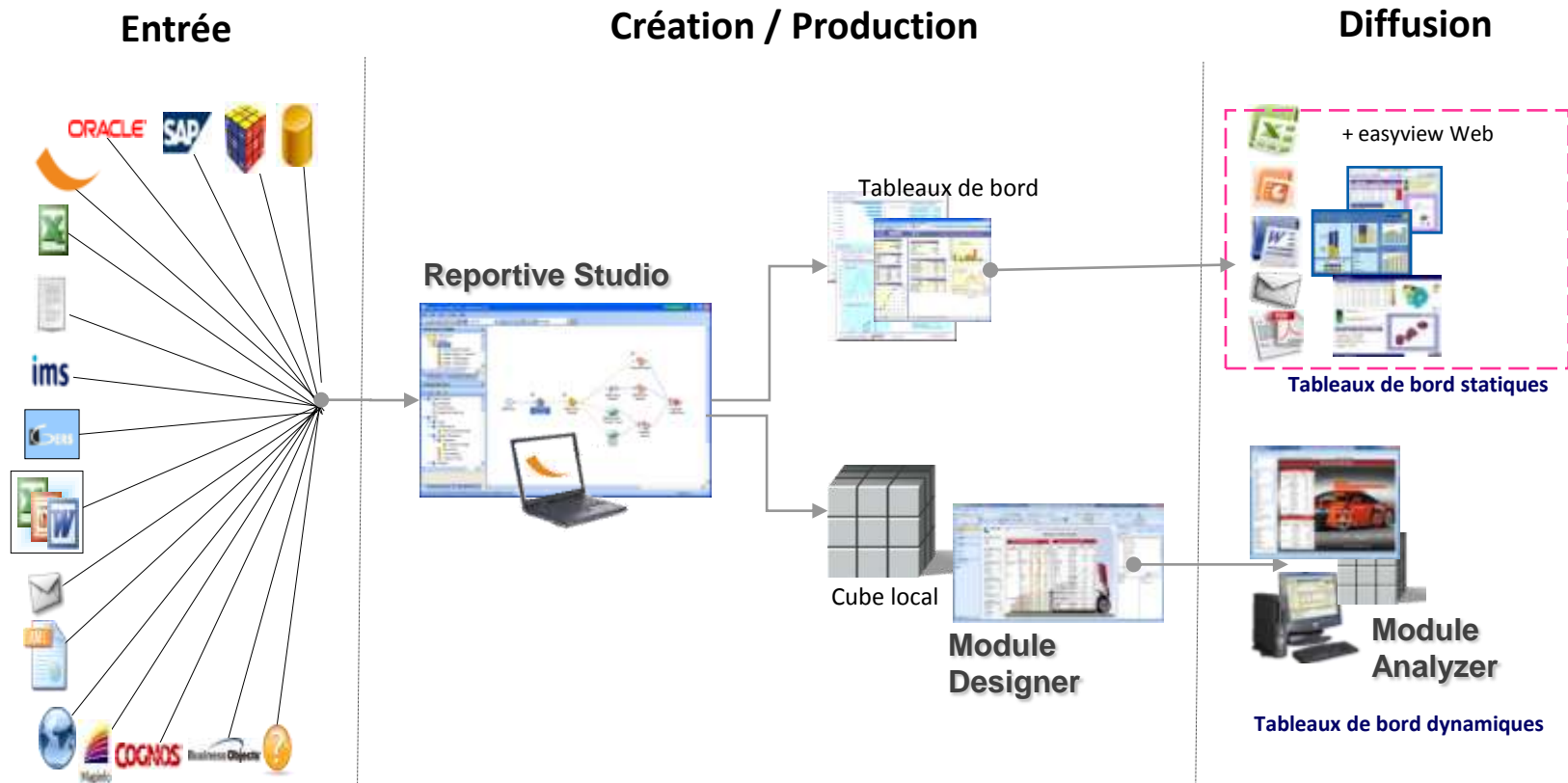
<http://ckan.org/>



The screenshot shows the data.gov.uk website. The header includes the CKAN logo and navigation menu. The main content area displays search results for datasets. The search bar shows '8572' results. The results are organized into columns for 'Search Results', 'Type', 'Publishers', and 'UK Location'. The 'Search Results' column shows a list of datasets with their titles and descriptions. The 'Type' column shows the category of the datasets. The 'Publishers' column shows the organizations that have published the datasets. The 'UK Location' column shows the geographical location of the datasets.

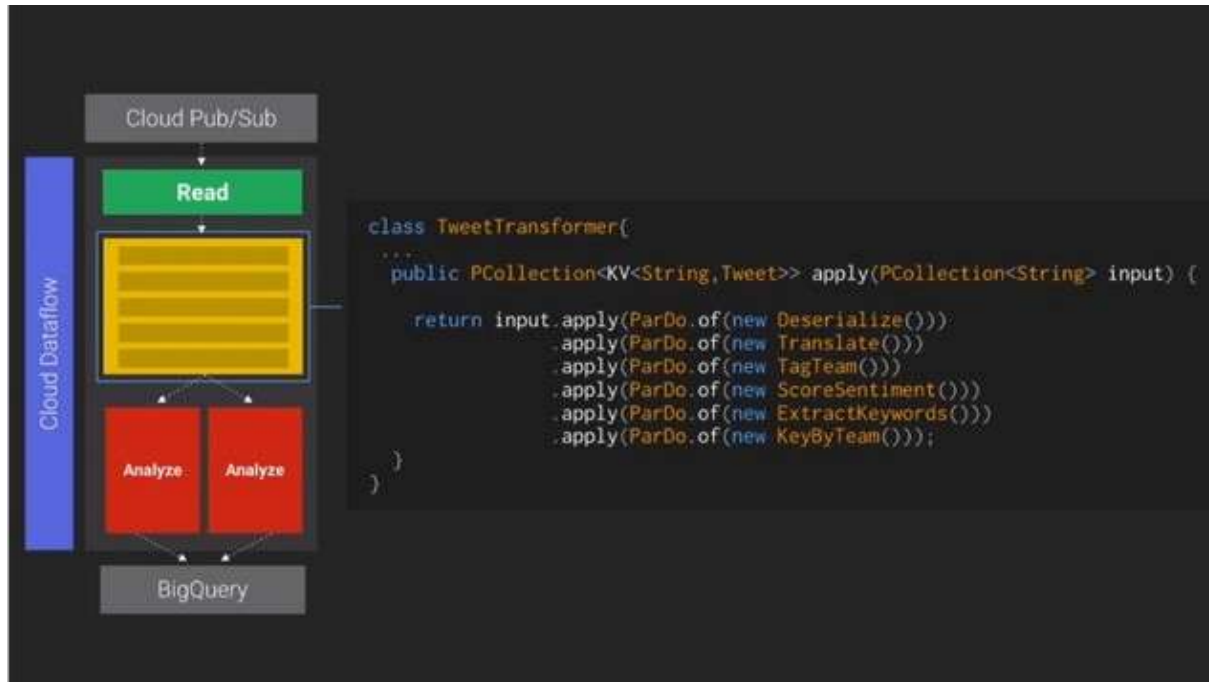
# Data architecture : Data world

- AGL Data
  - Reportive ... équipement de bout en bout



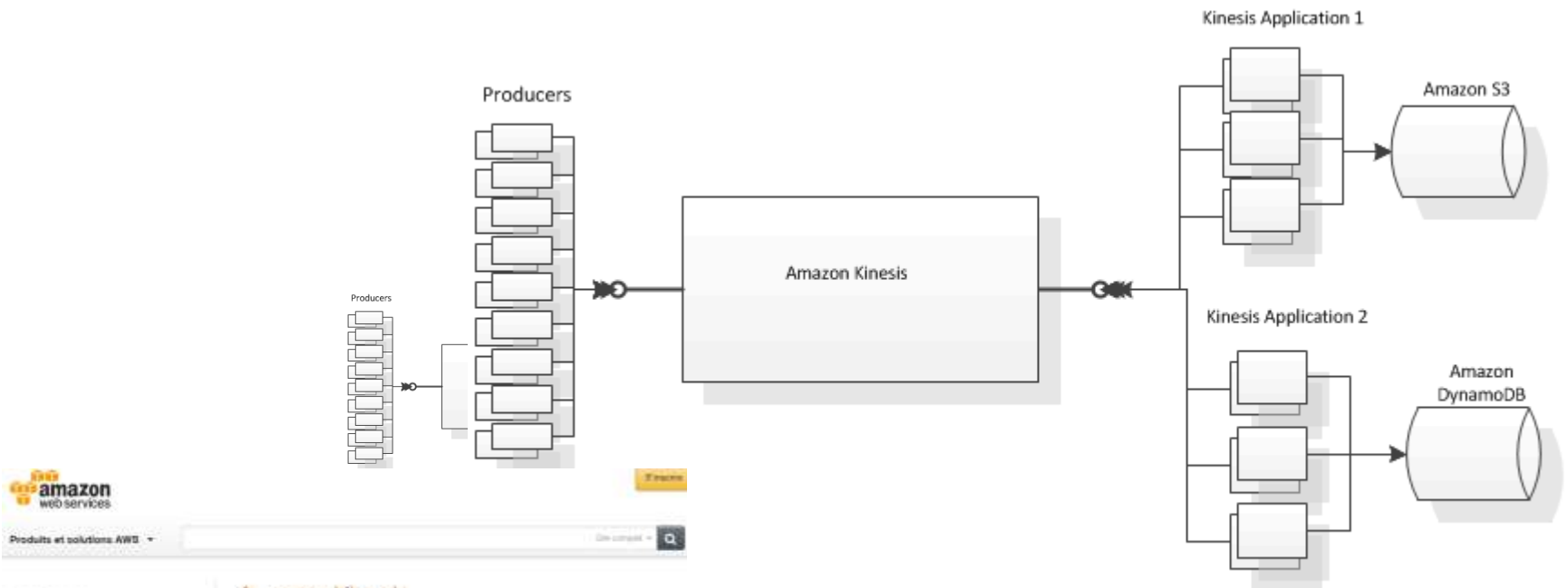
# Data architecture : Data world

- AGL Data ...
  - Pipeline (création de flux de données à la volée)
    - “Google Launches Cloud Dataflow, A Managed Data Processing Service”



# Data architecture : Data world

- AGL Data ...
  - Pipeline (création de flux de données à la volée)
    - Amazon Kinesis



The screenshot shows the Amazon Kinesis page on the AWS website. The page title is 'Amazon Kinesis'. The main content area contains the following text:

Amazon Kinesis est un service entièrement géré permettant de traiter en temps réel des données récoltées en continu à une échelle massive. Amazon Kinesis peut collecter et traiter des centaines de téra-octets de données par heure, provenant de centaines de milliers de sources. Vous pouvez ainsi développer des applications assurant le traitement d'informations en temps réel, à partir de sources telles que les flux de clics sur les sites Web, des informations commerciales et financières, des données provenant de l'instrumentation de systèmes de production et des réseaux sociaux, ainsi que des journaux opérationnels et des relevés de compteurs.

Courtesy: AWS Website | Amazon Kinesis High Level Architecture

<http://docs.aws.amazon.com/kinesis/latest/dev/key-concepts.html>

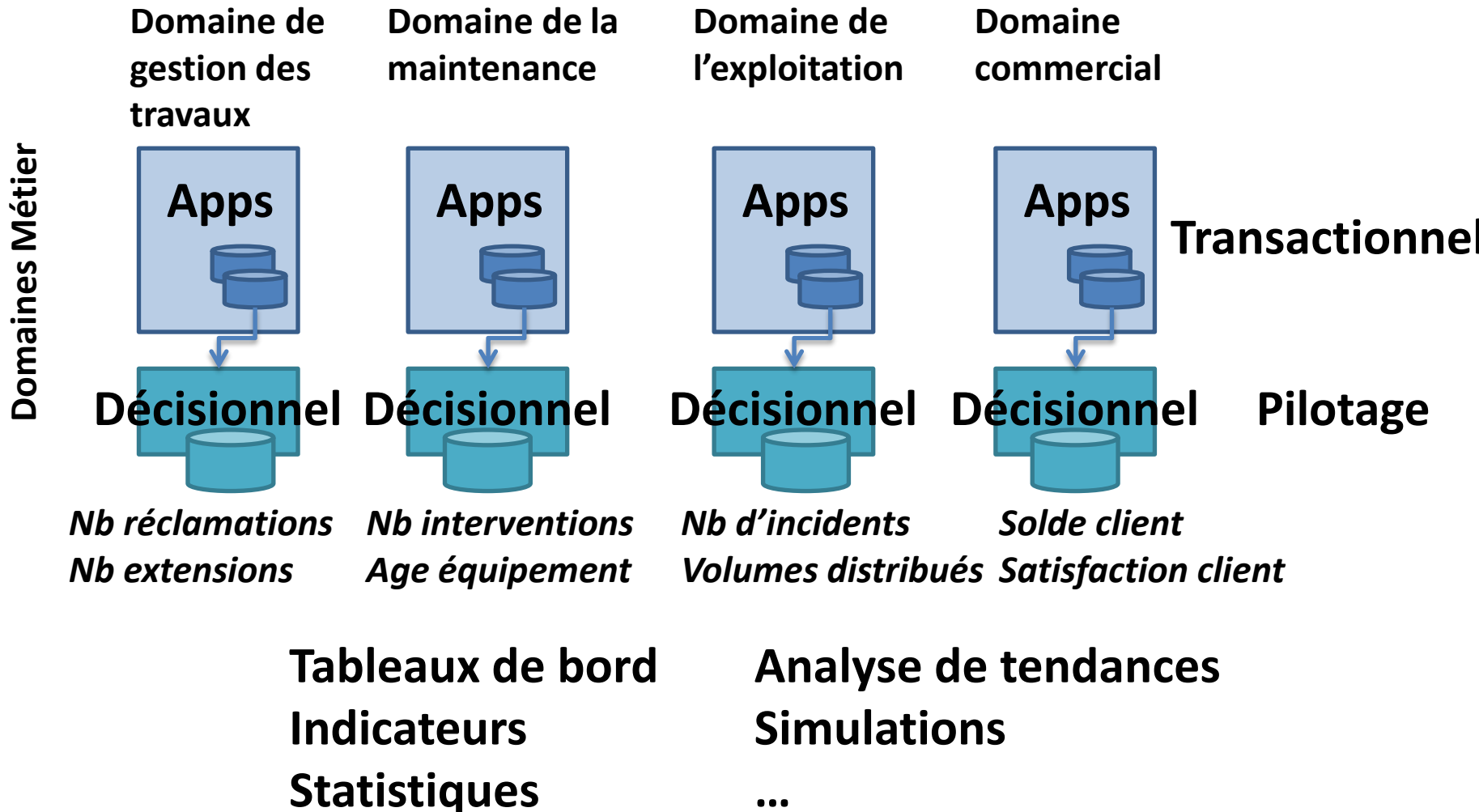
# Data architecture : Data world

- **Data Lake, Data Virtualization ... socle de données**
  - **Trop d'effort pour localiser les données et y avoir accès**
  - **Les données ne doivent pas seulement être extraites mais surtout combinées au travers différentes interfaces, applications et formats.**
  - **Cible idéale : un unique point de vue d'accès homogène à l'ensemble des données**



# Data architecture : Data world

- Data lake ... Data warehouse ... Data Lab



# Data architecture : Data world

- Data lake ... Data warehouse ... Data Lab

Domaine de  
gestion des  
travaux

Domaine de la  
maintenance

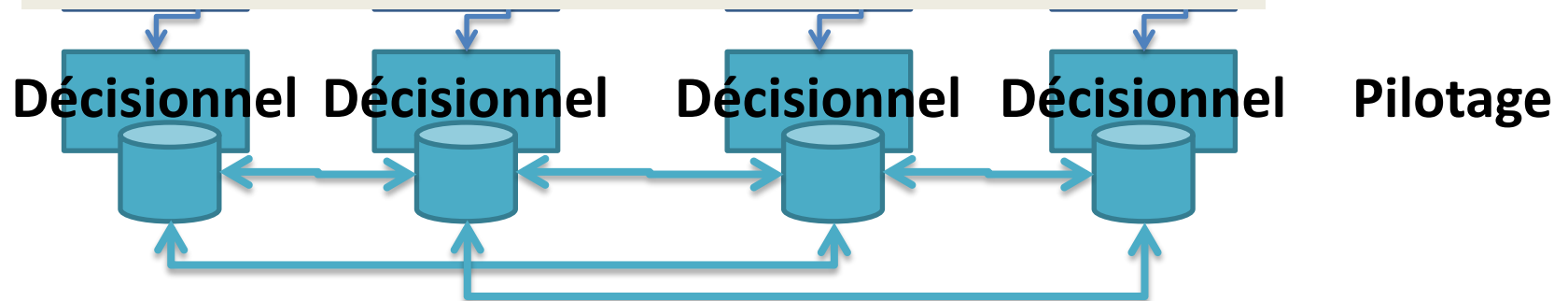
Domaine de  
l'exploitation

Domaine  
commercial

Domaines Métier

**Multiplication des échanges et des stockages de données – tout le monde veut les données de tout le monde**

Transactionnel



*Nb réclamations*

*Nb interventions*

*Nb d'incidents*

*Solde client*

*Nb extensions*

*Age équipement*

*Volumes distribués*

*Satisfaction client*

**Analyses transverses : croisement de données**

*Rapprochement solde client et incidents / réclamations / interventions*

*Rapprochement nb incidents et âge équipement*

# Data architecture : Data world

- Data lake ... Data quality ... Open Data / DaaS ...

