

Data Publica Sanitaria :

de l'Open Data à l'Open Report en santé
publique au service du chercheur, du
professionnel, du citoyen et du décideur

P. Staccini

Département d'Information et d'Informatique Médicale - **CHU Nice**
LabSTIC Santé - **UFR Médecine** - Université Nice-Sophia Antipolis
SESSTIM UMR 912 - Inserm - Université Aix-Marseille

Mercredi 4 juillet 2012 - Open Data : des paroles aux actes - Telecom Valley - Sophia Antipolis





Opening what ?

- Opening the programming code : **open source** (openOffice)
- Opening the access to pedagogical resources and scientific litterature : **open access** (openArchives) to knowledge (open data banks for medical resources – european project)
- Opening the access to elementary or agregated data : **open data** (validity, availability, formats, communicating)
- Opening the access to the analysis of data : **open report** for decision-making



Agenda 21 – Chap. 40

Dans le cadre du développement durable, chacun est un utilisateur et un fournisseur d'informations, au sens large. (...). Pour veiller à ce que **les décisions soient de plus en plus fondées sur des informations correctes**, il y a lieu d'appliquer les deux éléments ci-après du programme :

- élimination du fossé qui existe en matière d'information;
- amélioration de l'accès à l'information.

<http://www.un.org/french/ga/special/sids/agenda21/action0.htm>



Elimination du fossé

Assurer une collecte et une évaluation des données d'un meilleur rapport coût-efficacité en identifiant mieux les utilisateurs publics et privés et leurs besoins en matière d'information à l'échelon local, provincial, national et mondial;

Renforcer la capacité à l'échelon local, provincial, national et mondial de **recueillir des informations multisectorielles** et de les utiliser dans le processus de prise de décisions et promouvoir la capacité de collecte et d'analyse de données et d'informations pour la prise de décisions, notamment dans les pays en développement;

Mettre au point ou renforcer les moyens permettant, à l'échelon local, provincial, national et mondial, d'assurer que la planification du développement durable soit fondée, dans tous les secteurs, sur des informations opportunes, fiables et utilisables;

Présenter des informations pertinentes sous la forme et dans les **délais requis** pour faciliter leur utilisation.



Multi-Trans-Sectorialité

Outre le renforcement des activités actuelles de collecte de données concernant le développement, il faudrait accorder une attention particulière à des **domaines** comme les **facteurs démographiques**, **l'urbanisation**, **la pauvreté**, **la santé** et le **droit à l'accès aux ressources**, ainsi qu'à des groupes spéciaux tels que **les femmes**, **les populations autochtones**, **les jeunes**, **les enfants** et **les handicapés**, et à la **relation** qui existe entre ces domaines et les **problèmes d'environnement**.



Santé Publique

La science et l'art de la prévention des maladies, du prolongement de la vie et de la promotion de la santé d'un groupe ou d'une population grâce aux efforts organisés de la société .

Le champ d'action de la santé publique couvre tous les efforts sociaux, politiques, organisationnels qui sont destinés à améliorer la santé de groupes ou de populations entières : approches organisées, systèmes de promotion de la santé, de prévention des maladies, de lutte contre la maladie, de réadaptation ou de soins orientés en ce sens.

L'approche écologique de la santé publique met en avant les **déterminants environnementaux et économiques** de la santé et met l'accent sur la façon dont l'investissement économique devrait être guidé pour **obtenir les meilleurs résultats en santé pour la population**, la plus grande **équité en santé** et **l'utilisation soutenable des ressources**.

<http://asp.bdsp.ehesp.fr/Glossaire/>



Open Data in Healthcare

CDC Data & Statistics – The Centers for Disease Control & Prevention (CDC) have opened access to a wide range of data for public health purposes.

Google Public Data Explorer - Making large public data sets on healthcare, economics, and other subjects readily accessible and easy to explore & visualize.

HCUP Databases - Databases on healthcare cost & utilization in the U.S. available from the Agency for Healthcare Research & Quality (AHRQ).

Health.Data.Gov – The U.S. Department of Health & Human Services (HHS) has provided open access to a wide range of health information & data sets that are generated and/or held by the U.S. Government.

NationMaster - Compilation of health information & statistics from such sources as the UN, WHO, UNESCO, UNICEF, World Bank, etc.

WHO Global Health Observatory – The World Health Organization (WHO) Global Health Observatory provides data and analyses on global health priorities.

World Bank Health Data - Data collected by the World Bank on national health systems, disease prevention, reproductive health, nutrition, population, and more.

