

# Entreposer/ Classifier/ Penser

## Les Big Data: fragment d'un Net Savoir

Jean-Christophe Thalabard

MAP5, UMR CNRS 8145, Univ. Paris Descartes & APHP

18/07/2013

# Plan proposé

- Introduction : pourquoi intervenir ?
- Big Data : une question de terminologie(s)
- Big Data : Entreposer
- Big Data : Classer
- Big Data : Analyser- Penser
- Big Data : vos dépôts m'intéressent
- Quelques réflexions en guide de conclusion

"Le Cygne noir, la puissance de l'imprévisible" Partie II "Les prévisions sont tout bonnement impossibles" (Taleb, 2007)

*Quand je demande aux gens de citer trois technologies récentes qui exercent un impact considérable sur notre monde, ils pensent généralement à l'informatique, à l'Internet et au laser. Toutes trois n'avaient été ni planifiées, ni prévues ; elles furent décrites lorsqu'elles apparurent et continuèrent à l'être bien après qu'on eut commencé à les utiliser. Elles ont eu des conséquences majeures. C'était des Cygnes Noirs...*

# Je me souviens...

Un monde balbutiant réservé aux scientifiques, ingénieurs

- Des langages abstraits
- Des machines volumineuses
  - 1969- 1974 : les "monstres" IBM, Bull...
  - 1974- 1979 : Les conditions du dialogue homme- machine s'affinent
  - Supports et intermédiaires s'effacent : carte → ruban → console
  - Des modes d'emploi lourds et pesants d'accès ± facile !

1976 : Gilles de Gennes et les cristaux liquides ou l'imagination créatrice et visionnaire

## Je me souviens de l'accouchement... d'une discipline

- Un acteur majeur en France : **Pr. François Grémy**
- Le scepticisme des milieux médicaux : menace sur leurs savoirs ?
- 1982- 1983 : **IBM PC** *versus* Plan Mitra 15 et la tentation encore centralisatrice
- Intelligence artificielle et médecine, le centre mondial informatique (JJSS, JLFB). Le projet Medicine (Carnegie Mellon, Pittsburgh)
- La multiplication des développeurs
- Les bases de données autonomes DBASE, etc...
- L'instabilité des outils, des supports, des hommes...

Je me souviens aussi ...d'un échange avec un directeur d'**IBM**...

# Je me souviens...des balbutiements des réseaux locaux/ distants

- 1985 : **IBM** et le projet EARN : l'ouverture au monde académique non stratégique
- Le minitel et le réseau sentinelle : premier réseau français de connaissance sur maladies grippales
  - La cartographie renouvelée (C. Nicolle) : krigeage, distances temporelles...
  - Quel(s) objectif(s) ? Un mode d'approche nouveau en Santé Publique
  - Une nouvelle forme d'expertise avec des acteurs nouveaux
- Icônes et souris (**Douglas Engelbart**, 1964 !)
- CITI2 et premières banques d'information (P Lebeux, Compiègne)
- Le projet HEGP & Newton

# Hôpital public : une transformation...bien lente



Lieu de tradition



d'adaptation difficile



à la modernité

# Je me souviens...Un savoir plus ou moins organisé et mémorisé

- Le contact des maîtres et leurs paroles
- La force des écoles
- Le livre plus que le journal, la barrière de la langue
- La circulation hiérarchique des savoirs
  - Les abonnements intra- service, l'astreinte des *Currents contents*
  - L'échappée vers les bibliothèques médicales
  - L'accès hiérarchisés aux services/ vignettes/ INIST/...
  - Journal Clubs et ses contraintes
  - Demandes de tirés à part, contraintes, rêveries et occasions de partages distants/ locaux



# L'écran entre dans le décor

- De l'écran cathodique à l'écran plat
- L'entrée par le personnel non soignant
  - Le secrétariat : l'abandon difficile de la machine à écrire
  - La perte d'une maîtrise lentement acquise et des privilèges associés.
  - La transition des machines électriques : IBM, Olivetti. L'espace sonore se transforme.
  - Les demandes de formations, les maux induits, les nouveaux accessoires
- Le casse- tête de l'archivage, des incompatibilités matérielles, des services d'assistance

# Personnels soignants et Outils informatiques

- La dotation par ordre hiérarchique : l'impossible avec et ses conséquences
- La connexion contrôlée pour quel(s) secret(s)
- La défiance vis à vis d'accessoires ludiques : lecteurs CD, clés USB
- Le déploiement vers l'ensemble du personnel soignant
  - Des raisons économiques, le codage des actes...
  - La demande des nouvelles générations : Directions/ Soignants
  - Des contraintes mieux acceptées...car moins coûteuses
  - Les agendas électroniques : **Palm, Iphone**
  - La multiplication des courriers, médecin secrétaire. La question de l'authentification et de la responsabilité

# La connection internet, le service éclaté

- La hiérarchie des attributions
- L'inadaptation des locaux
- Filtres et Spams...
- La mise en commun des ressources bibliographiques & l'accès on line aux revues
- La généralisation des connections personnelles

# Publier pour happy few *versus* être lu par le plus grand nombre

- La multiplication des informations/ les recherches systématiques
- Une "modèle" économique qui vacille & un monde scientifique qui se re- configure
  - Des chercheurs et financements majoritairement publics/ fondations
  - Des media privés
  - Coûts pour publier  $\pm$  élevés
  - Coûts d'accès à ses propres productions
- Des solutions nouvelles
  - Abonnements collectifs : des crédits retirés pour être mutualisés
  - Payer pour rendre accessible gratuitement au plus grand nombre : **PloS**, BMC, etc...
  - Archives ouvertes (le modèle Archiv)
  - PED et fondations (Soros)

# Le savoir synthétisé : le mouvement EBM et ses zélotes...

Home About the CEBM Publications Training & Education **EBM Tools** Resource Centre Blog Contact
Search





**CEBM**  
CENTRE FOR EVIDENCE BASED MEDICINE



UNIVERSITY OF  
**OXFORD**

Quickfinder Go

[CEBM > EBM Tools > Overview](#)