Domaine de  
gestion des  
travaux

Domaine de la  
maintenance

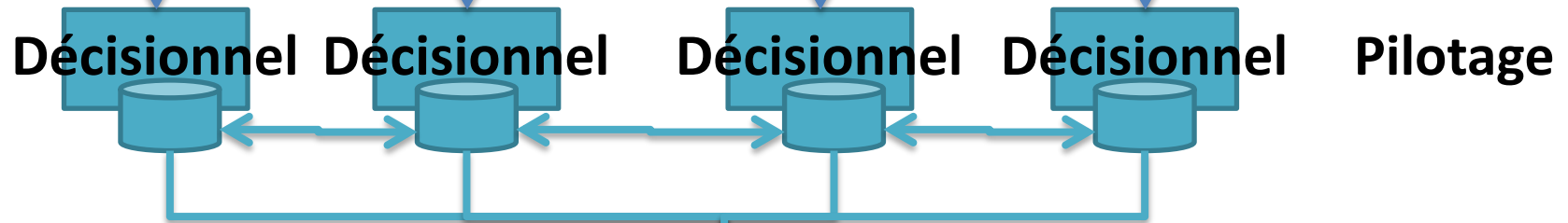
Domaine de  
l'exploitation

Domaine  
commercial

Domaines Métier

Contrôles, écarts, traçabilité de bout en bout,  
détails ... nécessité à nouveau de multiplier les  
échanges et les stockages des données

Transactionnel



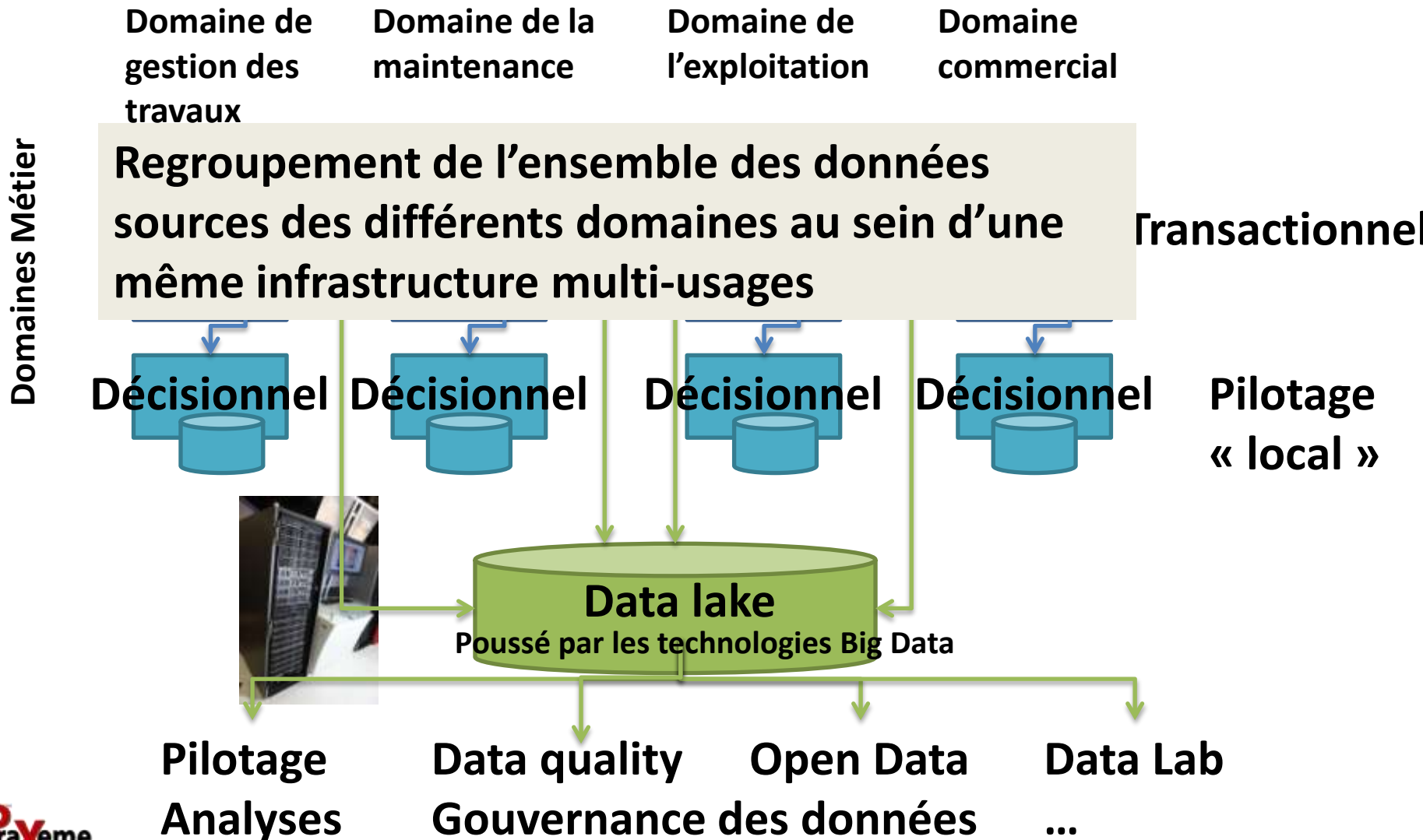
Mise en qualité  
des données  
Enrichissement

Point de vérité Publication

Une même données peut être stockées  
plusieurs dizaines de fois dans le S.I.

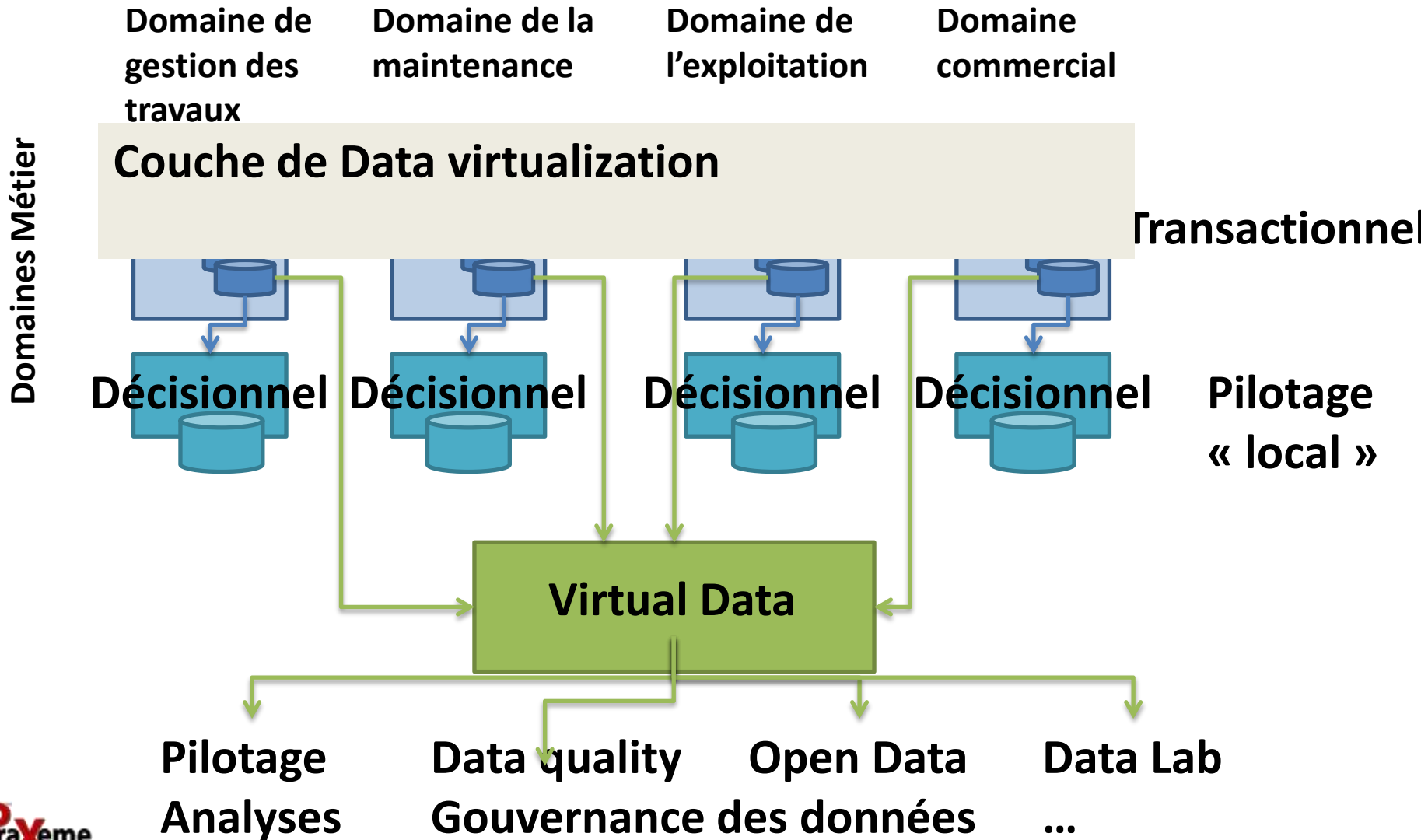
# Data architecture : Data world

- Data lake ... socle de données, puits de données...



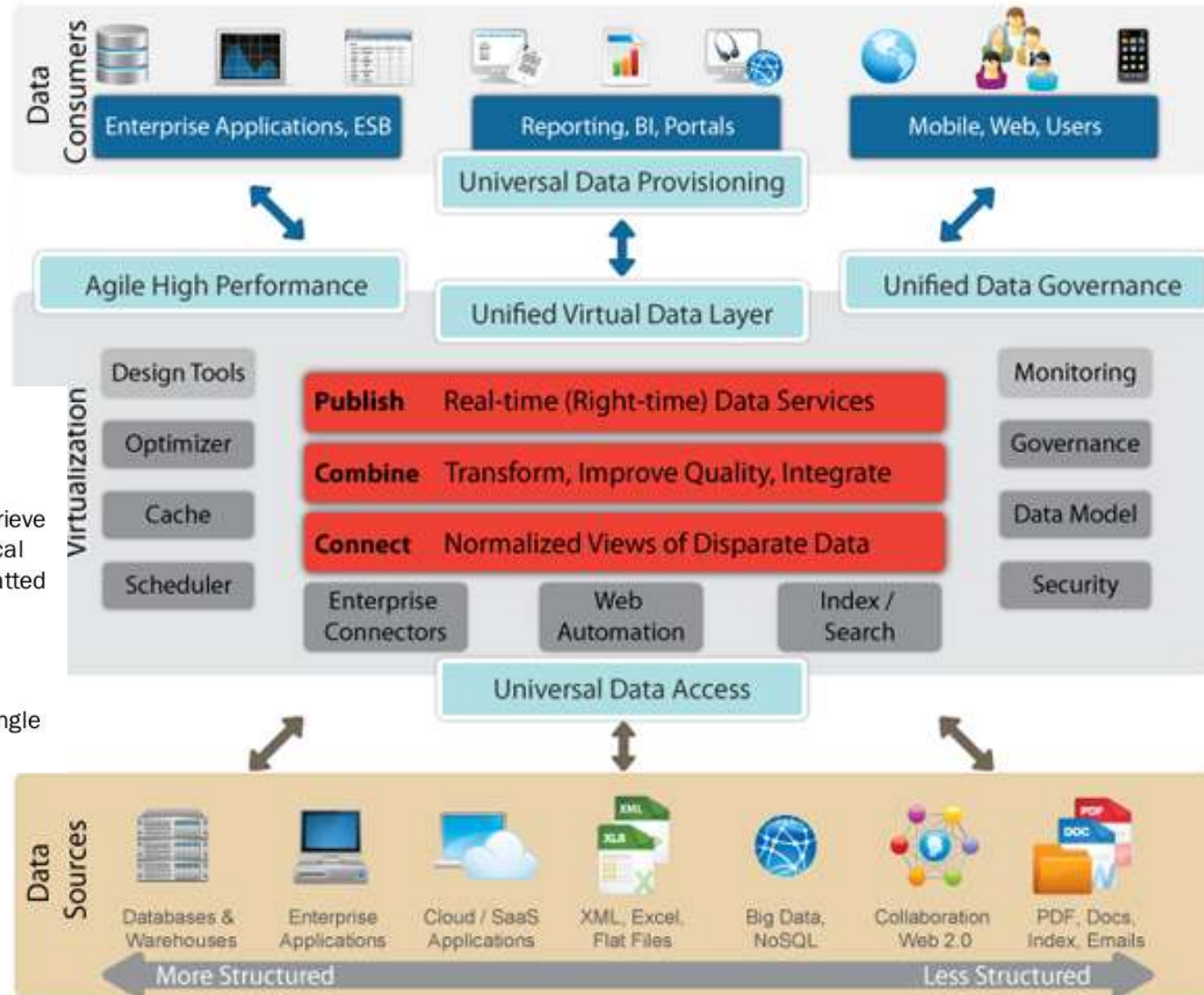
# Data architecture : Data world

- Data virtualization



# Data architecture : Data world

- Data virtualization



## What Is Data Virtualization?

- Wikipedia:  
 “Data virtualization is [...] an application to retrieve and manipulate data without requiring technical details about the data, such as how it is formatted or where it is physically located.”
- Or more simply:  
 A solution that sits in front of multiple data sources and allows them to be treated as a single SQL database

# Data architecture : Data world

- Data virtualization



Communities MySupport Marketplace Monde ▾

Vision

Produits

Solutions

Clients

Services et formation

Ressources

Société

Produits > Plate-forme Informatica > Vibe

## Informatica Vibe™ Virtual Data Machine

Vibe

### Nouvelles ressources

#### Livres blancs

Trois approches pour gérer la croissance des bases de données

Vibe est un moteur de gestion de données intégrable capable d'accéder à tout type de données, de les agréger et de les gérer, et sur lequel repose la plate-forme de données intelligente d'Informatica.