<http://www.openhealthnews.com/blogs/groenpj/2012-03-06/open-data-healthcare>



Open Data in Healthcare

Facteurs de risque

Consommation boissons alcoolisées

Consommation et prix relatif du tabac

Ressources

Médecins praticiens et équipement hospitalier

Caractéristiques socio-sanitaires

Bénéficiaires CMU, CMUC

Affections Longue Durée

Demandeurs d'emploi

Etats de santé

Mortalité et causes de mortalité

Santé au travail

Dépenses

Consommation médicale, dépense courante de santé

Recherche médicale et pharmaceutique

Biologie médicale (hors soins hospitaliers)

Fournisseurs

Caisse Nationale d'Assurance Maladie des Travailleurs Salariés (CNAMTS)

Agence Technique de l'Information Hospitalière (ATIH)

INSEE

EUROSTAT (pays européens)

OCDE



Données, Connaissance

- **Elémentaires**
- **Agrégées**
 - Tranches d'âge
 - Variables catégorielles diverses
 - Localisation (niveau national le plus souvent)
 - Problématiques de la validité, de l'indexation, de l'actualisation et de la diffusion
- **Consolidées**
 - Croisement de plusieurs sources
 - Attribut de consolidation pour différentes analyses : géoréférencement



Données géoréférencées et services d'information

(US) : adresses et coordonnées géographiques des centres participant au Head Start, grand programme du Département de la Santé, de l'éducation et des services sociaux des États-Unis ; opportunités de travail ou indicateurs de santé publique.

(UK) : informations détaillées sur l'ensemble des médecins généralistes, dispensaires ou pharmacies. **NHS publie aussi des données permettant au public de trouver les prestataires de soins les plus appropriés en fonction de nombreux critères comme : type, installations disponibles, efficacité du prestataire, état du patient ou traitement requis, proximité géographique.**

(FR) : Localisation des établissements de santé de Loir-et- Cher. produit par le Conseil Général de Loir-et-Cher, est actualisé annuellement (Open Data 41) (désormais de nombreux autres départements ont emboîté le pas... Open Data 71...)

Gabriella Salzano. Vers un observatoire de recherche sur les données publiques en santé. <http://www.aim2012.bem.edu/papiers/s42-4.pdf>



Applications grand public

Concours Dataconnexions

Le portail Fourmi Santé s'adresse ainsi au grand public. Il propose « d'aider les Français à maîtriser leur budget de santé » en comparant frais de santé, dépassements d'honoraires, etc., des professionnels de santé environnants. Mieux, Fourmi Santé prend en compte ce que les complémentaires remboursent pour **permettre à l'internaute de savoir « ce qu'il va payer de sa poche ».**

<http://www.lefigaro.fr/hightech/2012/05/29/01007-20120529ARTFIG00669-les-donnees-publiques-au-service-du-citoyen.php>



Applications grand public

Aux É. U., l'Agence de protection environnementale (EPA) tient à jour la base de données ouvertes Enforcement and Compliance History Online - ECHO qui permet de savoir rapidement si une entreprise enfreint les lois en vigueur. **Les journalistes peuvent ensuite recourir à ces informations pour sensibiliser le public aux risques d'un développement industriel incontrôlé.**

<http://quebecouvert.org/article/post/donnees-ouvertes-environnement-quebec>



Open Report

- A partir d'un exemple concret
- **Domaine** : « santé – environnement »
- **Objectif** : cartographier les niveaux de pollution et la densité de séjours hospitaliers pour problème respiratoire (Projet entrant dans le cadre du programme régional santé environnement, relayé par le CG 06)



Open Report

- **Idée :**
 - utiliser des données produites en «routine» (et non plus des données de cohorte spécifique)
 - proposer un outil dynamique et interactif
 - étude de faisabilité
 - démarche de prototypage



Open Report

- **Sources de données :**
 - données de consommation hospitalière (PMSI - ATIH)
 - données de pollution atmosphérique (ACRI-ST)
 - données de découpage et repérage géographique administratif (territoires, FINESS...)



Open Report

- **Cibles** :
 - initiales : le chercheur (méthodes d'estimation, analyse des corrélations, conclusion ; mais aussi extraction de données consolidées filtrées)
 - potentielles : le professionnel de santé, le citoyen, le décideur