|   |   |   |   |
|---|---|---|---|
| <ul style="list-style-type: none"> <li style="background-color: #e0e0e0; padding: 2px; margin-bottom: 2px;">Asking Focused Questions</li> <li style="background-color: #e0e0e0; padding: 2px; margin-bottom: 2px;">Finding the Evidence</li> <li style="background-color: #e0e0e0; padding: 2px; margin-bottom: 2px;">Critical Appraisal</li> <li style="background-color: #e0e0e0; padding: 2px; margin-bottom: 2px;">Making a Decision</li> <li style="background-color: #e0e0e0; padding: 2px; margin-bottom: 2px;">Evaluating Performance</li> <li style="background-color: #e0e0e0; padding: 2px; margin-bottom: 2px;">Designing Research</li> </ul> | <div style="background-color: #000080; color: white; padding: 2px; margin-bottom: 5px;"><b>Tools for each step of the EBM process</b></div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <div style="background-color: #000080; color: white; padding: 5px; text-align: center; margin-bottom: 5px;"> <span style="font-size: 2em; font-weight: bold;">?</span> </div> <p><b>Asking Focused Questions</b></p> <p>The translation of uncertainty into an answerable question</p> </div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <div style="background-color: #000080; color: white; padding: 5px; text-align: center; margin-bottom: 5px;"> <span style="font-size: 2em; font-weight: bold;">🔍</span> </div> <p><b>Finding the Evidence</b></p> <p>Systematic retrieval of the best evidence available</p> </div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <div style="background-color: #000080; color: white; padding: 5px; text-align: center; margin-bottom: 5px;"> <span style="font-size: 2em; font-weight: bold;">😊</span> </div> <p><b>Critical Appraisal</b></p> <p>Critical appraisal of evidence for validity, clinical relevance and applicability</p> </div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <div style="background-color: #000080; color: white; padding: 5px; text-align: center; margin-bottom: 5px;"> <span style="font-size: 2em; font-weight: bold;">↔</span> </div> <p><b>Making a Decision</b></p> <p>The application of results in practice</p> </div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <div style="background-color: #000080; color: white; padding: 5px; text-align: center; margin-bottom: 5px;"> <span style="font-size: 2em; font-weight: bold;">⚡</span> </div> <p><b>Evaluating Performance</b></p> <p>Audit and feedback of clinical practice to improve performance</p> </div> <div style="padding: 5px;"> <div style="background-color: #000080; color: white; padding: 5px; text-align: center; margin-bottom: 5px;"> <span style="font-size: 2em; font-weight: bold;">📐</span> </div> <p><b>Designing Research</b></p> <p>Designing trials or assessing research design</p> </div> | <div style="background-color: #000080; color: white; padding: 2px; margin-bottom: 5px;"><b>Popular Tools</b></div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <p><b>Finding the Evidence</b></p> <ul style="list-style-type: none"> <li>Levels of evidence</li> <li>Levels of Evidence #2</li> </ul> </div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <p><b>Explanations &amp; Examples</b></p> <ul style="list-style-type: none"> <li>Pre-test probability</li> <li>SpPins and SnNouts</li> <li>Likelihood Ratios</li> <li>NNTs</li> </ul> </div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <p><b>Calculators</b></p> <ul style="list-style-type: none"> <li>All-purpose 2x2 table analyser</li> <li>Interactive nomogram for interpreting likelihood ratios</li> <li>Excel spreadsheet for calculating 95% confidence intervals (by Dan Tandberg)</li> <li>Excel workbook for estimating hazard ratios from survival curves (implements method by Parmar et al.)</li> </ul> </div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <p><b>Research Design</b></p> <ul style="list-style-type: none"> <li>PaT Plot</li> </ul> </div> <div style="padding: 5px;"> <p><b>CATmaker</b></p> </div> | <div style="background-color: #000080; color: white; padding: 2px; margin-bottom: 5px;"><b>What's New</b></div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <p><b>Levels of Evidence #2</b></p> <p>Be the first to review our updated 'Levels of Evidence' table.</p> <p><a href="#">More Information &amp; FREE Download</a></p> </div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <p><b>Teaching Evidence-based Practice Workshop</b></p> <p>(5 days) 5th-9th September 2011</p> <p><a href="#">More Information</a></p> </div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <p><b>CEBM Workshop Videos</b></p> <div style="display: flex; align-items: flex-start; margin-top: 5px;"> <div style="flex: 1;"> <p><b>Carl Heneghan</b></p> <p>Diagnostic Tests</p> </div> <div style="flex: 1;">  </div> </div> </div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <p><b>Rafael Perera</b></p> <p>Appraisal of Clinical Trials</p>  </div> <div style="border-bottom: 1px solid #000080; padding: 5px;"> <p><b>Related Links</b></p> <p><a href="#">Understanding EBM</a></p> <p><a href="#">Powerpoint training presentations</a></p> <p><a href="#">EBM resources</a></p> </div> |
|---|---|---|---|



# Pour une diffusion des savoirs

**PLoS** PUBLIC LIBRARY of SCIENCE LEADING A TRANSFORMATION IN RESEARCH COMMUNICATION

HOME ABOUT PUBLICATIONS PUBLISH SUPPORT US

## Welcome to PLoS

We are a nonprofit [publisher](#) and [advocacy organization](#). Our [mission](#) is to accelerate progress in science and medicine by leading a transformation in research communication. [Everything](#) that we publish is [open-access](#) - freely available online for anyone to use. Sharing research encourages progress, from protecting the biodiversity of our planet to finding more effective treatments for diseases such as cancer.

[More >>](#)

[Our mission](#) [Publish](#) [Open Access](#) [Shop](#)

### Why publish with us?

- ✓ Rapid publication
- ✓ Unlimited readership
- ✓ High impact

### Latest from PLoS

#### PLoS Expands Mission

SEPTEMBER 28, 2011

Since 2000, PLoS's mission has been to make "the world's scientific and medical literature a freely available public resource." But more broadly, our goal has always been to reinvent research

Signup to get PLoS news

[Sign up](#)

Connect with us on:

[Twitter](#) [LinkedIn](#) [Facebook](#)

## C Bernard renouvelé : le nouveau laboratoire dans l'hôpital


- Des pratiques appropriées : analyse bibliométrique, lecture critique
- La recherche sur les recherches : l'approche méta- analytique
- Les recherches *in silico*
- De nouvelles richesses, de nouveaux défis
  - Banques de tissus biologiques & collections : une nouvelle *économie* qui s'organise et se structure/ les consortia/ les cohortes géantes (Oakland, Gazelle, NOWAC, E3N, Eden...)
  - Les croisements de bases de données
  - Les soutiers -"réducteurs" de dimensionnalité

# Informatisation du lieu de soin

- Une obligation contrôlée uniquement *pour l'exercice libéral*. Des résistances variables selon modes exercice
- Une source de connaissance monnayable et monnayée !
- Des initiatives stimulantes : Observatoire MG (SFMG). Les groupes de pairs
- Quid de l'exercice hospitalier ?



# L'union fait la force : l'exemple de la SFMG



omg.sfm.org

## Observatoire de la Médecine Générale

Informations épidémiologiques sur les pathologies et leur prise en charge en ville

>> [Contact](#) | [Presse](#)  
>> [Le Bulletin de l'OMG](#)

mercredi 23 novembre 2011

■ Présentation ■ Méthode ■ Dictionnaire ■ Réseau ■ Publications

🔍 **Données en consultation pour : HYPOTHYROIDIE**  
[Voir la définition](#) [Voir les Recommandations de Pratiques Francophones](#)

■ Généralités ■ Patients ■ Actes

**VOTRE SÉLECTION**

ANNEE  
2009

REGION  
TOUTES

**VALIDER**

**Données générales**

Répartition des patients par tranche d'âge

Répartition des actes par tranche d'âge

|                 | Nbe total | Par médecin |  |
|-----------------|-----------|-------------|--|
|                 |           | Moyenne     | Bornes interquartile (Q25 ; Médiane ; Q75) |
| <b>Patients</b> | 1346      | 26.4        | ( 14 ; 22 ; 36 )                           |
| <b>Actes</b>    | 3541      | 69.4        | ( 38 ; 57 ; 97 )                           |

🔍 [Pour faire une demande des données particulières, cliquez ici.](#)

**Données**

🔍 **Moteur de recherche**  
Pour trouver des données sur un diagnostic, tapez un ou plusieurs mots-clés

🔍 **Sélecteurs**  
Utilisez les menus déroulants pour faire une recherche par résultat de consultation.

**DOMAINE**  
Métabolisme-Endocrinologie-Ani

**RÉSULTAT DE CONSULTATION**  
hypothyroïdie

🔍 **Voir aussi**  
[asthénie - fatigue](#)  
[goitre](#)

🔍 **TOP 50**  
Pour visualiser les diagnostics les plus fréquents en médecine générale

🔍 **Glossaire**

<< < > >> 🔍

# Apprentissage médical renouvelé : le patient virtuel



**ilumens**  
mieux former pour mieux soigner

ACCUEIL I LUMENS NUMÉRIQUE SIMULATION FORMATIONS RECHERCHE COMMUNICATION PARTENARIAT PRATIQUE BLOG

Formation médicale des Professionnels de Santé

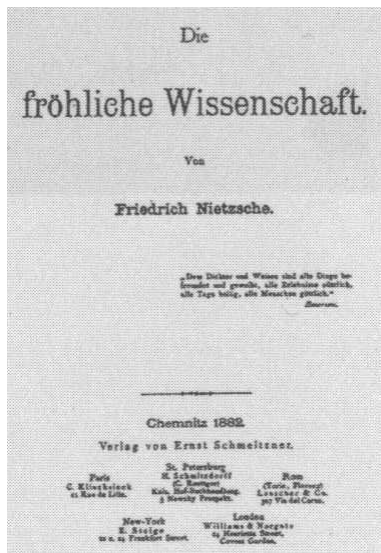
A travers ses programmes de simulation sur mannequin, ilumens participe à l'amélioration de la formation des professionnels de Santé. Ici, des étudiants s'entraînent sur un scénario d'arrêt cardiaque.