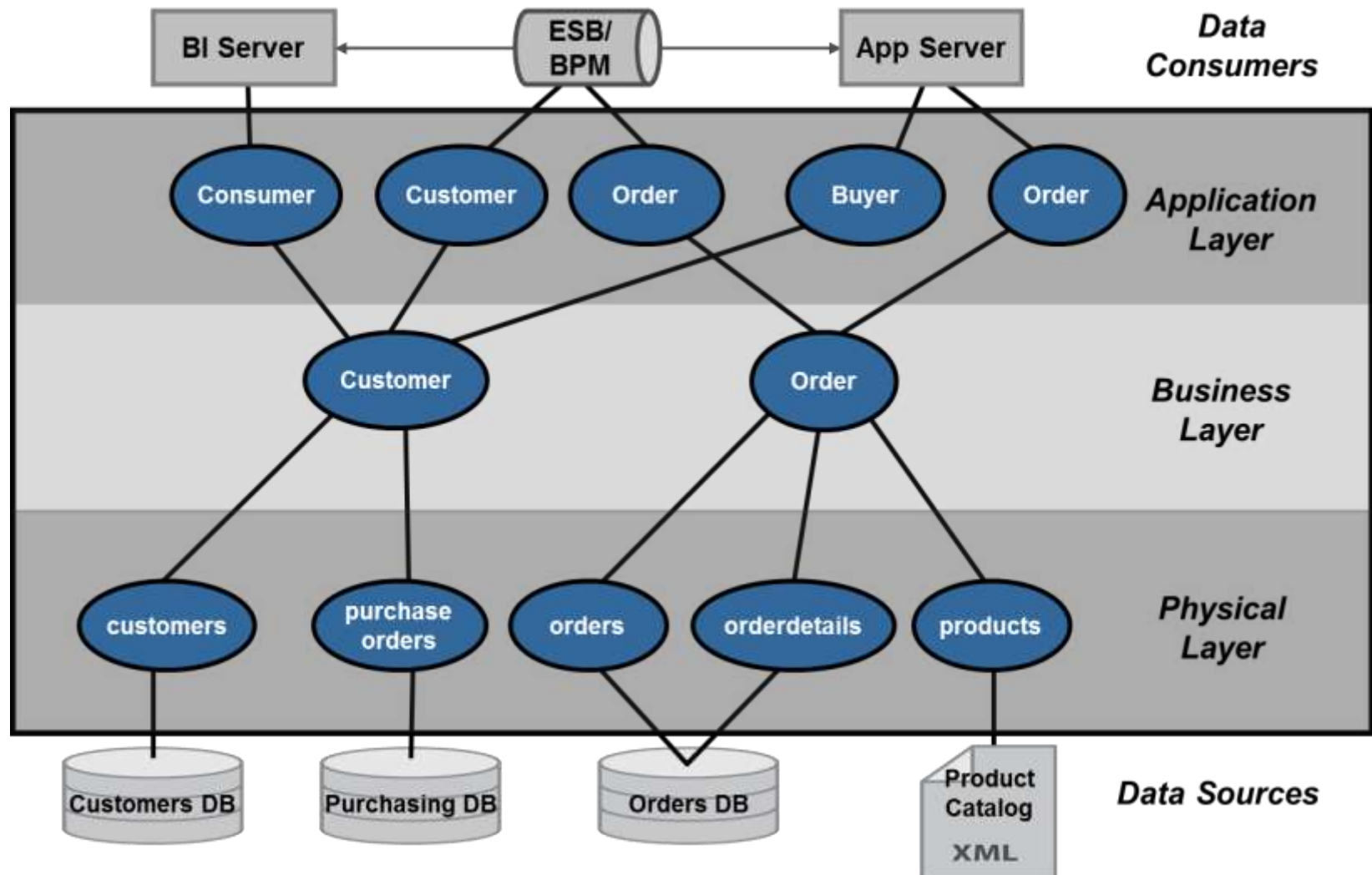
### Fonctionnement de Vibe

Vous devez vous adapter à l'émergence de nouvelles sources de données, à l'évolution des utilisateurs ainsi qu'à l'avènement de nouvelles technologies de données. Grâce à Vibe, vous le pouvez désormais — sans procéder à un recodage ou à une réécriture.



# Data architecture : Data world

- Data virtualization ... Data abstraction



<http://www.compositesw.com/data-virtualization/data-abstraction/>



# Data architecture : Data world

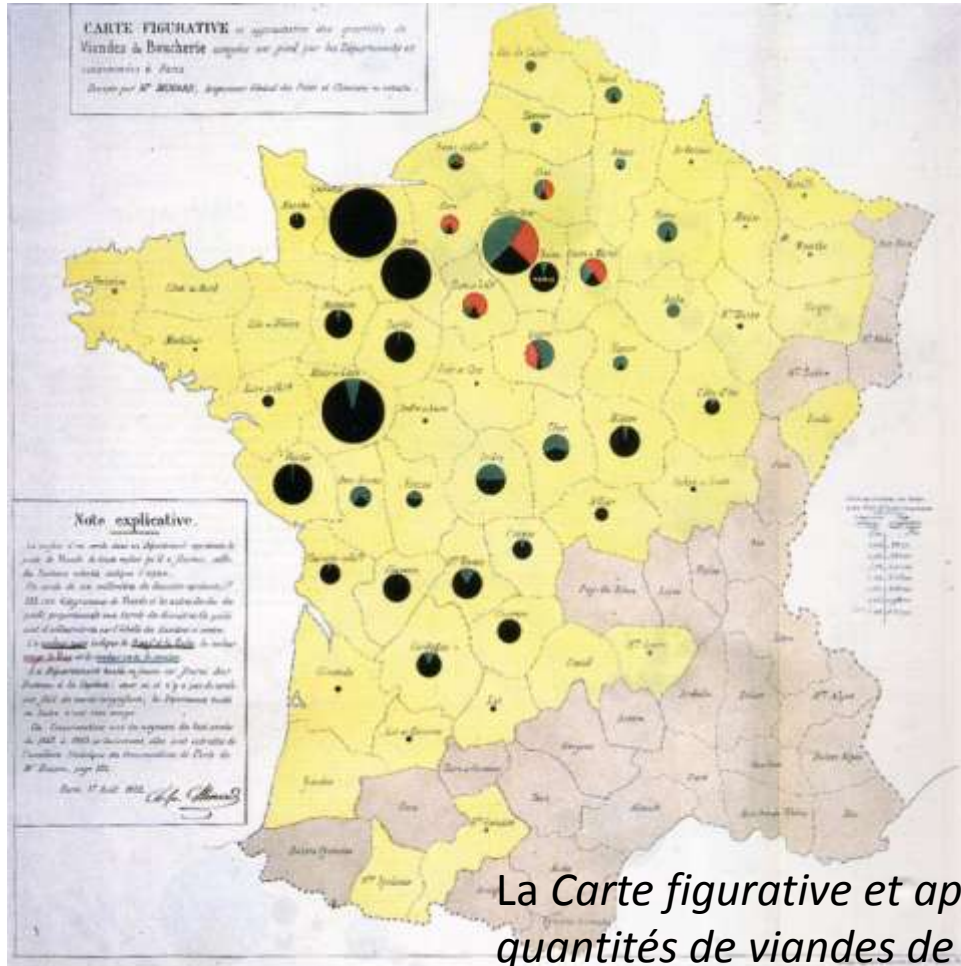
- **Metadata**
  - **Données sur les données**
    - **Auteur, localisation de l'acquisition, date et heure d'acquisition, moyens d'acquisition, degrés de confiance, statut (validation), usages, éléments de classements (ex tags), composition (résolution d'image, nombre de mots d'un texte), méta-données techniques (système logiciel d'acquisition, version, format de codage...), droits associés**
  - **Meta-Data architecture ... données qui qualifient la data architecture (qui sont les contributeurs, quelle qualité de données, quels formats sont utilisés, quelle fraîcheur des données, qui les utilisent, pour quoi faire... ?)**

# Data architecture : Data world

- **Metadata**
  - **Exemple SDMX – Statistical Data and MetaData Exchange (domaine de l'éducation – UNESCO – statistiques sur l'éducation dans les Etats membres - World Education Indicators)**
  - **Collecte de données des différents Etats**
    - **Problème de différence de qualité de collecte, d'interprétation des données à recueillir, de mode de calcul des indicateurs**
    - **Calendrier de collecte différent**
    - **Changement d'une année sur l'autre**
    - **...**
  - **Mise en œuvre de la norme SDMX pour homogénéiser, contrôler, éclairer la collecte**

# Data architecture : Data world

- Data vizualisation, data storytelling...



*La Carte figurative et approximative des quantités de viandes de boucherie envoyées sur pied par les départements et consommateurs à Paris*



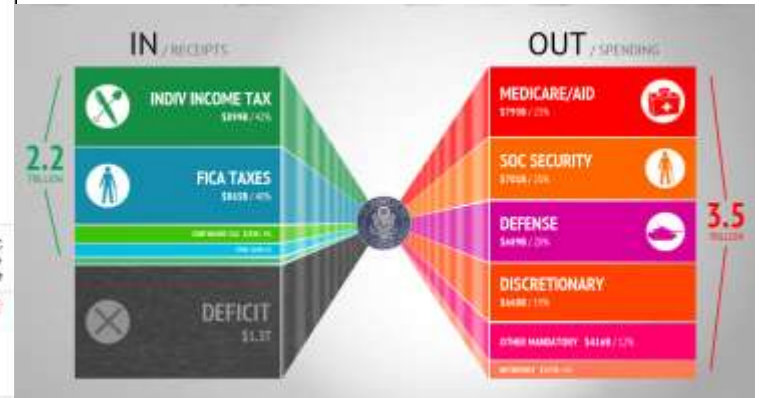
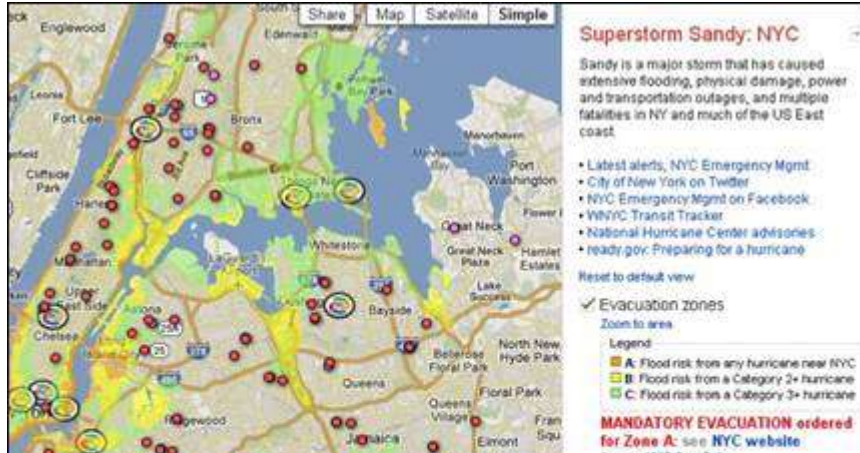
**Charles Joseph Minard** (né le [27 mars 1781](#) à [Dijon, Côte-d'Or](#))

[http://fr.wikipedia.org/wiki/Charles\\_Joseph\\_Minard](http://fr.wikipedia.org/wiki/Charles_Joseph_Minard)  
<http://www.datavis.ca/papers/>

# Data architecture : Data world

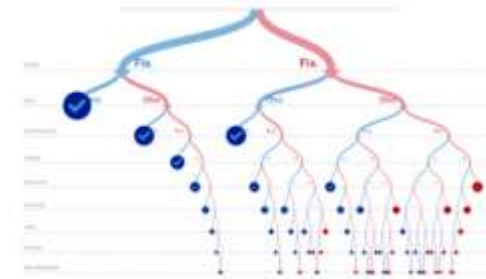
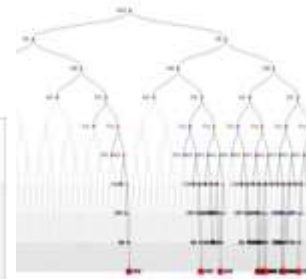
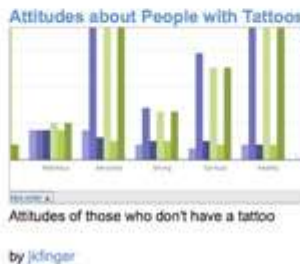
- Data visualization

*L'homme est fondamentalement visuel*



15 of 257

81 of 257



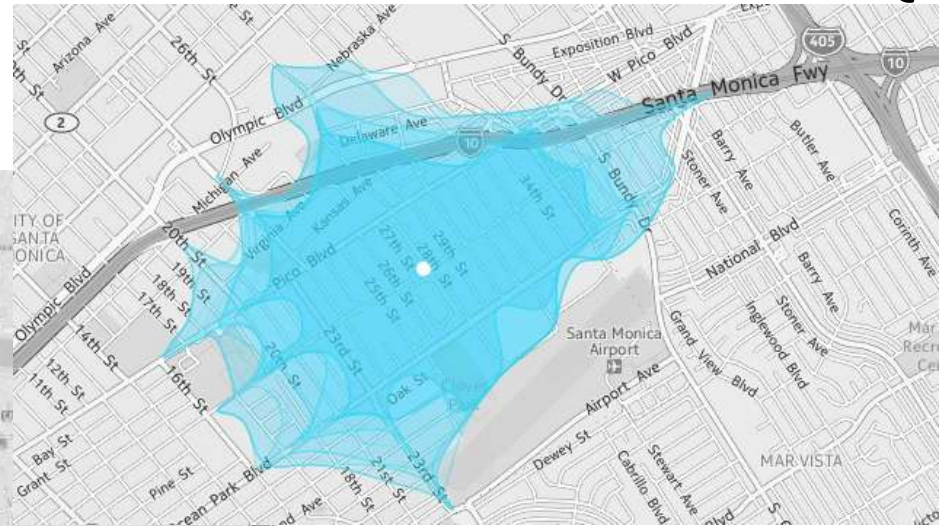
<http://www.youtube.com/watch?v=DtqSlpIGXOA&feature=youtu.be>

# Data architecture : Data world

- Data visualization

*L'homme est fondamentalement visuel  
Données -> Visualisation -> Sens -> Solution -> Problème*

## Isoscope Shows How Far You Can Drive in 10 Minutes from Where You Are



<http://www.flaviogortana.com/isoscope/>

# Data architecture : Data world

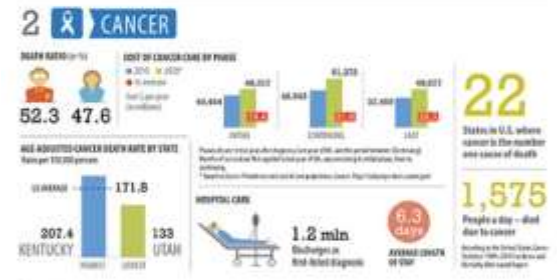
- Data storytelling

## WHAT KILLS US? THE REAL THREAT TO AMERICANS

Based on the key mortality reports of 2010 from National Vital Statistics System, a view on top 5 causes of death in the United States



44,981 PEOPLE DIE EVERY YEAR BECAUSE OF HEART DISEASE	40,996 PEOPLE DIE EVERY YEAR BECAUSE OF BREAST CANCER	6.3% PEOPLE DIE EVERY YEAR BECAUSE OF COPD	185,000 PEOPLE SUFFER FROM STROKE EVERY YEAR IN THE UNITED STATES	40.7% OVER 100 MILLION AMERICANS SUFFER FROM UNINTENTIONAL INJURIES
---	---	--	---	--



# Data architecture : Data world

- **Data Policy**
  - **Evolution réglementaire**
    - **Indicateurs : Quoi**
    - **Données détaillées : Pourquoi**
    - **Infrastructure et gouvernance : Comment**
  - **Services : consulter, corriger...**
- **Sécurité**
- **Data stratégie : ouverture, publication**
- **Exigences en termes de qualité de données**
- **Règles de gouvernance**