[Lire la suite](#)

⇒ Serious games (Nice!), Foldit (Firas et al., 2011, fol, 2008)

# Big Data et le trépied d'une connaissance

# Pour une science joyeuse



## JL Borges "De la rigueur de la science", 1999

"En cet empire, l'art de la cartographie fut poussé à une telle perfection que la carte d'une seule province occupait toute une ville et la carte de l'empire toute une province. Avec le temps, ces cartes démesurées cessèrent de donner satisfaction et les Collèges de Cartographes levèrent une carte de l'empire, qui avait le format de l'empire et qui coïncidait avec lui, point par point. Moins passionnées pour l'étude de la Cartographie, les générations suivantes réfléchirent que cette carte dilatée était inutile et, non sans impiété, elle l'abandonnèrent à l'inclémence du soleil et des hivers. Dans les déserts de l'ouest, subsistent des ruines très abimées de la Carte. Des animaux et des mendiants les habitent. Dans tout le pays, il n'y a plus d'autre trace des disciplines géographiques (Suarez Miranda, Viajes de Varones Prudentes, Livre IV, Chapitre XIV, Lérída, 1658)"

## G Perec, Tentative d'épuisement d'un lieu parisien

"Il y a beaucoup de choses place Saint-Sulpice, par exemple : une mairie, un hôtel des finances, un commissariat de police, trois cafés dont un fait tabac, un cinéma, une église à laquelle ont travaillé Le Vau, Gittard, Oppenord, Servandoni et Chalgrin et qui est dédiée à un aumônier de Clotaire II qui fut évêque de Bourges de 624 à 644 et que l'on fête le 17 janvier, un éditeur, une entreprise de pompes funèbres, une agence de voyages, un arrêt d'autobus, un tailleur, un hôtel, une fontaine que décorent les statues des quatre grands orateurs chrétiens (Bossuet, Fénelon, Fléchier et Massillon), un kiosque à journaux, un marchand d'objets de piété, un parking, un institut de beauté, et bien d'autres choses encore. Un grand nombre, sinon la plupart, de ces choses ont été décrites inventoriées, photographiées, racontées ou recensées. Mon propos dans les pages qui suivent a plutôt été de décrire le reste : ce que l'on ne note généralement pas, ce qui ne se remarque pas, ce qui n'a pas d'importance : ce qui se passe quand il ne se passe rien, sinon du temps, des gens, des voitures et des nuages."

# JL Borges "L'empire céleste de la connaissance bénévole" de Franz Kuhn

- Ceux qui appartiennent à l'empereur
- Les embaumés
- Ceux qui sont dressés
- Les cochons têtards
- Les sirènes
- Un être fabuleux
- Des chiens errants
- Ceux inclus dans cette classification
- Ceux qui tremblent comme s'ils étaient fous
- Les non dénombrés
- Ceux qui ont été extraits avec une brosse en poils de chameau.
- **etc,etc,...**
- Ceux qui viennent juste de casser le pot de fleurs
- Ceux qui de loin ressemblent à des mouches

# Perec, "Penser/ Classifier"

- Animaux sur lesquels on parie
- Animaux dont la chasse est interdite entre le 01/04 et le 15/09
- Baleines échouées
- Animaux soumis à une quarantaine aux frontières
- Animaux en possession partagée
- Animaux empaillés
- **Etcetera**
- Animaux susceptibles de transmettre la lèpre
- Chiens d'aveugle
- Animaux reçus comme dons dans le cadre du décès d'un oncle paternel
- Animaux qui peuvent être transportés en cabine
- Chiens errants sans collier
- Anes
- Juments avec un poulain



## Perec, "Penser/ Classer", 1985

"Que me demande-t-on, au juste ? Si je pense avant de classer ? Si je classe avant de penser ? Comment je classe ce que je pense ? Comment je pense quand je veux classer ? (...) Tellement tentant de vouloir distribuer le monde entier selon un code unique ; une loi universelle régirait l'ensemble des phénomènes : deux hémisphères, cinq continents, masculin et féminin, animal et végétal, singulier pluriel, droite gauche (...) Malheureusement ça ne marche pas, ça n'a même jamais commencé à marcher (...) N'empêche que l'on continuera encore longtemps à catégoriser tel ou tel animal selon qu'il a un nombre impair de doigts ou des cornes creuses"

# Echelles de compréhension des phénomènes humains

- La question universelle de la causalité individuelle/ collective (Les lumières, 1789 !)
- Echelle individuelle et ses limites
  - Les emprunts à psychométrie (méta- analyses, 1976)
- Echelle société
  - Emile Durkheim
    - La démarche causale en sciences sociales
    - Les régularités quantifiées
  - Erwin Goffman
    - La démarche individualisante
    - La stigmatisation locale
    - Une micro- dissection des rapports humains.
  - Les réseaux sociaux
    - Disparition ancrage géographique
    - Disparition ancrage corporel : second life et les réseaux virtuels

# Big Data : une question de terminologie

# Big Data : une question de terminologie

- 2 notions : collection massive de données × croisements de données
  - En filigrane : exploitation de données collectées à d'autres fins
- Quelles données ?
  - Larges collectes de données épidémiologiques ± accessibles
    - Registres : Danemark, Pays scandinaves, SEER, Francim & InVS
    - Cohortes : NOWAC, E3N, Paquid, Gazelle, Elfe, Constance, Epipage, ...
  - **Open Data** : ensemble de données disponibles pour être traitées par des machines contenant des Systèmes de Gestion de Base de Données (SGBD), puis par des "citoyens" qui en font libre usage

# Big Data : une question de terminologie

- **Big Data** : ensembles mouvants de données hétérogènes produites en temps réel à partir d'une multitude de sources publiques ou privées, locales ou mondiales
  - **V**olume, **V**itesse, **V**ariété (**V**éracité- **IBM**)
  - Unité : petabyte (PB)  $1.10^{15} = 1000\text{terabytes}$  (Doctorow, 2008)
  - Système à feedback permanent : Algorithmes analytiques  $\Rightarrow$  Résultats  $\Rightarrow$  Décisions  $\Rightarrow$  Evolution systèmes
- **"The end of theory ?"** (Chris Anderson, Wired, 2008) : pilotage d'un système complexe repose sur démarche empiriste (**Hume**) et non sur démarche ingénieur avec mise en équation (**Descartes**)

# Big Data en biologie et santé (Viari, 2012)

- 2001 : 1<sup>ière</sup> Cartographie du génome humain, progrès considérables des techniques de séquençages ⇒ Comment gérer, comment analyser ?
- Un phénomène ni récent, ni circonscrit aux sciences du vivant : Physique des particules (LHC, Cern) flux de données // Facebook
- Défis particuliers
  - (Alexandre, 2012) "il devient urgent que les spécialistes du cancer observent comment les astrophysiciens gèrent des exabytes (milliards de milliards) de données produites. Sinon, le **pouvoir** risque de changer de mains"
  - Données de natures très variées : information moléculaire < information tissulaire < corps entier (imagerie) < clinique < environnement
  - Expérimentation signification particulière ≠ schémas habituels de pensée
  - Des nouveaux outils/ métiers en biologie

# Big Data : Entreposer et le vocabulaire associé

- Banque et les nouveaux "banquiers"
- Fermes et les nouveaux "fermiers généraux"
- Containers et leurs "armateurs"
- Datamarts ...et la stratégie de Target exposée par Antoine Geissbuhler
- Le cloud...et le petit coutelier...

# Un changement d'objectif de l'outil

- Bouliers puis 1<sup>iers</sup> calculateurs mécaniques (Pascal)
  - Rendre calculs répétitifs automatiques & rapides
  - Pas vocation stockage  $\Rightarrow$  Trace écrite, registre
- Microfiches : simple réduction espace stockage
- Ere informatique : possibilité de stocker, récupérer, transformer information sous forme électronique, via codage
  - Densification constante  $\pm$  optimisation énergétique
  - Course permanente modernité *versus* obsolescence +++
  - Chute des coûts compensée par course technologique puissance accrue
- Logique de développement technologique  $\Rightarrow$  Quelle pérennité information ? Quel coût énergétique (une fée électricité éternelle !)  
Quelle dépendance ?