# Data architecture : Data world

- **Big Data - écosystème qui vise à :**
  - **Exploiter le flot de données :**
    - **Exhaustivité versus échantillonnage**
    - **Volume, Variété, Vitesse**
    - **Support Data science : corrélation (Quoi), statistiques, visualisation**
  - **Dépassez les technologies BI**
    - **Des technologies de stockage : NoSQL, Hadoop...**
    - **Des technologies de manipulation et d'exploitation des données : parallélisme - Map/Reduce, ...**



# Data architecture : Data world

- **Big Data**



**X**



**Volume**  
**Variété**  
**Vitesse**

**X**

**Données structurées**  
**Données semi-structurées**  
**Données non structurées**

# Data architecture : Data world

- Big Data

## Limites des «technologies» classiques

[Volume Variété Vitesse] -> Traitements

---

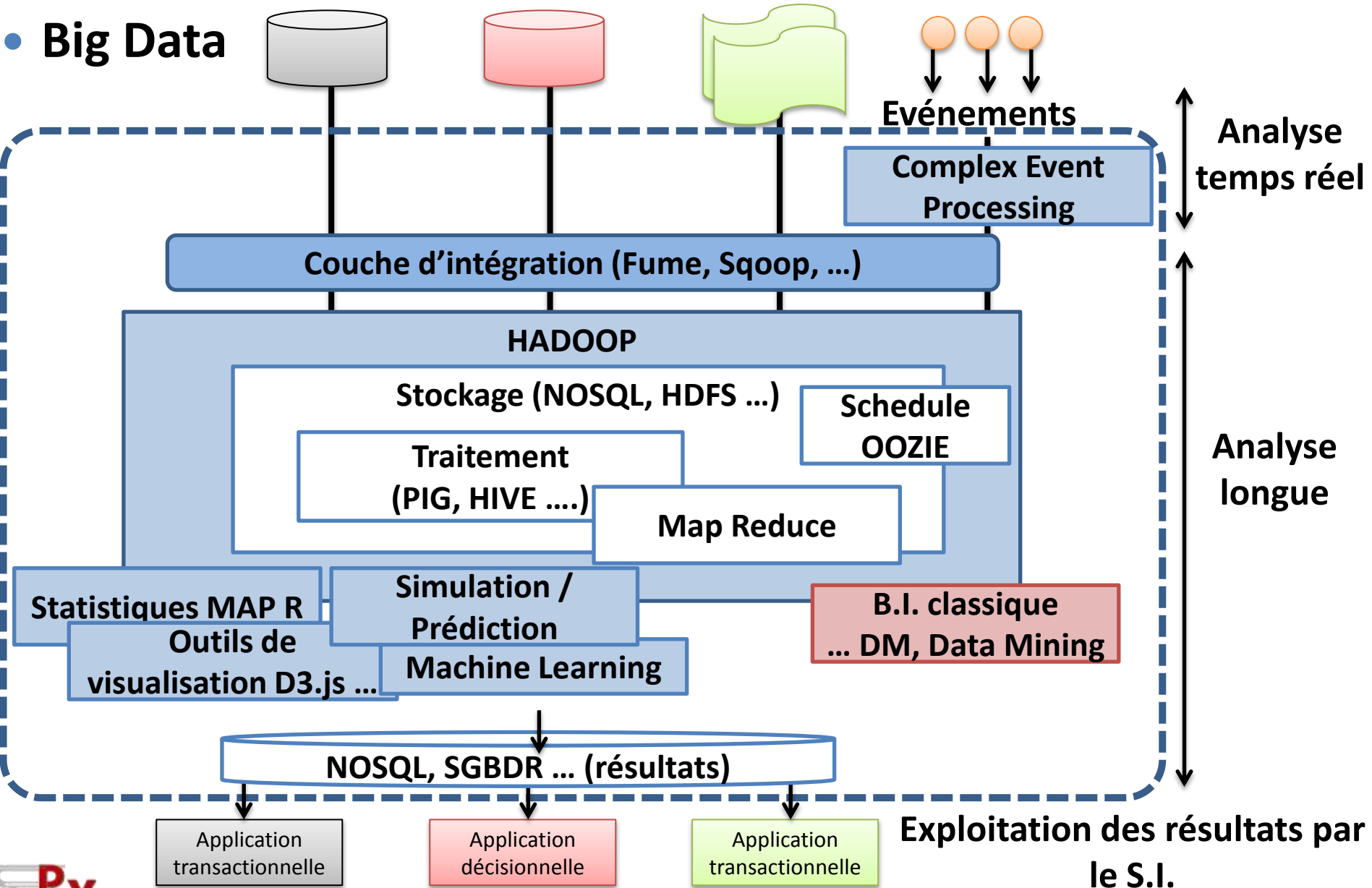
### Limites :

- Vitesses I/O
- Modèles physiques de stockage (SGBDR ... scalabilité, performances)
- Traitements majoritairement séquentiels
- Accès privilégiés DDUR vs RAM
- Modélisation SQL (transactionnel)
- Batches, ODS -> DW -> DM
- Silos données structurées – non structurées
- Normalisation a priori (impact variété)

➔ Emergences d'un ensemble / écosystème de «nouvelles» technologies Big Data via les acteurs du web (Google, Yahoo, Amazon ...)

# Data architecture : Data world

- **Big Data**



# Data architecture : Data world

- Big Data

## Des exemples : n°1

### Le bon prix d'un vêtement (chaîne Textile au US) – élasticité des prix en fonction :

- De la localisation
- De la météo locale
- Des données de population de la zone d'achalandage (type INSEE)
- De l'historique de vente
- Du prix de revient des vêtements
- Des stocks et approvisionnements
- Etc



# Data architecture : Data world

- **Big Data**

**Des exemples : n°2**

**Walmart impact sur le business model et sa relation fournisseurs :**

- **fourniture en temps réel des données de consommation à ses fournisseurs qui se charge d'optimiser leurs stocks**
- **versus prise en charge des stocks par Walmart !**

# Data architecture : Data world

- **Big Data**

## **Des exemples : n°3**

### **Organisme de crédit :**

**Détection de fraude en temps réel inscrite dans les processus transactionnels.**

#### **Avant**

- Temps de la transaction et temps de la détection dissociés
- Recherche de fraude à partir de données d'identité – segmentation / profiles

#### **Après :**

- Un même temps – analyse temps réel
- Recherche de patterns comportementaux : liens familiaux - géographiques, viralité de la fraude, données d'environnement socio-économique ...

**« Intelligent Business Operation » - Gartner**

# Data architecture : Data world

- **Big Data**

## **Des exemples : n°4**

**Aéroports : Précision de l'heure d'arrivée des avions (conditionne de multiples flux financiers) ... la minute de parking est précieuse !**

**Avant : heure estimée d'arrivée donnée par les pilotes en approche finale ... 10% en écart > 10' et 30% en écart d'au moins 5'**

**Après : PASSUR modèle prédictif**

- Données des pilotes,
- Météo locale
- Données du réseau de radars ([www.flightradar24.com](http://www.flightradar24.com))
- Données horaires des compagnies
- Historique d'occupation des parkings, d'encombrement du ciel du moment ...
- Configuration des aéroports

**Prédiction en temps réel de l'heure d'arrivée des avions**

# Data architecture : Data world

- **Big Data**

## **Des exemples : n°5**

**Campagne d'OBAMA 2012 ... faite voter vos amis pour Obama !**

**Quelles personnes doivent être convaincues ?**

« Les différentes équipes de campagne ont ainsi largement fait appel au processus pour cibler au mieux les électeurs potentiels. "Alors qu'auparavant les équipes téléphonaient un peu au hasard, cette année, je n'ai reçu ni coup de fil, ni e-mail de promotion des candidats", raconte l'Américain Chuck Hollis. "Il faut dire que j'ai 53 ans et que je suis un Républicain convaincu. Par contre ma fille de 22 ans, qui votait pour la première fois, a été sollicitée de toute part tout au long de la campagne ! »



# Data architecture : Data world

- **Big Data**

## **Des exemples : n°5**

**Campagne d'OBAMA 2012 ... faite voter vos amis pour Obama !**

**Quelles personnes doivent être convaincues ?**

**Sources :**

- **Registre des votants**
- **Sympathisants/supporters démocrates, adhérents**
- **Logs de visites du site de campagne (cookie tracer)**
- **Données Facebook – réseaux d'amis ...**
- **Sondages**
- **Résultats des appels sortants des équipes de campagne**
- **Historique des votes par Etat**
- **Historique des équipes Obama 2008**
- ...

# Data architecture : Data world

- **Big Data ... et aussi**

**Grande distribution (MACY) estimation des ventes avant l'enregistrement des ventes ... à partir de l'analyse de l'occupation des parkings.**

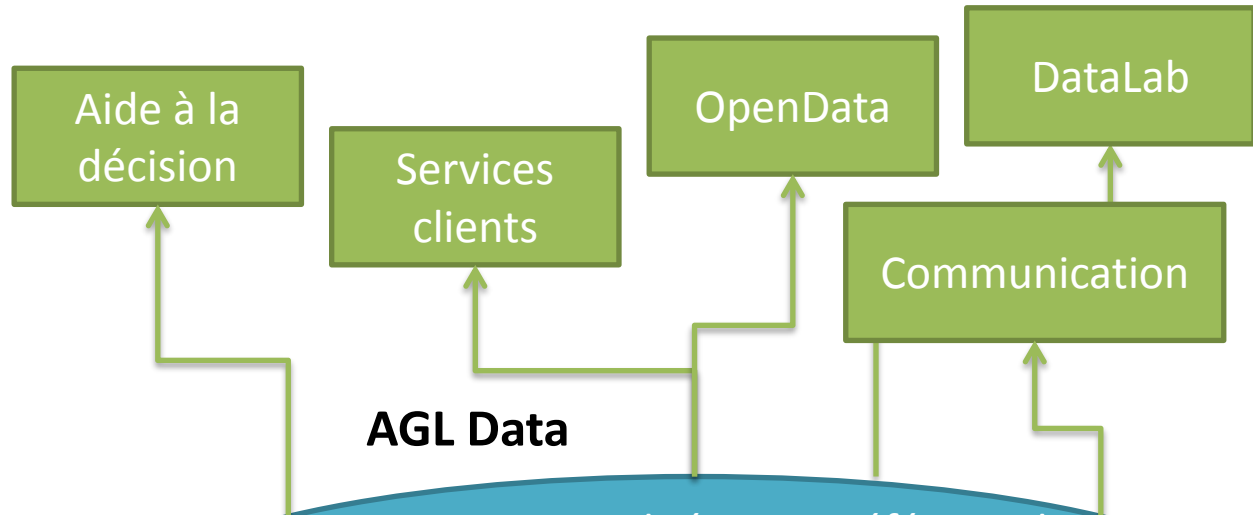
**Apple, suivi en temps réel des ventes d'iPad et comparaison avec le sentiment général exprimé sur les réseaux sociaux (détection de signes négatifs).**

**Ville de NYC, détection des fraudes fiscales ... équipe de 5 data scientists (données d'inspection, données fiscale, tickets de parking, contraventions, plaintes , cadastre...), appartements illégaux, comportement commerçants illégaux ...**

**Publicité temps réel (recommandation à partir des commentaires postés). Réaction en temps réel au churn ...**

# Data architecture

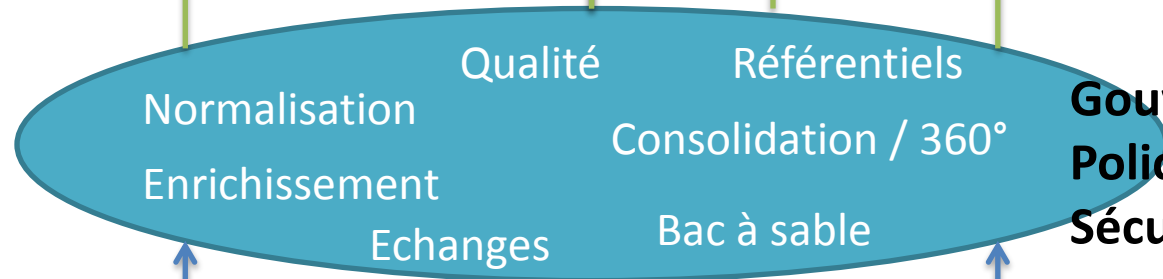
## Services



AGL Data

## Socle de données :

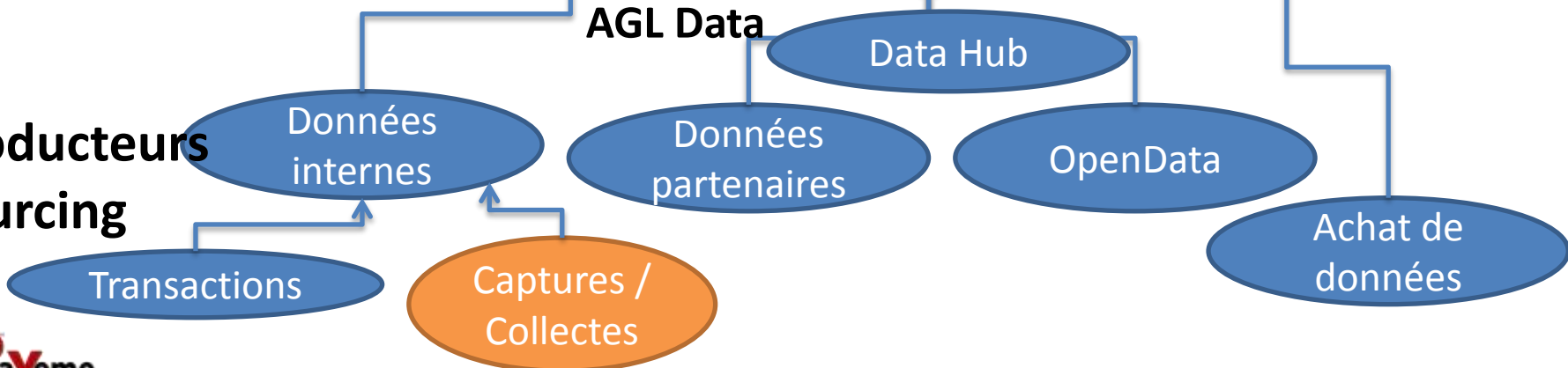
- Préparation
- Stockage / Virtualisation (dont big data)



Gouvernance  
Policy  
Sécurité

AGL Data

## Producteurs Sourcing



# Data architecture : Data war

- Comment allez-vous ?

Je vais



# Data architecture : Data war

- **Comment allez-vous ?**

**Je vais**



<http://www.numerama.com/magazine/28489-la-medecine-personnalisee-ou-la-sante-confiee-a-google-et-apple.html>

# Data architecture : Data war

- Mais qui est le médecin ?



**PHILIPS**

**SONY**



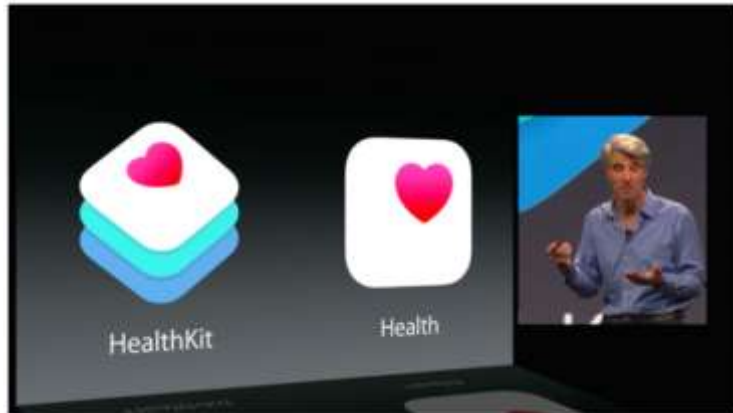
# Data architecture : Data war

- **Bataille autour des données de santé**



# Data architecture : Data war 03/06/14

- Bataille autour des données de santé – Apple



Today at WWDC, Apple formally introduced HealthKit for [iOS 8](#)

Apple propose aux développeurs d'applications et aux fabricants d'objets connectés destinés à la médecine personnalisée de **stocker toutes les données** issues des capteurs de santé dans HealthKit, pour que l'utilisateur puisse toutes les consulter ensemble, dans l'application "Health".