# Des sites stratégiques : les entrepôts de "connaissances"



AM3

Amsterdam : Corridor froid, confinement aérolique, green computing, green IT  
(optimisation des infrastructures IT, bâtiment, énergie, exploitation)

# Des sites stratégiques : les entrepôts de "connaissance"

**Comment Equinix a bâti le plus grand datacenter virtuel de France**



De 12 000 mètres carrés, le nouveau datacenter d'Equinix à Pantin ouvrira officiellement ses portes le 31 juillet 2012. Il est raccordé au site du groupe à Saint-Denis, et ses deux centres de données (de 13 000 mètres carrés au total), via 1 000 fibres noires sur deux chemins différents. Un dispositif qui permet ainsi à Equinix de disposer d'un datacenter virtuel de 25 000 mètres carrés. En louant ces fibres, les clients qui le souhaitent pourront dis... [Lire la suite](#)

© Equinix

|   |   |   |   |   |
|---|---|---|---|---|
|  |  |  |  |  |
| 25 000 mètres carrés de...  | Sécurité du datacenter  | Critères de choix du lieu   | Générateurs d'électricité   | Electricité de sources...   |

AM3 Saint- Denis (France) <http://www.equinix.fr>

# Big Data et Stockage : la question du coût énergétique

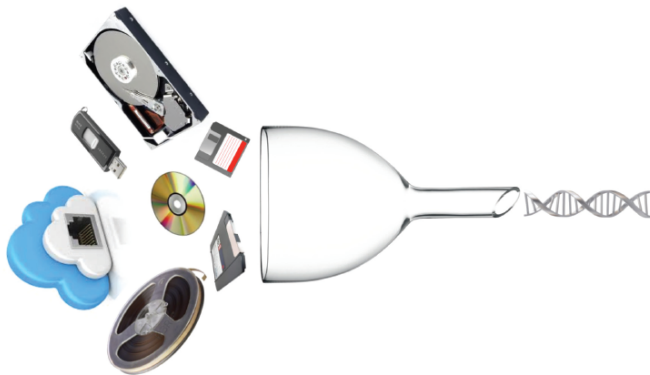
- AM3 (St Denis) : par mois, équivalent consommation ville 20000-50000 habitants
- Europe : 2% de la consommation énergétique mondiale (Greenpace)/ $104 \times 10^9$  en 2020
- Réduction des coûts : enjeu financier vis à vis concurrence
  - Utilisation des eaux profondes pour le refroidissement et de l'eau chaude générée pour le chauffage urbain, les serres, etc...
- Sommet de Rio 1992 : 40 entreprises au sein Business Environmental Leadership Council (BELC)

# Big Data et Stockage : la question de la pérennité

"Nous ne pouvons résoudre des problèmes en reproduisant le type de pensée à l'oeuvre pour les créer" (Einstein)

- Arnano (<http://www.arnano.fr/>)
  - Alain Rey (CEA) : Sur 113 DVD enregistrés gravés entre 2004-2008, 12% signes vieillissements +++ Durée vie : ~ qlqs dizaines d'années (Larousserie, 2012)
  - ⇒ Gravure particulière sur diamant : (couche nitrure + résine photosensible) exposés laser  $\propto$  texte. Couche verre II. Projet
  - Projet "Mémoire" ANDRA. Pas conditions particulières de conservation.
- Penser plus petit et non plus Grand ! : Utilisation ADN synthétique

# Big Data et Stockage : ADN synthétique (Driscoll and Sleator, 2013)



# Big Data et Stockage : ADN synthétique (Driscoll and Sleator, 2013)

- Enzymes de restriction, PCR, Ligases  $\Rightarrow$  codage éléments comme mots
- ADN molécule très stable (cf. découverte ADN espèces disparues depuis 10000 ans)
  - JCV1-syn1.0 : bactérie avec génome synthétique codant pour adresse internet, nom des auteurs (**Joyce, Oppenheimer, Feynman**)
  - Church (Harvard) : bouts ADN chevauchant. Codage livre entier avec texte, images(html), séquence binaire de 5.27 megabit transformée en 54898 blocs.
  - Goldman : amélioration taux erreur lecture  $\propto$  codage par triplet
  - Forte capacité de stockage : 455 exabytes/g  $\Rightarrow$  40 ZB data : 90 g ssADN
- Mais **Read Only Memory**  $\Rightarrow$  Lecture nécessite code // **Pierre de rosette**
- Le futur : ADN ré-inscriptible

# Big Data : Classifier

# Big Data : Big = Complex

- Expérimentation Big Data : agglomération données hétérogènes (cf. méta-analyse en pire!)
- Filtrages, approximations, petits arrangements pour gérer incomplétude des données
- Nécessité interactions **non hiérarchisées** entre "end users" (médecins, pharmaciens, biologistes), DM, bio- informaticiens, statisticiens
- Monde non stabilisé, mouvant avec avancées technologiques constantes
- Une nouvelle discipline : **la biocuration**



## Exemple \*omics : RNA non codant (Liu et al., 2012)

- Intérêt initial pour ARN codant pour protéines
- Développements technologiques permettent recueil ARN non codant(ncRNA)
  - miRNA (19- 25 nucléotides) : rôle de régulateur sur traduction/ répression RNA codants
  - lncRNA (~ 200 nucléotides) plus récents : rôle sur modification chromatine et traduction ARN codants ?
- Mais...plus 45 formats pour séquences. Formats dominants < EMBL, GenBank, SwissProt, PIR
- ⇒ Manque standardisation, faiblesse validations croisées

## Big = Large. Goliath contre David ?

- SciDB <http://www.scidb.org>
- Le projet 1000 génomes (<http://www.1000genomes.org>)
- Le projet Huma-Num <http://www.huma-num.fr>
  - TGIR vise à faciliter tournant numérique en SHS. Echelle nationale & européenne
  - TGIR promeut consortia, offre dispositif unifié (ISIDORE), archivage
  - TGIR promeut "guides de bonnes pratiques technologiques"
  - TGIR porte participation France dans projet DARIAH [www.dariah.eu](http://www.dariah.eu)
- Des conséquences en terme ré- organisation milieu, leadership, production, carrières...
- ⇒ Collecter des fonds & (se) valoriser : 20 patients × 100+ signataires !

# La biocuration (Howe et al., 2008)

<http://www.biocurator.org>

- Définition : activité d'organiser, représenter et rendre l'information biologique accessible aussi bien pour les personnes que pour les ordinateurs
- Trois maux : financement, développement et reconnaissance.
- Enjeux
  - Tout dispositif de gestion de données repose sur modèle conceptuel (Viari, 2012)
  - Pas modèle universel que ce soit pour un même type de données ou données de natures différentes
  - Importance **interopérabilité** sur plans technique (compression et pertes d'information) & **++ sémantique** (ontologie)

# La biocuration (Howe et al., 2008)

<http://www.biocurator.org>

- Les multiples rôles d'un *biocurateur*
  - Extraire l'information des journaux publiés
  - Connecter des informations de sources variées d'une manière cohérente et exhaustive
  - Inspecter et corriger automatiquement les structures et séquences
  - Développer un vocabulaire adapté
  - Intégrer des bases de connaissance pour représenter des systèmes complexes
  - Corriger inconsistances et erreurs dans la représentation des données
  - Aider les utilisateurs pour les rendre plus productifs
  - Piloter le plan de ressources reposant sur le web
  - Interagir avec les recherches
- Augmenter la visibilité des biocurateurs en tant que profession reconnue

# La curation est idéalement collective

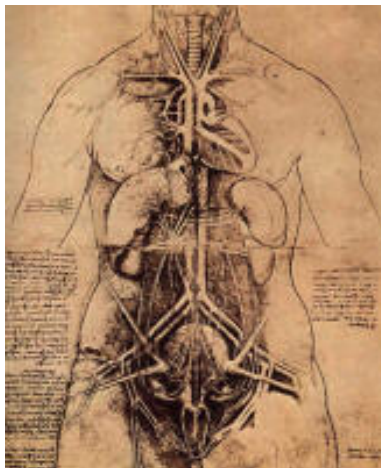
- Reconnaître la tâche d'annotation comme partie normale du travail académique et comme le prolongement naturel de la tâche de publication
- Développer structures de reconnaissance
  - Exemples : Daphnia Genomics Consortium ; International Glossina Genomics Initiative ; Sloan Digital Sky Survey
  - WikiProfessional Life Sciences <http://www.wikiprofessional.org>
  - Mise en place d'organes de diffusion cf. (Latour and Woolgar, 1979, Bensaude-Vincent, 1993)
    - The Journal of Biological Databases and Curation
    - Journal en ligne **Gigascience**
    - Actions communes entre auteurs, journaux et curateurs pour faciliter échanges de données et mise à disposition pour validation par pairs (cf. débat OGM)
  - Pouvoir mobiliser le public général

# Big Data : Penser

# De la pré- connaissance à la connaissance

- Un groupement d'information n'est pas neutre, forme de pré- connaissance (cf. pro- hormones ?)
- Passage pré- connaissance à connaissance nécessite modélisation puis interprétation résultats
  - Suppose une finalité et un angle d'analyse, qui n'est pas neutre (cf. choix des variables, hypothèses de méta- analyses, modèles de risques compétitifs, modèles d'expression génique, etc...)
    - Exemples : Nowac, E3N et âge ménopause ; THM et risques cancer sein...
  - **Desrosières** (Desrosières, 1993) : Comment prendre en compte à l'étape analyse/ discussion d'une étude les effets incitatifs des résultats sur le public concerné ?

# Voir c'est interpréter





# Ré- analyses de larges BDD : des succès riches de promesses

- Données néo- natales et risques CV adultes (Barker and Osmond, 1988, Barker et al., 1989, 1993)
  - Une source papier de données professionnelles sur 50+ ans : biométrie néonatale du carnet santé remplie par SF (terme, genre, poids placenta, poids et morphotype enfant)
  - Croisement avec système de santé NHS/ Registres ⇒ Facteurs risque majeurs
  - Initiation domaine recherche intense (Maternité Haguenau..., Eden)
  - Quel serait le niveau de preuve d'une telle recherche aujourd'hui ?
- Emmanuel Todd (Todd, 1976) : thèse de la chute de l'empire soviétique à partir reconstitution tableaux et graphes à partir statistiques officielles concernant mortalité néonatale, sous réserve de quelques hypothèses sur données manquantes

# Big Data : croisements registres × DMA

## ● Registres

- Registres scandinaves : 1ier registre danois décès 1875- Système citoyen démocratique. Croisement possible entre jeux de données
- France : émergence lente, volontés locales, financements complexes...
  - Une définition précise, des critères de labélisation, rôle **InVS** ++
  - 53 registres qualifiés, 15% population. **Particularité registre : AVP-RA** (Martin et al., 2004, Laumon, 1998)
  - Les évolutions de codifications internationales, les enjeux sociétaux (BAAC *versus* registre RA), les aspects juridiques (maladies rares, AVP-RA), ajustements et pondérations...

## ● Données des bases médico- sociales

- Les données couplées SNIIRAM- CNAMTS : Histoire médicale à travers proxy actes ayant donné lieu à remboursement (dépistage, chirurgie bariatrique, nodule thyroïde, CO 2-3...)
- Données Assurances Privées : globalement fermées (INCA- Aeras ; MA & IFSSTAR)

# Croisement de données et méthodologies statistiques

Un domaine très actif, qui évolue très vite

- Les techniques de capture - recapture (Amoros et al., 2007a,b, 2008)
- Les techniques à classes latentes
  - Techniques fréquentistes (Walter et al., 2012, 2013)
  - Techniques bayésiennes (Joseph et al., 1995, Demissie et al., 1998, Sewitch et al., 2013)
    - Des méthodes et algorithmes performants grâce aux puissances de calcul
    - L'incorporation de **croiances a priori**
- L'importance du monde libre, des échanges directs (cf. P Degoulet) : R, Bioconductor...

# Big Data : Le Data Mining

- Suppose définitions modalités d'extraction (distance, logique, etc...)
- Permet, en théorie, faire ressortir **signaux invisibles** via voies métaboliques communes ou gènes communs : exemple des maladies rares, exploration systématique banques de données génomiques ⇒ Rapprochements, Identifications voire Emergence de thérapeutiques
- *a priori* Open Data  $\neq$  Big Data
  - Exploration hypothèses pour programmes d'action ⇒ avec des limites
  - Difficultés  $\propto$  fluidité, intemporalité, recomposition permanente
  - Accumulation de données, aussi large et hétérogène qu'elle soit, ne constitue pas une connaissance en soi
  - Résultat : **alchimie subtile** entre diverses intentionnalités : celles des collecteurs et celles des exploiters
  - Le rôle des intermédiaires  $\pm$  obscurs
- Le nouveau laboratoire...(GK), les nouvelles règles de transparence

# Nouveaux outils, nouvelles classifications $\Rightarrow$ nouveaux savoirs conceptuels ?

## D Raoult (Marseille) et les virus géants

- Un pêcheur collectionneur enthousiaste et passionné, dont la science repose sur les outils qu'il manipule, ouvrant vers des mondes nouveaux
- Des moissons toujours plus grandes et plus belles  $\propto$  séquenceur : de la pêche à la traine à la pêche au chalut
- Nécessité de reconceptualiser pour penser : l'information ne respecte plus les entités patiemment répertoriées. Elle doit être prise globalement, communautairement

# Big Data, techniques marketing & humanitaire :

## Datakind, Jack Porway

### Evaluation programmes actions & interventions

- Keyfund : évaluation des programmes soutien scolaire
- Croix Rouge Chicago : prévision des quartiers à risque d'incendie
- Utilisation réseau mobile
  - *Mobilizing Health* (Indes) analyse appels & niveau/ type aide médicale
  - *Grameen Foundation* et aide en zone rurale fermiers
  - ONU/ Banque mondiale : usage portable & mesure niveaux pauvreté selon zones
  - Adolescents & prévention MST, VIH

# Big Data, Data Mining. Quelques exemples

- Nate Silver (Silver, 2012), "sabermétricien" (SABR : Society for American Base Ball Research) **Nerds** ("intello"= no one really dies) *versus* **Geeks** !
- Cartographie des épidémies (Hay et al., 2013) sur les traces de Charles Nicolle à Tunis
  - Projet HealthMap et survenue en temps réeldes cas  
<http://www.healthmap.org/en/>
  - <http://www.flunearyou.org>, <http://www.influenzanet.eu>
- Utiliser les ressources du cloud pour comprendre l'impact de la variabilité génétique sur le fonctionnement du cerveau : action jointe Inria (MapReduce), Microsoft (Azure) (Antoniou et al., 2012)
- Dr Watson (IBM) en oncologie (Sillig, 2013) Le jeu "**Jeopardy**" et histoire clinique d'un patient, "justification" du raisonnement dans tous les cas (Hume) !

# Big Data : illusion d'une histoire vraie ?

- (Bovet and Gedeon, 2013) à propos d'un article de J Salomon (Salomon et al., 2012) sur poids des maladies
- Espérance de vie à naissance aux Seychelles, 1990- 2010 : 62.0 ↘ 61.3 ans (hommes), 72.4 ↘ 71.8 ans (femmes) ≠ registre local indique 63 ↗ 69 ans (hommes), 74 ↗ 79 ans (femmes)
- Réponse auteurs (Wang et al., 2013) : **travail n'a pas cherché autres données** que données OMS. Standardisation indirecte  $\propto$  variations strates genres  $\times \times$  âges

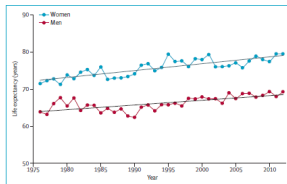


Figure Life expectancy at birth in Seychelles between 1976 and 2012



# Big Data : un progrès ?

- Accumulation de données, open ou big data : illusion source de connaissances en soi ?
- Données jamais neutres
- Nouvelles technologies ne se substituent pas simplement aux anciennes : elles témoignent monde nouveau et nouveaux modes de communauté qui rejette monde ancien  $\pm$  brutalement
- Monde futur  $\neq$  monde ancien amélioré
  - Nouveaux modes de penser/ vivre contenus dans fonctionnalités de ses technologies
  - Economie des nouvelles technologies contenue dans leurs conception et modalités de diffusion, sans distanciation possible