<http://www.numerama.com/magazine/29561-healthkit-l-inquietante-base-de-donnees-medicales-d-apple.html>



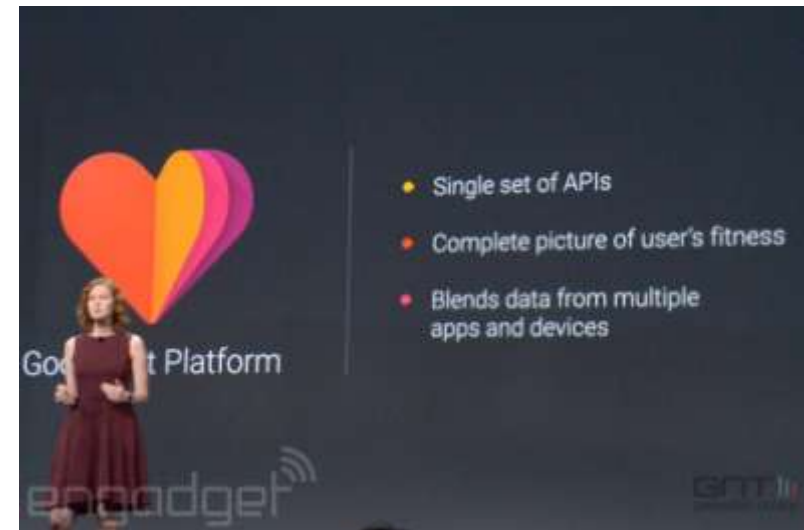
# Data architecture : Data war

27/06/14

## • Bataille autour des données de santé – Google



**Google Fit veut être votre principale source de données de santé**



« S'il fallait résumer Google Fit, nous dirions qu'il s'agit simplement d'une énorme API ouverte permettant à toute les applications et autres trackers d'activité d'envoyer et de partager leurs données, **données qui seront centralisées chez Google.**

La plate-forme Android devient un « **collecteur** » de **données** de hardware venant d'autres fabricants. Par exemple, des capteurs d'Adidas fonctionneraient avec Google Fit, et même Nike va ouvrir ses apps pour intégrer Nike Fuel avec Fit. »

<http://www.gizmodo.fr/2014/06/26/google-fit-suivi-constantes-vitales.html>

<http://fr.ubergizmo.com/2014/06/google-fit-veut-etre-votre-principale-source-de-donnees-de-sante/>

# Data architecture : Data war

## • Bataille autour des données de santé

27/06/14

### Salesforce – Philips

« "Nous proposons à nos clients du secteur de la santé de pouvoir agréger les données médicales de leurs patients en provenance de l'Internet des objets", résume Frans van Houten, CEO de Royal Philips. Au programme : les équipements d'hospitalisation à domicile (outil de monitoring, d'oxygène...), mais aussi les balances connectées... sans oublier les nouvelles apps mobiles qui commencent à apparaître sur le terrain de la santé.

La solution du groupe est livrée avec des API pour faciliter les intégrations avec les objets connectés ou apps de partenaires.

Pour l'heure, Philips prévoit de livrer cet été deux premières applications pour tablette tirant profit de l'environnement : Philips eCareCoordinator et Philips eCareCompanion. Destinée au personnel de santé, la première permet le pilotage à distance de l'hospitalisation à domicile, jusqu'à l'accès aux données des systèmes de monitoring installés au domicile du patient, au contrôle de l'usage des containers d'oxygène ou de la pression du sang. La seconde application, elle, est à l'usage du patient, pour communiquer avec le médecin sur son état de santé. »

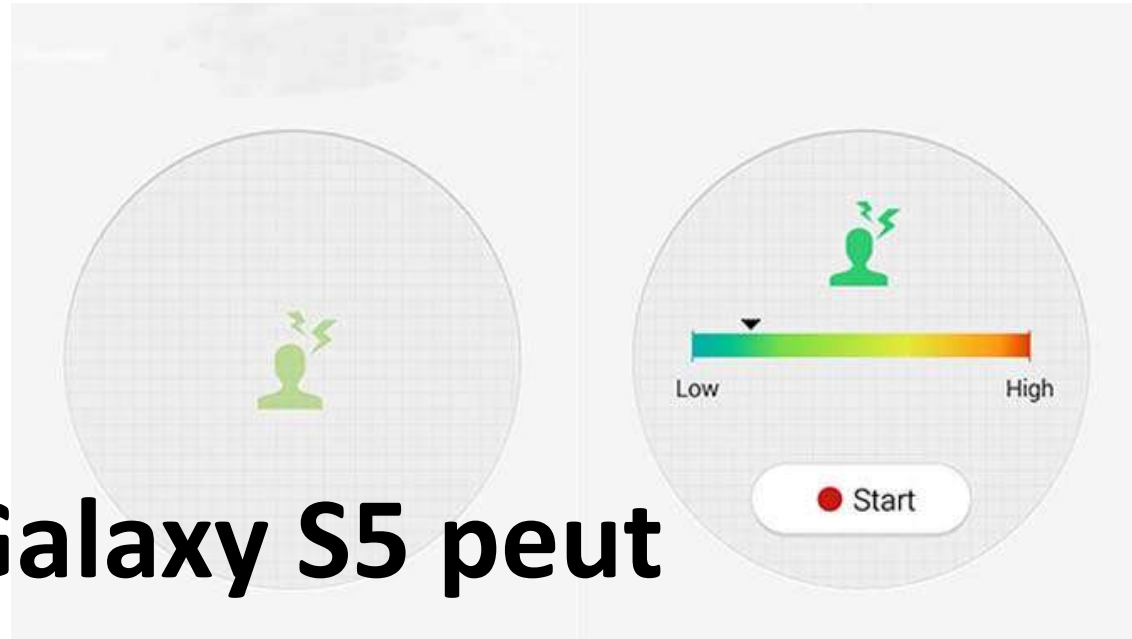


An advertisement on a blue background. At the top right is the "PHILIPS + salesforce" logo. On the left, a tablet displays the "PHILIPS eCareCompanion" interface with a "NEW" sticker and six "Start Now" buttons. Below it, the text reads "Connect Devices, Patients &amp; Providers Real-Time". On the right, a laptop displays the "PHILIPS eCareCoordinator" interface with a "NEW" sticker and a line graph. Below it, the text reads "Collaborate &amp; Monitor 1000 's of Patients Real-Time".

<http://www.journaldunet.com/solutions/dsi/accord-philips-salesforce-0614.shtml>  
<http://www.informatiquenews.fr/accord-philips-salesforce-linformatique-medicale-17751>

# Data architecture : Data war

- **Bataille autour des données de santé**



## Le Samsung Galaxy S5 peut mesurer votre stress

À l'occasion d'une mise à jour de son application mobile S Health, Samsung a ajouté une nouvelle fonctionnalité tirant avantage du capteur mesurant le rythme cardiaque présent dans le Galaxy S5. Désormais, le logiciel peut évaluer le niveau de stress de l'utilisateur observant les battements de son cœur.

<http://www.numerama.com/magazine/29508-le-samsung-galaxy-s5-peut-mesurer-votre-stress.html>

# Data architecture : Data war

30/05/14

- Bataille autour des données de santé - Lab Samsung



## SIMBAND

### YOUR BODY IS TALKING TO YOU

Samsung Simband is not a product. It's a reference design. It's our concept of what a smart health device should be. Devices based on the Simband platform will be able to gather vital diagnostic information - from your heart rate to your skin's electrical conductivity, 24 hours a day, seven days a week.

Now, for the first time you will be able to ask specific wellness questions and get clear, insightful answers direct from the source: you. Simband is being designed to power you on your way to better health and a richer, more fulfilling life.



# Data architecture : Data war

30/05/14

- Bataille autour des données de santé - Lab Samsung



## WE'RE OPENING THE ECOSYSTEM

Samsung Architecture Multimodal Interactions (SAMI) will be a data broker that will enable wearable devices like those based on Simband to upload information to the cloud. From there, developers can access the data and leverage it to create entirely new applications. SAMI will be the first truly secure, open, diverse data platform of its kind. Its open APIs enable it to collect data from any type of device or online source and deliver it to any other device or application in real time. These APIs are simple, which allows sensor developers to recruit data in the way they want.

## INTELLIGENCE FOR SMARTER HEALTH

The Samsung Digital Health Initiative leverages an array of hardware and software technologies designed to bring developers, healthcare professionals, academics and health enthusiasts together to create a healthier world. The key component of the Digital Health Initiative is a new health open reference design platform tailored to take advantage of the latest sensors, behavioral algorithms, battery technologies and displays. This is an open data platform and open sensor platform that's attracting the brightest minds and the most creative innovators to work with Samsung to accelerate innovation in personal health technology.

SIMBAND    SAMI    PARTNERS

# Data architecture : Data war

02/06/14

- Bataille autour des données de santé – AXA

## Ça y est, AXA conditionne un avantage santé à un objet connecté

Aurez-vous encore votre liberté lorsqu'il faudra choisir entre payer cher sa mutuelle santé, ou accepter de voir son comportement scruté par des machines au service de l'assurance ? Pour le moment sous forme d'un jeu donnant droit jusqu'à 100 euros de médecine douce, AXA montre l'avenir de l'assurance santé liée aux objets connectés et à la médecine personnalisée



Pour **7 000 pas /jour**  
pendant 1 mois



1 chèque de médecine  
douce d'une valeur de  
**50 €**

**+**  
**15%**  
de remise  
sur vos achats  
sur le site  
withings.com

Pour **10 000 pas /jour**  
pendant 1 mois



2 chèques de médecine  
douce d'une valeur de  
**50 €** chacun

**+**  
**20%**  
de remise  
sur vos achats  
sur le site  
withings.com

<http://www.withings.com/fr/withings-pulse.html>

<http://www.numerama.com/magazine/29556-aa-y-est-axa-conditionne-un-avantage-sante-a-un-objet-connecte.html>

# Data architecture : Data war

- **Bataille autour des données de santé – SFR**

The screenshot shows the SFR Business Team website. At the top left is the SFR logo and 'BUSINESS TEAM'. A dropdown menu is set to 'Toutes nos solutions'. On the right, there are links for 'Acheter', 'Assistance', 'Business Room', and 'Espace client'. Below this is a navigation bar with five buttons: 'TOUT EN UN' (highlighted in red), 'TÉLÉPHONER', 'INTERNET ET RÉSEAU', 'RELATION CLIENT', and 'CLOUD & COLLABORATION'. The main content area has a breadcrumb trail: 'Accueil > Nos solutions Entreprises > Services Managés Santé'. The main heading is 'SERVICES MANAGÉS SANTÉ' with the subtitle 'L'offre d'hébergement de votre plateforme santé'. Below this is a large banner with a background image of doctors in a clinical setting. The banner text reads: 'LES DONNÉES DE SANTÉ SONT VITALES !' and 'Confiez leur hébergement à un opérateur de confiance'.

# Data architecture : Data war

## • Bataille autour des données de santé – Cloud

### What is OncologyCloud™?

#### The Problem

Only 4% of patients diagnosed with cancer in the United States participate in clinical trials. For the rest, their data is captured in silos and inconsistent/unstructured formats. Because of this, the industry is not learning from the experiences of 96% of patients' journeys.

#### Our Approach

Flatiron's OncologyCloud platform aggregates and transforms clinical and financial data from EMR and billing systems in real-time. The result is a data platform that provides the most comprehensive view of a patient's experience in the oncology office in real-time.

### Why We Built OncologyCloud™

#### Big Data

In today's world, it is increasingly critical for organizations to harness and leverage their data. "Big Data" has transformed entire industries into being smarter and faster. However, the oncology industry has not yet seen this revolution. Despite the growth of electronic medical record (EMR's), a lack of standards, information silos, fragmentation of information and abundance of unstructured data has prevented the oncology industry from learning from real-world patient experiences.



#### Oncology is complicated

Most of the "clinically valuable" data in cancer care resides in doctor and nurse notes, pathology reports, PDF's, and other unstructured forms. Even structured data has major integration issues, with EMR's having adopted their own custom ontologies and standards. Traditional population health analytics are built using claims data, which can attain quick results but lack the depth required to understand a disease as complex as cancer. To run analytics for oncology off of claims data would be the equivalent of analyzing an iceberg by only looking at what's above the surface. If you want the "clinical truth," you have to get into the details.



- 1) Collecte de données sur les personnes atteintes de cancer
- 2) Recueillies auprès des compagnies d'assurance, cabinets médicaux et les hôpitaux via le Cloud
- 3) Mises à disposition des chercheurs, médecins, centres de soins pour aider aux décisions thérapeutiques.

<http://www.flatiron.com/#/>  
<http://www.dsih.fr/article/1044/google-investit-dans-les-big-data-en-sante.html#sthash.CLOuHVuR.dpuf>

09/07/2014



# Data architecture : Data war

- Bataille autour des données de santé –  
Corps médical - Apps



<http://www.e-soins.fr/>

# Data architecture : Data war

## • Bataille autour des données de santé – Corps médical – Capteurs - Apps

### Cardiauvergne : un hôpital virtuel pour les patients insuffisants cardiaques

« La prise de poids rapide, liée à une rétention hydrosodée, est un marqueur fort de la maladie. Tous les patients nous transmettent donc leur pesée chaque matin », explique le Pr Cassagnes. Un dossier médical informatisé, comprenant trois volets (clinique, biologique et thérapeutique), génère des alertes et des alarmes. Lesquelles sont gérées « dans les 4 heures » par la cellule de coordination, composée d'un - et bientôt deux - cardiologue, d'un cadre de santé, d'une infirmière et d'une secrétaire. En cas de problème, le médecin traitant, qui reste le pivot du système, est contacté.

#### ALARMES CLINIQUES

absence de renseignements > 2 jours après un RDV infirmier programmé  
> 1 mois

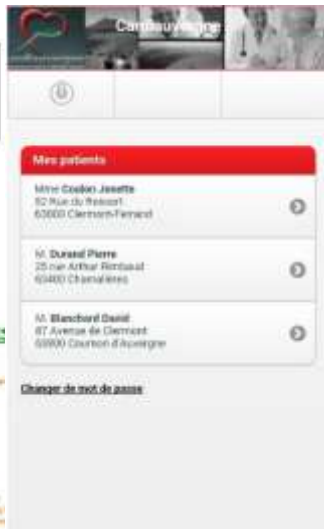
- Poids : +/- 2 Kg / poids de référence  
absence de mesure > 3 jours  
+/- 5 Kg / poids de référence
- Dyspnée : augmentation de 1 grade
- Œdèmes : augmentation de 2 grades (ou ++)
- Douleurs – Symptômes – Palpitations
- Fréquence cardiaque : > 110 - < 50
- TA : Systolique : > 140 / < 90  
Diastolique : < 60 / > 120  
Systolique > 180 / < 75

#### ALARMES BIOLOGIQUES

- Créatininémie : > 30% / valeur antérieure
- Kaliémie : < 3,5 < 3  
> 5,5 > 6,5
- Natrémie : < 130  
> 150
- NT pro-BNP : > 100% / valeur antérieure
- Hémoglobine : < 10g/dl
- INR : < 2 > 4
- Digoxinémie : > 2,5

#### ALARMES THERAPEUTIQUE

- Adaptation des doses : TEC  
R<sub>e</sub> : Positif & 3 jours et plus
- Contre-indication : - TEC  
= R<sub>e</sub>  
= Autre contre-indication  
= Autres théromédicaments : - Lithium  
- AINS - Anticoag  
- COX-2 - Sérologie  
- Digoxine - Digoxine



<http://www.hopital.fr/Hopitaux/Espace-medecin-liberal/Initiatives/Cardiauvergne-un-hopital-virtuel-pour-les-patients-insuffisants-cardiaques>

<http://www.cardiauvergne.com/>

Data Architecture - PxDATA



# Data architecture : Data war

## • Bataille autour des données de santé – L'Etat

The image displays two screenshots of government health data portals. The left screenshot shows the HealthData.gov website with search results for 'Food and Drug Administration--Recalls' and 'CMS Medicare and Medicaid EHR Incentive Program, electronic health record products used for attestation'. The right screenshot shows the HealthIT.gov website with a 'Security Risk Assessment' tool and a list of datasets.

**HealthData.gov Search Results:**

- Food and Drug Administration--Recalls**  
This feed describes all new items that are being recalled by the FDA.  
Agency: U.S. Food and Drug Administration  
Subject: Safety  
Date Released: Jan 1, 2004  
Geographic Granularity: Country
- CMS Medicare and Medicaid EHR Incentive Program, electronic health record products used for attestation**  
Data set merges information about the Centers for Medicare and Medicaid Services Medicaid EHR Incentive Programs attestations with the Office of the National Coordinator Certified Health IT Products List. This new dataset enables systematic analysis certified EHR vendors and products among those providers that have attested to the CMS EHR Incentive Programs. The data set can be analyzed by state, provider specialty, and practice setting.  
Agency: Department of Health & Human Services  
Subject: Medicare  
Date Updated: Aug 1, 2013  
Geographic Granularity: State
- Community Health Status Indicators (CHSI) to Combat Obesity, Heart Disease and Cancer**  
Community Health Status Indicators (CHSI) to combat obesity, heart disease, and components of the Community Health Data Initiative. This dataset provides key communities and encourages dialogue about actions that can be taken to improve (e.g., obesity, heart disease, cancer). The CHSI report and dataset was designed for health professionals but also for members of the community who are interested

**HealthIT.gov Security Risk Assessment:**

Download the HHS Security Risk Assessment Tool to help your practice evaluate its risks and adherence to the HIPAA Security Rule.

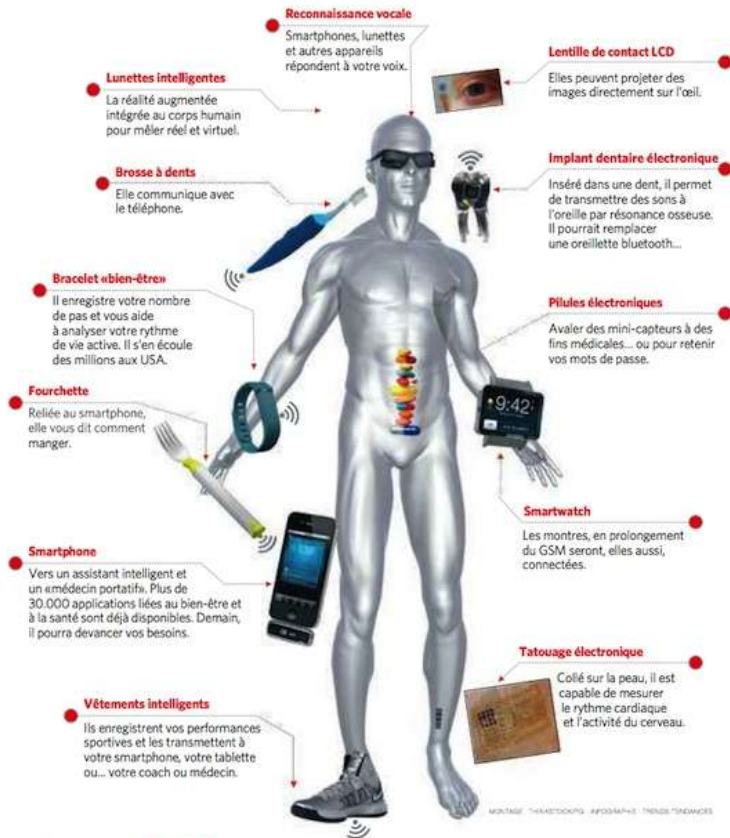
**HealthIT.gov Datasets:**

- Certificate of Need Applications: Beginning 1974
- Child and Adult Care Food Program Participation: Beginning 2007
- Child Health Plus Program Enrollment by County and Insurer: Beginning 2009
- Child Health Plus Program Enrollment: Beginning 2009
- Community Health: Obesity and Diabetes Related Indicators
- Congenital Malformation Rates (CMR): Beginning 1992
- Creating Healthy Places Intervention Locations
- Diabetes Type 2 Prevention Tips
- Donate Life Organ and Tissue Donor Registry Enrollment by County: Beginning September 2009

# Data architecture : Data war

## • Bataille autour des données de santé – Moi Quantified self (moi quantifié)

### Chaussures, lunettes, tatouages :



42 18 JUILLET 2013 | WWW.TRENDS.BE

<http://www.terminauxalternatifs.fr/tag/quantified-self/>

<http://dev.af83.com/2014/04/15/quantified-self.html>

# Data architecture : Data war

## • Bataille autour des données de santé – Nous



patientslikeme®

Username or Email Password Sign in  
Remember me I forgot

conditions, symptoms, treatments...

Live better, together!™  
Making healthcare better for everyone through sharing, support, and research

Join now  
(it's free!)

Learn from others  
Compare treatments, symptoms and experiences with people like you and take control of your health

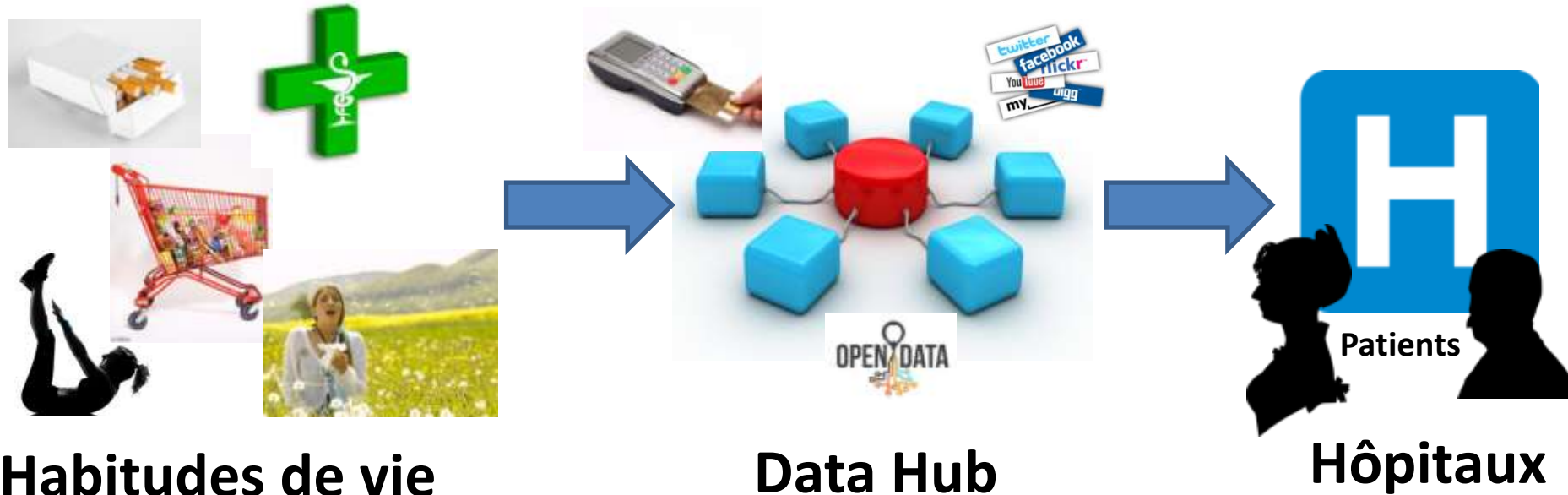
Connect with people like you  
Share your experience, give and get support to improve your life and the lives of others

Track your health  
Chart your health over time and contribute to research that can advance medicine for all

# Data architecture : Data war

## • Bataille autour des données de santé – Eux

Your Doctor Knows You're Killing Yourself. The Data Brokers Told Her



**Habitudes de vie**

**Data Hub**

**Hôpitaux**

“Information compiled by data brokers from public records and credit card transactions can reveal where a person shops, the food they buy, and whether they smoke. The largest hospital chain in the Carolinas is plugging data for 2 million people into algorithms designed to identify high-risk patients. For a patient with asthma, the hospital would be able to score how likely they are to arrive at the emergency room by looking at whether they’ve refilled their asthma medication at the pharmacy, been buying cigarettes at the grocery store and live in an area with a high pollen count”

<http://www.bloomberg.com/news/2014-06-26/hospitals-soon-see-donuts-to-cigarette-charges-for-health.html>

# Data architecture : Data war

- **Bataille autour des données de santé**



## Tedi is high level AI toy.

It corresponds differently as baby develops.

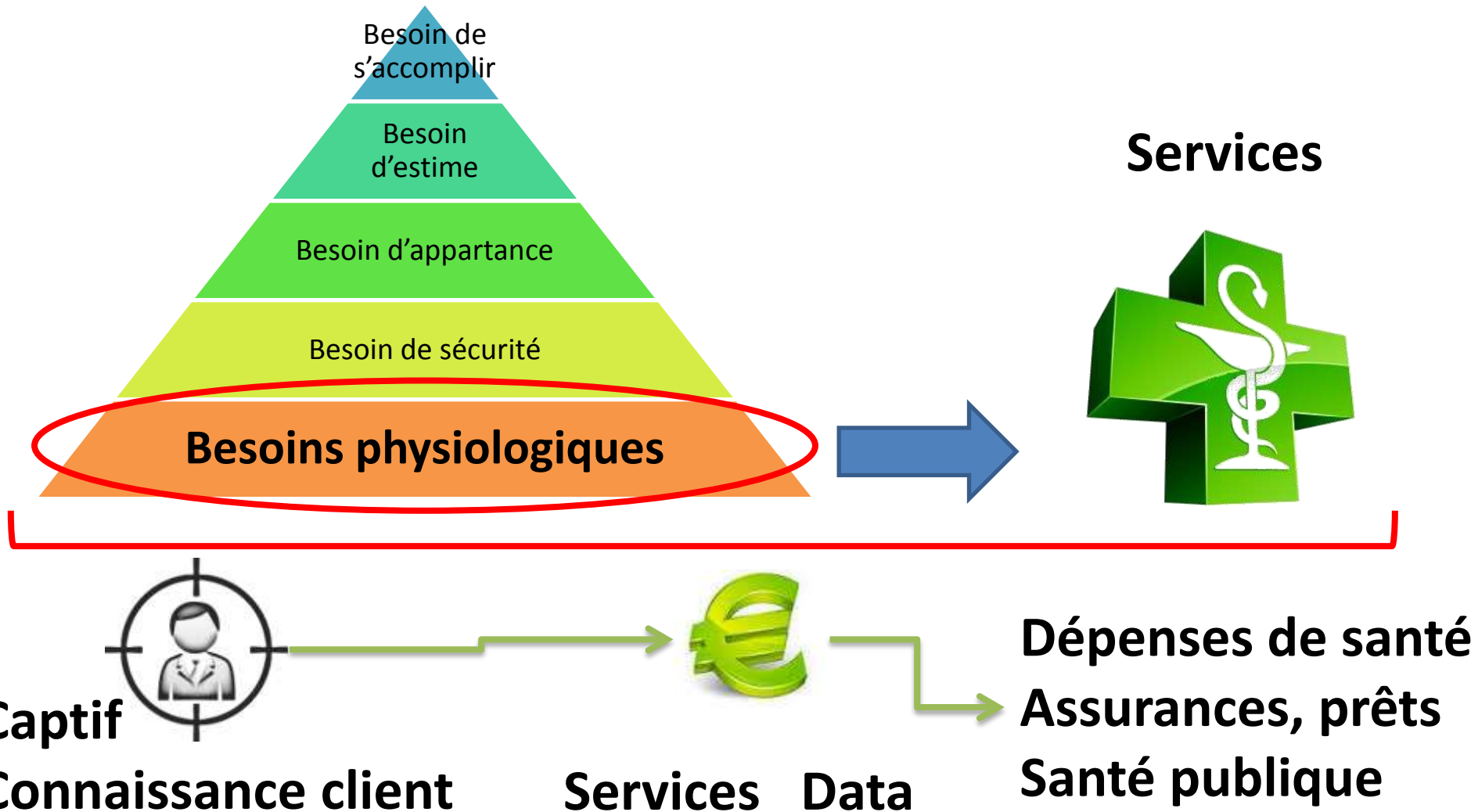
It has grip sensors to pick up the intensity of the grip, sensor which measure your baby's body temperature and pulse.

Tedi corresponds differently as baby develops



# Data architecture : Data war

- **Bataille autour des données de santé**





# Data architecture : Data war

## ● Bataille autour des données de santé

- Apps Assurance : prendre soin de votre santé
- Apps activités sportives : paramètre d'activité et de santé
- Quantified Self : capter/mesurer ses paramètres santé
- Apps pour les professionnels de la santé
- Opérateur Telecom : hébergement sécurisé de ses données de santé
- Fabricants de smartphone (ex Digital Health Samsung)
- L'Etat –Dossier médical
- Etablissements médicaux (Hôpitaux)
- Objets connectés : toilettes au Japon, pèse personne, oreillette Apple, siège du conducteur (Nissan), bracelets, ours en peluche bardé de capteurs...

*A qui confier ses données de santé .... ?*

# Data architecture : Nouveaux métiers

- Data manager
- Chief data officer
- Data analyst
- Data scientist
- Data miner
- Data strategist
- Data sourcer
- Data broker
- Data architect
- Data steward
- Data protection officer
- Data journalist
- Data coach
- Data artisan ;-) ....
- Data artist
- Et organisation : Datalab



## Le big data créera 4,4 millions d'emplois IT dans le monde d'ici 2015



Gartner estime que plus de 4 millions de jobs IT liés au big data seront créés dans le monde d'ici trois ans, mais que les compétences dans ce domaine ne seront pas suffisantes pour répondre à la demande, faute d'un système éducatif inadapte.