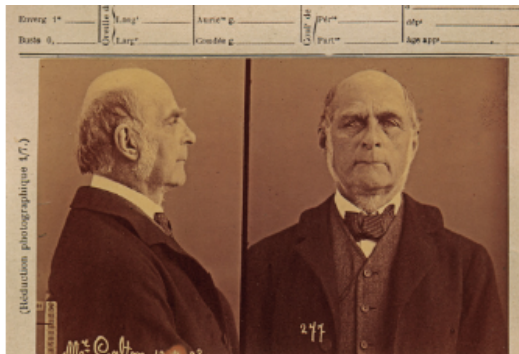
# Big Data & ses outils d'accès : **Komboloi** des temps modernes

- Nouveaux modes d'apprentissage d'un savoir. Que devient le processus de mémorisation ?
- Acquérir un savoir ne passe plus forcément par reconstruction selon sa propre logique d'une chronologie de savoirs
- Des nouveaux modes d'échanges
- Sophistication des modes d'extraction, augmentation données disponibles : naturalisation de notre vision du monde et de nos décisions à travers prisme de ces instruments ?
- Des pratiques professionnelles profondément changées où l'avalanche d'information remplace le temps de réflexion ⇒ **Une perte de l'épaisseur du temps**

# Big Data et vie privée

## Vos données m'intéressent...

# 1911-2011 : Le spectre de la biométrie eugéniste

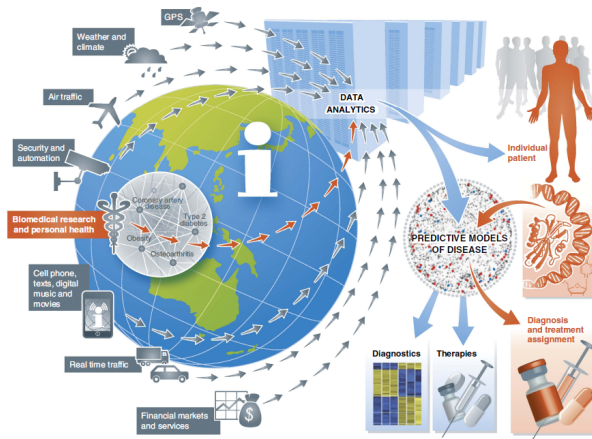


*Du génie héréditaire et de l'intelligence naturelle*

# Big data et vie privée : un paradigme nouveau ? (Schadt, 2012)

- Le consentement à l'exploitation de données personnelles fortement contesté (Ioannidis, 2013) au nom de la science !
  - Biais de consentement et ses corrections possibles
  - Absence de perturbation vie privée lors d'un traitement à l'insu de la personne
  - Statut exploitation Big Data  $\neq$  "vraies recherches" = Oxymoron (Ioannidis) car recherches qui ne sont pas des recherches

# Big data et vie privée : un paradigme nouveau ? (Schadt, 2012)



# Big Data : l'anonymat a-t-il encore un sens ?

- 1975 : Les trois jours du condor (Sydney Pottier)
- 2013 : Verizon
- Modifications d'un profil d'achat, images, mobilité (Carrefour, Amazon, Fnac, péages, velib) et informations intimes (grossesse, etc...)
- Profil abondance ARN plasmatique et inférence ADN  $\Rightarrow$  Identification

# Cloud computing et espionnage industriel (Editor, 2013)

- NSA et les programmes PRISM (US) et TEMPORA (UK)
  - Duplication données du net pour utilisation rétroactive (site de Bluffdale, Utah)
  - Arsenal légal : loi Foreign Intelligence Surveillance Amendments Act (FISAA), 2008 permet de "siphonner" les données du net, portée extra-territoriale, prolongée de 5 ans
  - Permet action sur données **citoyens non US**. Derniers amendements FISAA (1881-a) étendent surveillance à toutes données sur le cloud.
    - Amazon, Google, Microsoft, Apple, compagnies américaines
    - Accord USA-CE : centres cloud = "safe harbors" dispensées accord CNIL ou équivalents EU
- Salon Milipol (Salon international de la sécurité intérieure des états) Paris (2013), Qatar (2014) sponsors officiels (Renault Trucks Defense). Domaines couverts
  - **Traitement des fichiers et bases de données**



# Big Data : l'anonymat a-t-il encore un sens ?

- Facebook et le logiciel Graph Search
  - Identification possible liste individus français, RATP, lecteur mein Kampf
  - Soldats français partis en Afganistan (Le Monde, 11/07/2013)
  - Des "jeux" aux conséquences imprévisibles
    - Résultats non garantis en précision mais stigmatisants +++ (cf. harcèlements au travail via courriel)
    - ⇒ Réponse Facebook (Erin Egan, responsable vie privée) : l'internaute doit être responsabilisé sur les contenus le concernant !
- MOOCs (Massive Open Online Course)
  - Coursera, Udacity, Edx, ....
  - Des enjeux importants, une compétition intense
  - Des aspects éthiques : le profil psychologique des étudiants...

# Multiplication des services en ligne et ré- utilisation données

- Réutilisation non-autorisée de données n'est pas une invention de l'ère Internet
- Premiers services d'annuaire inversés proposés sur Minitel : réutilisation non-autorisée des données publiées France Telecom
- Listes multiples disponibles sur Internet  $\pm$  directement accessibles
- Outils logiciels disponibles librement
- "Barrière à l'entrée pour la collecte non-autorisée de données s'est donc très largement abaissée" (Campbell, 2013)

# Donner ses données en les contrôlant ? [we.consent.org](http://we.consent.org) "Donate your data"

- Projet directive européenne 2013 (Directive, 2013) qui modifie directive 2003/98/UE sur réutilisation informations du secteur public
  - "les documents produits par les organismes du secteur public des Etats membres constituent une réserve de ressources vaste, diversifiée et précieuse, dont peut bénéficier **l'économie de la connaissance**
  - Données concernées : bibliothèques publiques, dont universitaires, musées et archives
  - Apparition du terme **méta-données**

## Solution : les Open Data ?

- Ville de Rennes (Xavier Crouan) : prix pour des applications intelligentes
- Fourmi- Santé : utilisation des données AM (site Ameli-Direct) pour proposer comparateur prix pratiqués par médecins. Procédure contestée par CNAMTS
- (de Kervasdoué and Sicard, 2013) "Plus grave que la pilule, l'affaire des données de santé" publique"
- Un intérêt mutuel valorisation développeur/ détenteur BD à ré-utilisation dans cadre juridique précis

## Le spectre du panopticon Nicolas Demorand (Demorand, 2013)

"La sphère privée n'existe plus....Rien de ce qui est montré, dit ou écrit sur Internet n'échappe à ce Panopticon électronique qui scanne, copie et stocke tous les échanges. Que cette masse inimaginable de données **puisse ou non être exploitée** finit par sembler mineur au regard de l'infraction brutale, durable, sans doute même irréversible contre ce qui fonde toute démocratie : le respect de la vie privée. ...[Le vieux continent] déjà incapable de collecter l'impôt des géants américains qui commercent sur leur sol et ouvrant grands leurs serveurs aux agences de renseignements de leur pays...comment même songer qu'is puissent créer, appliquer et défendre un *Habeas Corpus numérique*. ...La souveraineté numérique fonde désormais une part essentielle de la puissance d'un état et **s'impose de fait comme le sujet éthique, politique, économique, international majeur de l'époque**"

## Quelques réflexions pour conclure

- A propos de la notion de progrès
- Un renouveau du dialogue soignant/ soigné
- Un plaidoyer pour la recherche : savoir utiliser, savoir imaginer, savoir réaliser
  - En respectant la volonté des participants volontaires
  - Dans le respect les règles et conventions sociales
  - En laissant des traces ouvertes aux autres (des cahiers ouverts)

## Progrès et Humanisme (Wismann and Kahn, 2012)

- Notion Progrès liée à la Renaissance en Europe et la possibilité d'un futur pour l'homme et non une simple révolution (Copernic, découverte du nouveau monde,...)
- Pascal : "Toute la suite des hommes, pendant le cours de tant de siècles, doit être considérée comme un seul homme qui subsiste toujours et qui apprend continuellement".
- Opposition toujours présente
  - Socrate : le Vrai (= Savoir), conduit au Bien. Seul l'ignorant peut s'engager dans la voie du Mal
  - Protagoras, le sophiste : quête du Bien séparée quête du Vrai  $\Rightarrow$  Une innovation scientifique belle et brillante peut ne pas être bonne
- A Kahn "Progrès = (Intelligence + créativité  $\Rightarrow$  Connaissances + Techniques)  $\cap$  (Richesses dans but Humaniste)"

# Un dialogue renouvelé entre soigné et soignant





# Un échange renouvelé entre soigné, ses proches et soignant

- La consultation : nouveau lieu de confrontation des savoirs
- Les patients s'organisent et proposent
  - Associations de patients et leurs familles (rôle  $\pm$  en coulisse Big Pharma ?)
  - L'initiative Patients Like Me & Value for Openess
  - Bona Dona
- Les registres d'études <http://www.ClinicalTrials.gov>
- Les "cécités collectives" professionnelles au risque du droit commun (médiator, c.o., GH, etc...)

# Google, internet et la santé

The screenshot displays the PatientsLikeMe website interface. At the top, there is a navigation bar with the logo, a search bar, and a login section. Below the navigation bar, a banner features a large green question mark icon and the text "Get your health in order. Join PatientsLikeMe." The main content area is divided into several sections:

- 120,833 patients 500+ conditions**: A blue box highlighting the number of users and conditions.
- Who's like you?**: A section with a "Share your experience." prompt, stating "The more you share, the easier it will be to find patients like you. Start by adding a condition, symptom or treatment."
- I have**: A section for sharing conditions, with a text input field "Type at least 3 letters of a condition".
- I take**: A section for sharing treatments, with a text input field "Type at least 3 letters of a treatment".
- I am**: A section for gender selection, with buttons for "Male" and "Female".
- My Age**: A section for age selection, with a horizontal slider ranging from 0 to 100.
- Join Now! (It's free)**: A yellow button at the bottom of the registration options.

On the right side of the page, there is a search bar and a list of navigation links:

- Find Patients Like You
- Explore our Treatment Reports
- Learn about Symptoms
- Review our Research
- Check for your Conditions
- Patient Testimonials


Below the testimonials, there is a section titled "About PatientsLikeMe" with links for "Read about the company", "How we make money", "Be part of our team", "Contact Us", and "Information for Industry Partners".

# Google, internet et la santé

## The Value of Openness

patientslikeme<sup>®</sup>  
Patients helping patients live better every day.

Home | Archives | What is our Openness Philosophy?



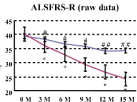
Search

### Does It Work? Lithium and ALS

by James Heywood

Update (March 7, 2008): [PatientsLikeMe ALS Lithium Research](#) released.

**Does it work?**  
On February 12th of this year, Proceedings of the National Academy of Sciences (one of the leading science journals) published a paper entitled — Lithium Delays Progression of Amyotrophic Lateral Sclerosis. After 10 years researching ALS, I believe it is fair to say this paper includes the most promising suggestive set of data from a clinical trial ever published. I say "suggestive" because there are many flaws with both the information presented and with the publication process itself. These flaws make it so that patients and their doctors are left trying to draw conclusions about the use of Lithium to treat ALS, without actually having any realistic confidence in the data or its meaning.



| Month | Score |
|-------|-------|
| 0 M   | 40    |
| 3 M   | 38    |
| 6 M   | 35    |
| 9 M   | 32    |
| 12 M  | 28    |
| 15 M  | 25    |

#### About

PatientsLikeMe is the leading online community for people with life-changing conditions. Patients embrace the open sharing of personal health data because they believe that information can change the course of their disease.


With a focus on patients and research, our blog reflects knowledge resulting from the shared real-world experiences of our community. Welcome to the genesis of patient-led research.

---

#### Stay Connected

- [Subscribe via RSS](#)
- [Subscribe via Email](#)
- [Twitter](#)
- [Facebook](#)


# Google, internet et la santé

 23andMe genetics just got personal.

Search 23andMe   [Log In](#) | [Register Your Kit](#) | [Blog](#) | [Help](#) | [Cart](#)

welcome | ancestry | health | how it works | store

Start filling in the gaps with your DNA




"Because I had given my doctor information from 23andMe, he got to a diagnosis much faster. 23andme saved my life." Kirk C.

\$99\* **Our new low price for all!**  
Was \$199

**Order Now >>**

\*Requires a 1-year commitment to the Personal Genome Service® at \$9/mo. **Order for \$399** without commitment.

- Get Your Kit
- Provide Saliva
- Learn About Yourself
- Get Monthly DNA Discoveries



Gain insight into your traits, from baldness to muscle performance. Discover risk factors for 97 diseases. Know your predicted response to drugs, from blood thinners to coffee. And uncover your ancestral origins. [start tour >](#)

Overview | Discover Health & Ancestry | Keep Your Doctor Informed | Participate In Research

# La santé devenue marché prometteur !



le rendez-vous des télécoms et de la santé

CONFÉRENCE ET EXPOSITION  
15 et 16  
Maison de l'...

ACCUEIL CONFÉRENCE SPEAKERS EXPOSITION ANIMATIONS PARTENAIRES



Vous souhaitez... inté

## Le M-Health, l'e-santé, un marché en pleine éclosion !

Pratiquement ignoré il y a encore 2 ans, le marché de la santé mobile ou "M-Health" devrait représenter **un marché de 6 milliards de dollars d'ici 2015.**

Depuis 2008 et le lancement de l'App Store, **17 000 applications de m-santé** ont déjà été développées sur les appareils mobiles de dernière génération : les smartphones. Face au vieillissement de la population, à l'évolution des maladies chroniques et à l'augmentation du coût des soins, le développement des applications de santé apporte aujourd'hui des réponses concrètes aux attentes des patients et des professionnels de la santé.

M-Health market size by revenue type 2010-2015



| Revenues (US \$ Millions) | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---------------------------|------|------|------|------|------|------|
| Advertisement             | 5    | 17   | 31   | 45   | 58   | 71   |
| Transaction               | 0    | 63   | 156  | 247  | 395  | 595  |
| Paid download             | 148  | 279  | 398  | 484  | 552  | 618  |
| Device sales              | 0    | 205  | 936  | 809  | 1201 | 1867 |
| Services                  | 0    | 110  | 987  | 1544 | 2100 | 2950 |

Saisissez votre email pour vous li



# Conceptions des régulations et savoirs ambiants

- La porosité des domaines vis à vis des concepts
- L'analogie électrique et l'homéostasie. Les circuits RLC, les modèles compartimentaux
- Le concept hormone- récepteur et l'information distribuée du réseau à jeton IBM
- L'horloge hypothalamique et les tops d'horloge de la logique digitale



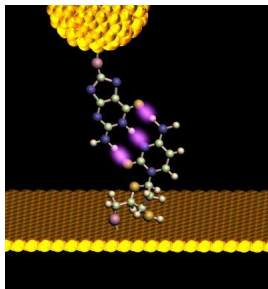
# Gurdon : Cellules Souches Pluripotentes Embryonnaires ESC

- 1952 : Noyau C embryon de têtard (*blastula*) + oeuf énuclée de grenouille ⇒ Têtard normal
  - Dogme : la différenciation cellulaire entraîne des changements irréversibles
    - ....mais difficultés techniques : pas de preuve énucléation totale, stade exact embryon mal connu
- Etapes suivantes
  - Gurdon (1975) : Cellules totalement différenciées ; marquage génétique hôte/ donneur
  - Campbell & Wilmut : **Dolly**, 1997 : extension aux mammifères
- ⇒ Espoirs, Difficultés, QS dans production des **ESC**



# La *mode* de la complexité : décryptage du génome et de son expression & biologie des systèmes

- Une information massive : des concepts et outils nouveaux



**FIGURE:** Science Daily, 2009. **Microscope à effet tunnel** et lecture des paires A-T ou C-G

**IBM** : un investissement raisonné sur le long terme!

# Yamanaka : les cellules souches pluripotentes **PSC**

- 2006 : Production de PES  $\propto$  cellules de peau de souris adultes
- Transfert pluripotence aux fibroblastes  $\propto$  présence groupe 24 gènes
  - Oct4, Sox2, Klf4, c-Myc transfectés ensemble  $\Rightarrow$  **iPS**
- Scepticisme...mais **réplication** et extensions
  - Correction hémoglobine dans la drépanocytose à partir d'iPS provenant d'une biopsie de peau
  - Fibroblaste peau humaine : 4 gènes particuliers  $\rightarrow$  iPS
- $\Rightarrow$  Yamanaka a renversé le point de vue sur la programmation nucléaire

# Art & Technique : un parallèle possible



**FIGURE:** De gauche à droite : Raphaël 1511, Titian 1543, Velazquez 1650, Bacon 1953, Brown 2008. (Goldstein, 2009)

Merci de votre attention !