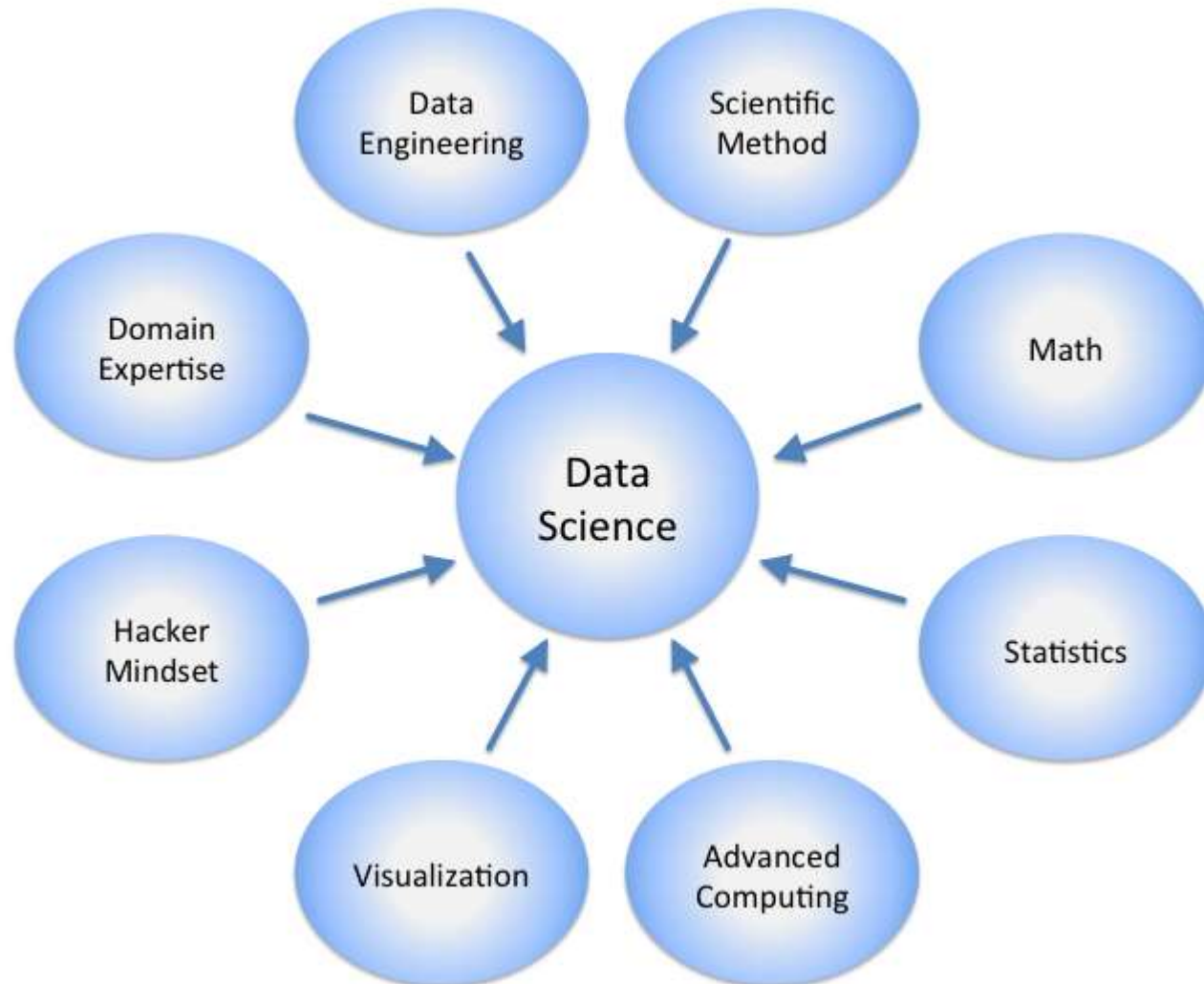
Les dépenses mondiales en IT devraient dépasser 3,7 milliards de dollars en 2013, soit une hausse de 3,8% par rapport aux 3,6 milliards attendus d'ici la fin de l'année; estime le cabinet Gartner. Mais ce sont les prévisions concernant le big data qui suscitent beaucoup

d'enthousiasme. « D'ici 2015, 4,4 millions d'emplois informatiques seront créés dans le monde pour soutenir le big data, dont 1,9 millions aux Etats-Unis », a déclaré Peter Sondergaard, senior vice-président et responsable mondial de la recherche chez Gartner lors du SymposiumITExpo qui s'est déroulé du 21 au 25 octobre à Orlando. En outre,

**LE PREMIER SALON DE  
RENCONTRE POUR  
LES METIERS  
DE LA DATA**

# Data architecture : Nouveaux métiers

- **Data Scientist**



Source : wikipedia

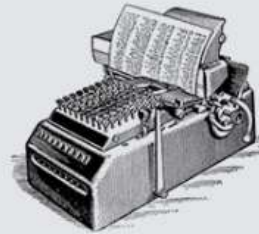
# Data architecture : Nouveaux métiers

- **Data scientist**

## Mettez votre talent de data scientist au service des entreprises

Participez à des challenges excitants, trouvez le meilleur modèle, gagnez des prix, soyez reconnus pour votre expertise, soyez visibles.

Entreprises : [Comment créer une compétition ?](#)



### Je m'inscris

J'ai lu et j'accepte les [conditions d'utilisation](#) du site [datascience.net](#)

[Je m'inscris](#)

[Je suis déjà inscrit](#)

1.



Téléchargez les données

Pour démarrer, rien de plus simple ! Lisez le descriptif du challenge proposé, récupérez les données fournies d'un simple clic, et vous êtes paré.

2.



Construisez votre modèle

Vous travaillez sur votre machine, chez vous, avec les logiciels et les méthodes que vous aimez. Quand vous pensez avoir obtenu un modèle performant, il ne vous reste plus qu'à le tester.

3.



Testez la performance

Appliquez votre modèle sur l'échantillon de test fourni, et postez les résultats sur le site. C'est tout. Il n'y a plus qu'à regarder si votre modèle est plus performant que celui des autres :-)

# Data architecture : Nouveaux métiers

- **Data scientist**



## Prédiction de la fréquentation des gares SNCF en Ile-de-France

Le challenge consiste à construire un modèle permettant d'estimer efficacement le nombre de voyageurs montant par gare pour un jour de semaine, dans les gares SNCF du réseau Ile-de-France, à partir de données open data caractérisant la gare d'un point de vue SNCF (nombre de lignes, nombre de trains par jour, services en gare, etc.), mais également de données externes (caractéristiques socio-démographiques de la ville, présence de transports urbains, etc.)

Ce challenge sera lancé **très prochainement**

**des lots**

à gagner

Ouvrir

## Rejoignez un challenge



## Estimation du prix de vente de véhicules d'occasion

Il s'agit de fournir un modèle d'estimation du prix de vente de véhicules d'occasion, à une maille d'analyse la plus fine possible, à partir du descriptif et des caractéristiques du véhicule, des éléments connus de sa vie (kilométrage...), et des historiques de transaction.

 il reste **22 jours**

**12 000 €**

à gagner

Rejoindre



## Modéliser l'impact des conditions économiques sur le comportement des clients

L'objectif d'AXA France est de disposer de modèles prédictifs permettant d'appréhender le comportement des clients selon différentes situations de marché (évolution des taux du marché, des prix de l'immobilier...).

**10 000 €**

à gagner

Rejoindre

<http://datascience.net>

# Data architecture : Nouveaux métiers

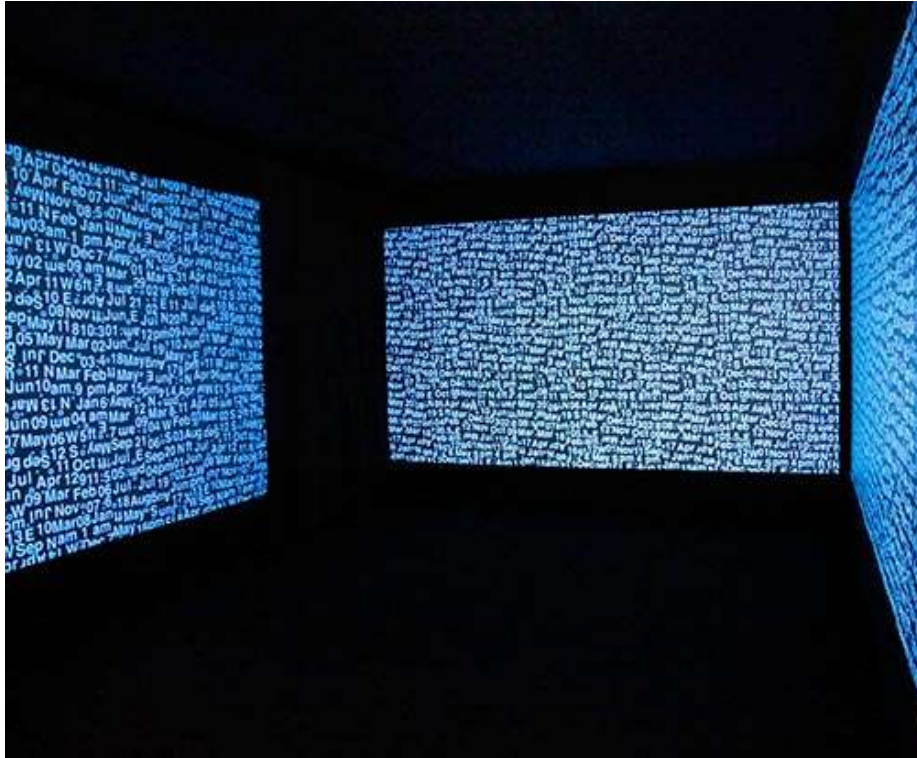
- **Data journalisme**



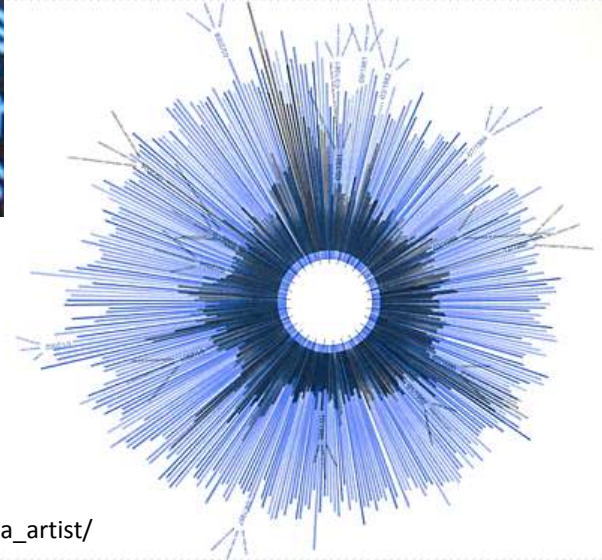
<http://www.globaleditorsnetwork.org/programmes/data-journal-awards/>

# Data architecture : Nouveaux métiers

- Data artist



[http://www.cccb.org/en/album\\_descobreix--46304?idg=46304;sn=12](http://www.cccb.org/en/album_descobreix--46304?idg=46304;sn=12)



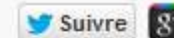
[http://islamicartsmagazine.com/magazine/view/raw\\_data\\_1999-2013\\_exhibition/](http://islamicartsmagazine.com/magazine/view/raw_data_1999-2013_exhibition/)

[http://www.datanami.com/2012/03/26/inside\\_the\\_mind\\_of\\_a\\_data\\_artist/](http://www.datanami.com/2012/03/26/inside_the_mind_of_a_data_artist/)

# Data architecture : Nouveaux métiers



Publié par Julien L., le Mardi 08 Juillet 2014



## La France espère 137 000 emplois en 2020 grâce au Big Data

L'émergence du Big Data en France est suivie avec attention par le ministère de l'économie et des finances, car ce nouveau secteur pourrait contribuer à créer 137 000 emplois d'ici 2020. Un plan visant à encourager l'adoption du Big Data a par ailleurs été validé la semaine dernière.

5<sup>e</sup> comité de pilotage des 34 plans de la nouvelle France industrielle

<http://www.numerama.com/magazine/29944-la-france-espere-137-000-emplois-en-2020-grace-au-big-data.html>

### BIG DATA



Un marché estimé à

**9 Md€**

en France en 2020

Un potentiel de

**137 000**

emplois créés  
ou consolidés

<http://proxy-pubminefi.diffusion.finances.gouv.fr/pub/document/18/17721.pdf#page=9>

Données informatiques, lignes de code



# Data architecture : Nouveaux métiers

*Et si vous recherchez un emploi*



MARYLISE LEBRANCHU  
MINISTRE DE LA DÉCENTRALISATION,  
DE LA RÉFORME DE L'ÉTAT ET DE LA FONCTION PUBLIQUE

*Communiqué de presse*

*Communiqué de presse*

Paris, le 21 mai 2014

## **Ouverture des données publiques : création de la fonction d'administrateur général des données (*chief data officer*)**

A l'occasion d'une communication en Conseil des ministres ce mercredi 21 mai, Marylise LEBRANCHU, ministre de la décentralisation, de la réforme de l'Etat et de la fonction publique, a dressé un bilan de la politique d'ouverture des données publiques menée par le Gouvernement depuis deux ans.

# Data architecture

- **1ère partie : Data World**
  - **Point de bascule : signaux autour des données**
  - **Tour d'horizon de thèmes Data : Open Data, Data Vizualisation, Big Data, Data policy, « AGL » Data, Data provider, Master Data, MetaData, Data Lake...**
  - **Du Data game à la Data war ... quel sera le monde d'après ?**
  - **Je suis Data Scientist, Data Manager... inventaire des métiers Data**

- **2ème partie : Data architecture**
  - **Nouveau paradigme ?**
  - **Intérêt d'un apport méthodologique ?**
    - **L'initiative PxDATA**

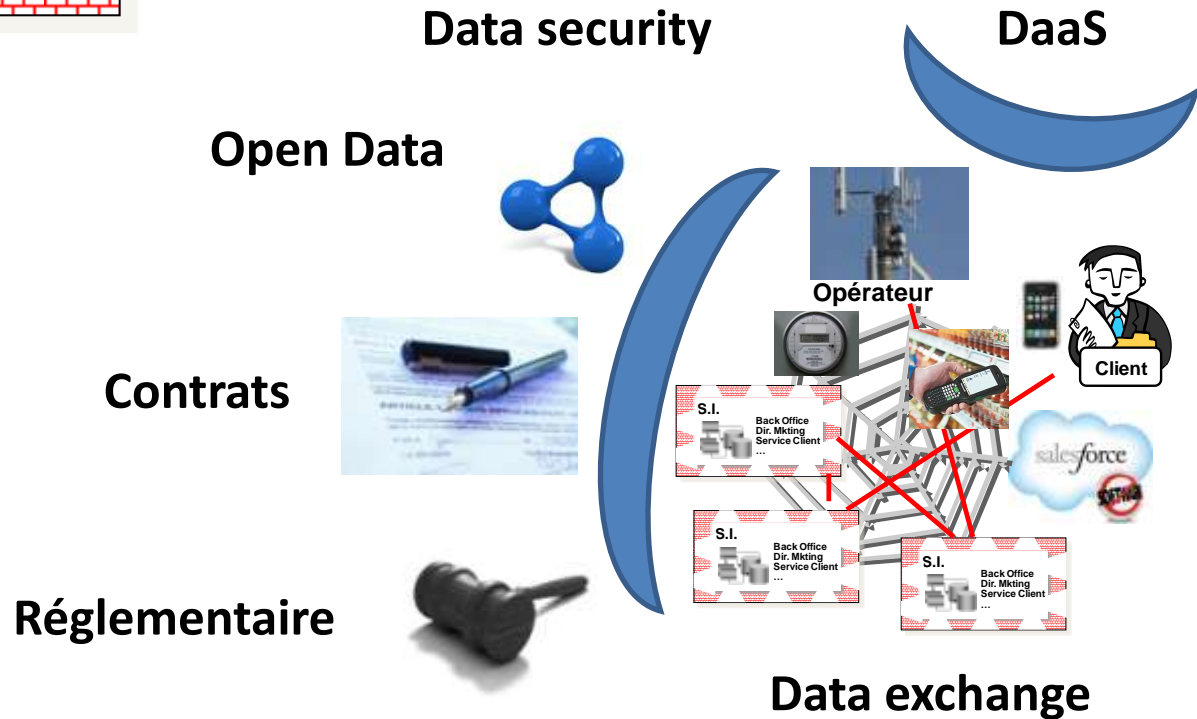
# Data architecture : Nouveau paradigme ?

## De la muraille de chine à l'ouverture



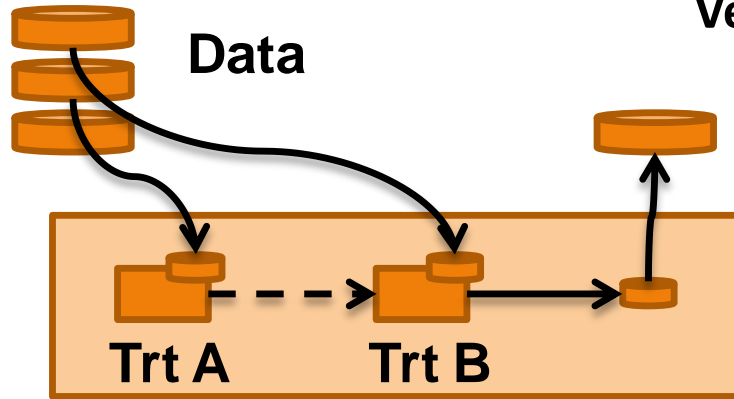
**Historique** : Les S.I. et leurs données sont circonscrit dans les murs de l'entreprise, voire au sein des applications.

**Data centric** : Les S.I. sont ouverts (entreprise étendue) et les données voyagent de S.I. en S.I., de client en client... Les sources ne sont plus uniquement au sein de l'entreprise



# Data architecture : Nouveau paradigme ?

## Du traitement centric au data centric Vers les données autonomes

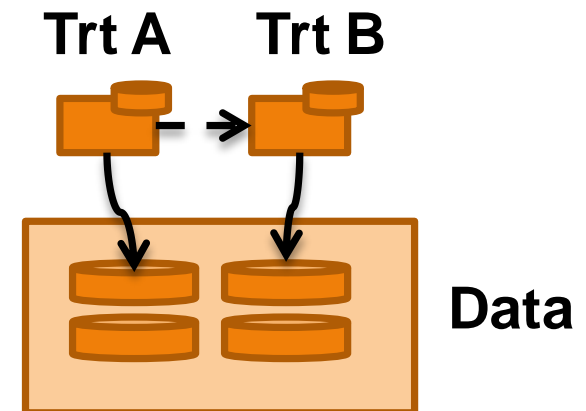


**Classique** : Les données sont dans des silos de traitements (elles appartiennent à des applications pour un usage donné).

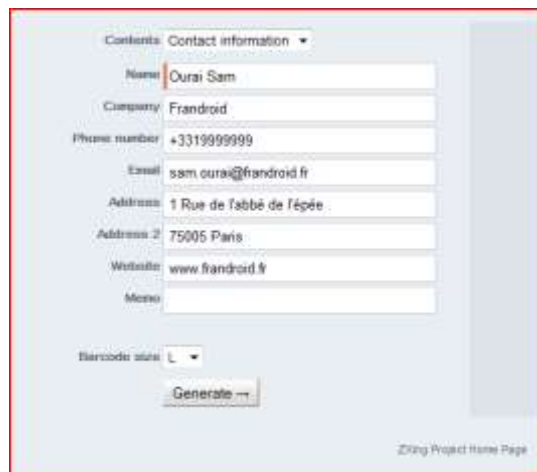
Chaque besoin transverse amène à s'échanger les données.

**Data centric**: Les données sont les fondations du S.I. (elles ne dépendent plus des applications). Les traitements (agiles) exploitent les données. On transporte les traitements vers les données.

On pense données avant traitement.



# Data architecture : Nouveau paradigme ?



Contents: Contact information ▾

Name: Ourai Sam

Company: Frandroid

Phone number: +3319999999

Email: sam.ourai@frandroid.fr

Address: 1 Rue de l'abbé de l'épée

Address 2: 75005 Paris

Website: www.frandroid.fr

Barcode status: L ▾

Generate →

Zing Project Home Page

**La capture des données fait l'objet d'une stratégie explicite**

**Classique** : dans les environnements de gestion, saisie, collecte des données via les applications transactionnelles.  
Peu de métadonnées sont recueillies.

**Data centric** : possibilité d'une stratégie volontariste de capture de données (mobile, collecte applicative, capteurs, « enquête », cartes...) ... approche de type télémétrie  
La géolocalisation devient naturelle.  
Les métadonnées ont autant d'importance que les données recueillies.



<http://de.wikipedia.org/wiki/Sensor>



# Data architecture : Nouveau paradigme ?



La capture des données fait l'objet d'une stratégie explicite

<https://www.uspsoig.gov/sites/default/files/document-library-files/2014/rarc-ib-14-002.pdf>

“The possibility of equipping the postal network (vehicles, mailboxes, mail pieces and parcels, sorting centers, etc.) with low-cost sensors will exponentially expand the capability of postal operators to collect valuable data. This new rich data sources could help the Postal Service improve operational performance, customer service, create new products and services, and support more efficient decision-making processes. The “Internet of Postal Things”, experts say, could also have a positive spillover effect on other adjacent non-postal sectors, as the information collected by and for the Postal Service could be useful to others.”

# Data architecture : Nouveau paradigme ?

Du traitement centric au data centric  
Vers les données autonomes



versus



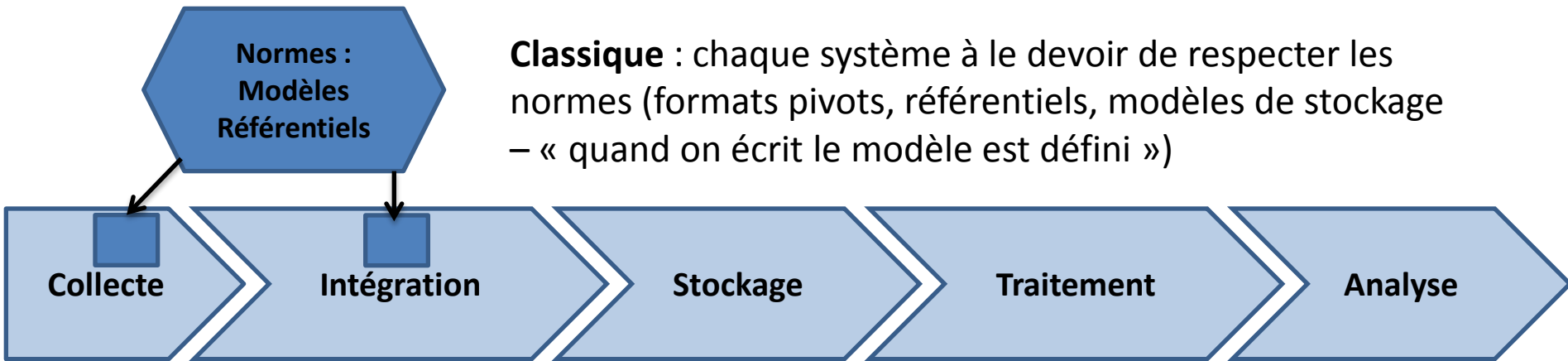
Every airplane in flight can act as a sensor to help estimate large-scale wind conditions.

**Les données ont une seconde, troisième...vie.**

<http://research.microsoft.com/en-us/news/features/windflow-060314.aspx>

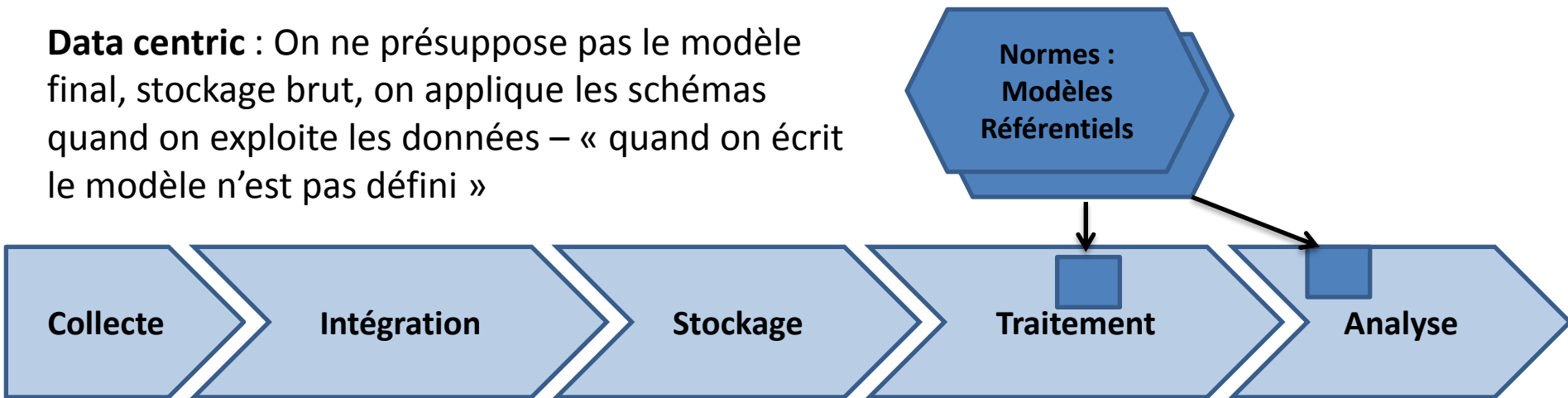
# Data architecture : Nouveau paradigme ?

## De la recherche de la standardisation à l'hétérogénéité



**Classique** : chaque système a le devoir de respecter les normes (formats pivots, référentiels, modèles de stockage – « quand on écrit le modèle est défini »)

**Data centric** : On ne présuppose pas le modèle final, stockage brut, on applique les schémas quand on exploite les données – « quand on écrit le modèle n'est pas défini »





# Data architecture : Nouveau paradigme ?



**On ne répond plus directement à un besoin,  
on met à disposition une capacité à**

**Classique** : quel est le cahier des charges ?  
Décrire la solution

**Data centric** : mettre en place les moyens pour  
offrir une capacité à.  
Découvrir, explorer, essayer afin d'identifier une  
solution.



# Data architecture

- **Ouverture**
- **Data centric – penser data sans traitement a priori**
- **Stratégie de capture**
- **2<sup>ème</sup> vie des données**
- **Standardisation à la consommation**
- **Capacité à**

**Est-on face à un nouveau paradigme ?**

# Data architecture

- Océan de données : un nouveau cap à prendre
  - Nouveau terrain de jeu .... Nouveaux métiers
  - Data eldorado
  - Innovation
  - ... et au-delà impact sociétal et économique vers des changements profonds... ruptures



# Data architecture

- Océan de données : un nouveau cap à prendre
  - Qui est l'architecte ?
  - Que pourrait-on attendre d'une méthode ?

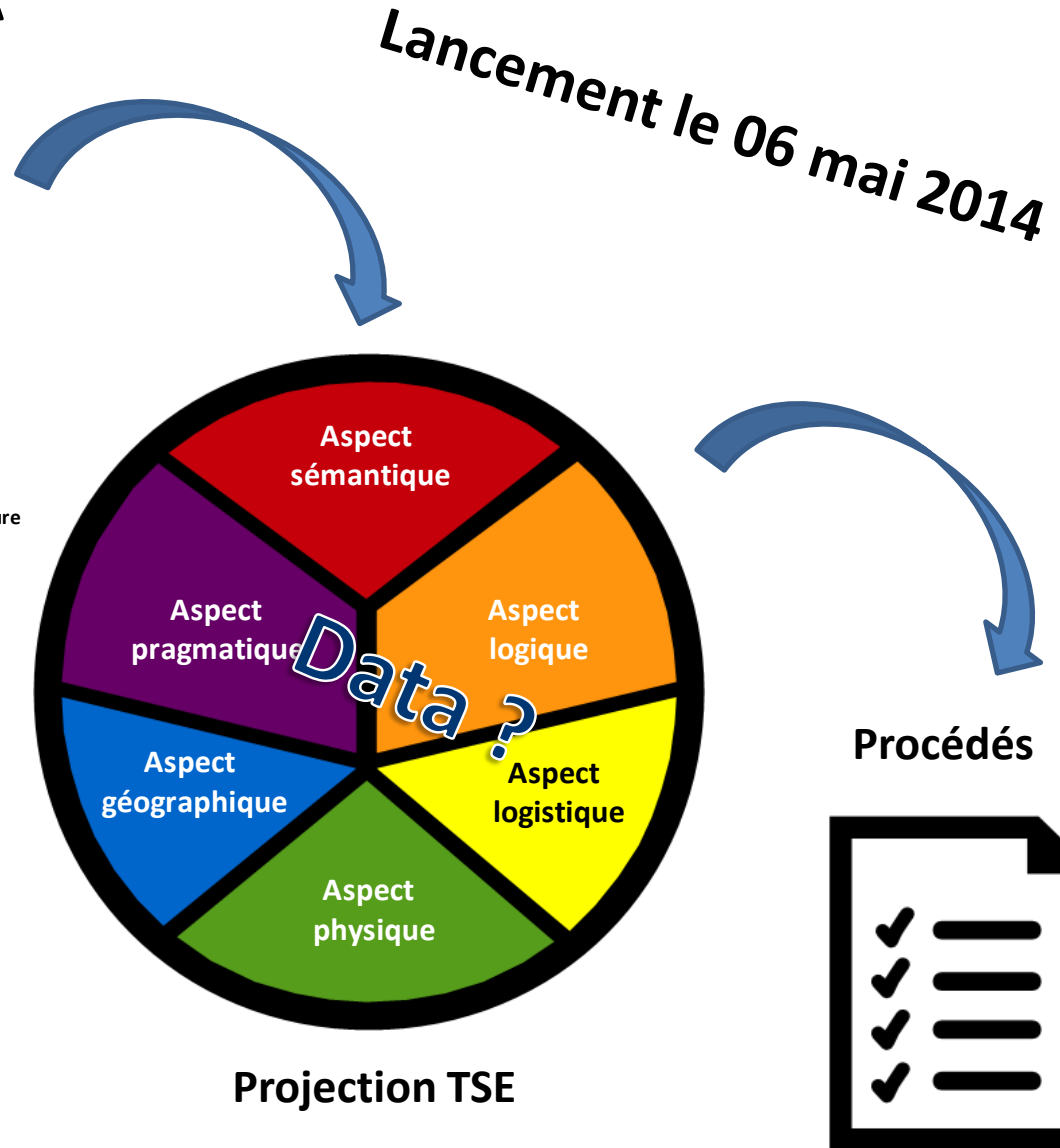


# Data architecture

## • L'initiative PxDATA

- Data sourcing, achat de données
- Open Data
- Réglementaire sur les données
- Big Data
- Data as a service
- SI Data centric socle/puit de données
- BI nouvelle génération – Data analysis
- Référentiels de données
- Data quality
- Data security
- Echanges de données – Data exchange
- Data visualisation
- Data storytelling
- Data policy
- Smart Data
- Data virtualization
- Master Data
- Data scraping
- AGL Data
- Data Lake
- Data provider
- Data Hub
- Data grid
- Unstructured Data
- Data API
- Data crawler
- Data broker
- Data marketplace
- Data cleansing
- Data science
- Data warehouse
- Metadata
- Linked Data
- Data portal
- Enterprise Data Architecture
- Data gouvernance

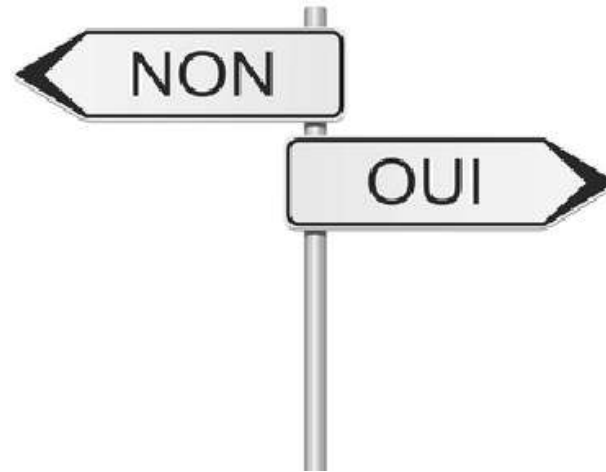
### Terminologie DATA



### Projection TSE

# Data architecture

- L'initiative PxDATA



## Discussion