## References I

- Foldit : solve puzzles for science, 2008.
- L. Alexandre. Le docteur et l'astrophysicien. *Le Monde*, 20977, Juin 2012.
- E. Amoros, J.-L. Martin, M. Chiron, and B. Laumon. Road crash casualties : characteristics of police injury severity misclassification. *J Trauma*, 62(2) :482–490, 2007a.
- E. Amoros, J.-L. Martin, and B. Laumon. Estimating non-fatal road casualties in a large french county, using the capture-recapture method. *Accid Anal Prev*, 39(3) :483–490, 2007b.
- E. Amoros, J.-L. Martin, S. Lafont, and B. Laumon. Actual incidences of road casualties, and their injury severity, modelled from police and hospital data, france. *Eur J Public Health*, 18(4) :360–365, 2008.
- G. Antoniu, A. Costan, B. da Mota, B. Thirions, and R. Tudoran. A-brain : using the cloud to understand the impact of genetic variability on the brain. *ERCIM News*, 89, 2012.

## References II

- D. J. Barker and C. Osmond. Low birth weight and hypertension. *BMJ*, 297(6641) :134–135, Jul 1988.
- D. J. Barker, P. D. Winter, C. Osmond, B. Margetts, and S. J. Simmonds. Weight in infancy and death from ischaemic heart disease. *Lancet*, 2 (8663) :577–580, Sep 1989.
- D. J. Barker, C. N. Martyn, C. Osmond, C. N. Hales, and C. H. Fall. Growth in utero and serum cholesterol concentrations in adult life. *BMJ*, 307(6918) :1524–1527, Dec 1993.
- B. Bensaude-Vincent. *Lavoisier. Mémoires d'une révolution*. Flammarion, 1993.
- P. Bovet and J. Gedeon. Life expectancy in seychelles. *Lancet*, 382 :23, 2013.
- D. Campbell. Le royaume- uni, maître- espion. *Le Monde*, page 18, 02 Jul 2013.

## References III

- J. de Kervasdoué and D. Sicard. Plus grave que le débat sur la pilule : l'affaire des données de santé publique. *Le Monde*, page 18, 2013.
- K. Demissie, N. White, L. Joseph, and P. Ernst. Bayesian estimation of asthma prevalence, and comparison of exercise and questionnaire diagnostics in the absence of a gold standard. *Ann Epidemiol*, 8(3) : 201–208, Apr 1998.
- N. Demorand. Panopticon. *Libération*, 9994, 02 Jul 2013.
- A. Desrosières. *La politique des grands nombres : histoire de la raison statistique*. La Découverte, 1993.
- Directive. Directive 2013/98/ue du parlement européen et du conseil du 26/06/2013 modifiant la directive 2003/98/ue concernant la réutilisation des iinformation du secteur public. *Journal Officiel de l'Union européenne*, 2013.

## References IV

- C. Doctorow. Big data : Welcome to the petacentre. *Nature*, 455(7209) : 16–21, Sep 2008.
- A. O. Driscoll and R. D. Sleator. Synthetic dna : The next generation of big data storage. *Bioengineered*, 4(3) :123–125, 2013.
- Editor. L'oncle sam se comporte très, très mal. *Le Monde*, 21290 :1, 02 Jul 2013.
- K. Firas, F. Dimaio, F. C. Group, F. V. C. Group, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popovic, M. Jaskolski, and D. Baker. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat Struct Mol Biol*, 18 :1175–1177, 2011.
- J. L. Goldstein. Lasker awards and papal portraiture : turning fields upside down. *Nat Med*, 15(10) :1137–1140, Oct 2009.



## References V

- S. I. Hay, D. B. George, C. L. Moyes, and J. S. Brownstein. Big data opportunities for global infectious disease surveillance. *PLoS Med*, 10(4) :e1001413, 2013.
- D. Howe, M. Costanzo, P. Fey, T. Gojobori, L. Hannick, W. Hide, D. P. Hill, R. Kania, M. Schaeffer, S. S. Pierre, S. Twigger, O. White, and S. Y. Rhee. Big data : The future of biocuration. *Nature*, 455(7209) : 47–50, 2008.
- J. P. A. Ioannidis. Informed consent, big data, and the oxymoron of research that is not research. *Am J Bioeth*, 13(4) :40–42, 2013.
- L. Joseph, T. W. Gyorkos, and L. Coupal. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol*, 141(3) :263–272, Feb 1995.
- D. Larousserie. Des microfiches en saphir pour affronter l'éternité : une solution française promet un stockage de l'information sans perte pendant des millénaires. *Le Monde*, 2012.

## References VI

- B. Latour and S. Woolgar. *Laboratory Life : the Social Construction of Scientific Facts*. Princeton University Press, 1979.
- B. Laumon. Epidemiologic research and road traffic accidentology in europe. *Rev Epidemiol Sante Publique*, 46(6) :509–521, Dec 1998.
- C. H. Liu, D.-Y. Wu, and J. D. Pollock. Bioinformatic challenges of big data in non-coding rna research. *Front Genet*, 3 :178, 2012.
- J.-L. Martin, S. Lafont, M. Chiron, B. Gadegbeku, and B. Laumon. Differences between males and females in traffic accident risk in france. *Rev Epidemiol Sante Publique*, 52(4) :357–367, Sep 2004.
- J. Salomon, H. Wang, and M. Freeman. Health life expectancy for 187 countries : a systematic analysis for the global burden of diseases 2010. *Lancet*, 380 :2144–2162, 2012.
- E. E. Schadt. The changing privacy landscape in the era of big data. *Mol Syst Biol*, 8 :612, 2012.

## References VII

- M. J. Sewitch, M. Jiang, L. Joseph, R. J. Hilsden, and A. Bitton. Developing model-based algorithms to identify screening colonoscopies using administrative health databases. *BMC Med Inform Decis Mak*, 13 :45, 2013.
- L. Sillig. Les conseils du docteur watson : l'ordinateur qui a appris le langage des humains, au point de les battre lorsdu jeu télévisé "jeopardy", se lance dans l'oncologie. *Le Monde/ Le Temps*, 2013.
- N. Silver. *The Signal and the Noise*. 2012.
- N. N. Taleb. *The Black Swan, The Impact of the Hyghly Improbable*. Les Belles Lettres, 2007.
- E. Todd. *La Chute finale : Essai sur la décomposition de la sphère soviétique*. Robert Laffont, 1976.
- A. Viari. Big data en biologie. *Médecine/ Sciences*, 28 :1027–1028, 2012.

## References VIII

- S. D. Walter, P. Macaskill, S. J. Lord, and L. Irwig. Effect of dependent errors in the assessment of diagnostic or screening test accuracy when the reference standard is imperfect. *Stat Med*, 31(11-12) :1129–1138, May 2012.
- S. D. Walter, C. A. Riddell, T. Rabachini, L. L. Villa, and E. L. Franco. Accuracy of p53 codon 72 polymorphism status determined by multiple laboratory methods : a latent class model analysis. *PLoS One*, 8(2) : e56430, 2013.
- H. Wang, J. A. Salomon, and C. J. L. Murray. Life expectancy in seychelles - authors' reply. *Lancet*, 382(9886) :23–24, Jul 2013.
- H. Wismann and A. Kahn. Repenser le progrès dans une perspective humaniste. *Sciences et Avenir*, pages 46–50, 2012.