Année 2007 Département 06

Niveau de pollution

Polluant: DIOXYDE AZOTE - NO2

0 25

Seuil min max

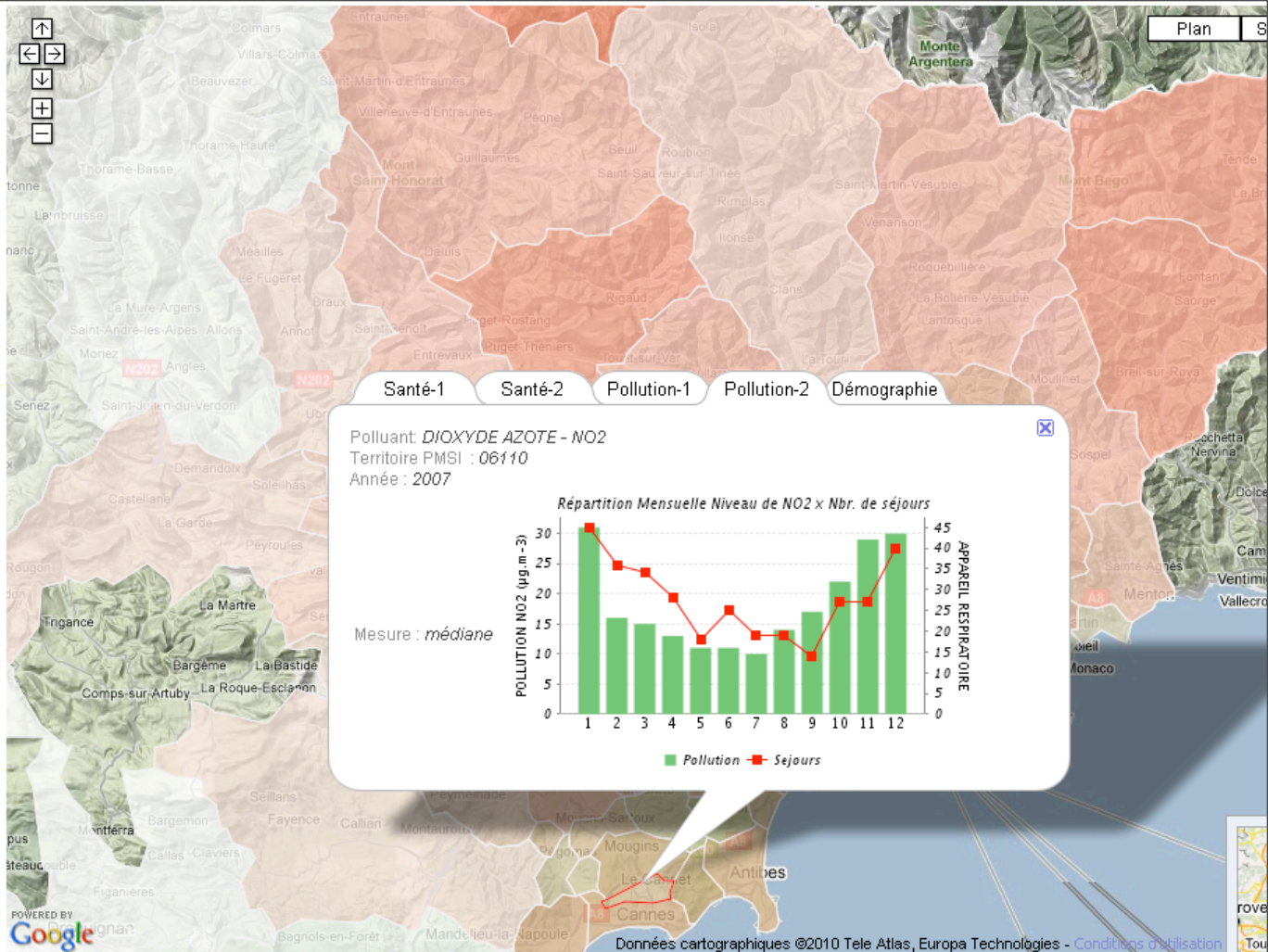
Densité de séjours

Affections: APPAREIL RESPIRATOIRE

0 173

Seuil min max

Opacité 50%



Cliquez sur le territoire 06710, pour visualiser les données ...

Année d'étude (2007)

Territoire PMSI - Démographie (06110)

Commune(s) : LE CANNIET
Superficie : 7 km2
Population : 42 531 (19 641 hommes - 22 890 femmes)
Densité territoriale : 5 566 hab./km2

Polluant (DIOXYDE AZOTE - NO2)

Valeur minimale : 0 µg.m-3
Valeur maximale : 107 µg.m-3
P50 (Médiane) : 17 µg.m-3
P98 : 85 µg.m-3
Ecart-type : 25 µg.m-3

Fonction (APPAREIL RESPIRATOIRE)

Densité de séjours pour 10000 habitants : 78.0
Nb. de séjours : 332 (41.9 % hommes - 58.1 % femmes)
Durée moyenne de séjour : 8.3 jour(s)
Moyenne d'âge : 63.6 ans.

Résultat ? Conséquences ?



Verrous

- **Techniques :**
 - géoréférencement différent
 - contrainte du PMSI avec agrégation de codes postaux
 - recalcul des indices de pollution
 - travail sur le vocabulaire (non réalisé)
 - validité de la corrélation « facteur(s) - consommation » pour traduire la conséquence santé (séjours et non individus)



Verrous

- **Juridiques :**
 - effectif de séjours par type de pathologie et par zone géographique > 5 (CNIL : ne pas permettre de «reconnaitre» l'individu)
- **Ethique :**
 - «juxtaposition des données»
 - transparence de la construction de l'information et de sa diffusion



Perspectives

- **Recherche «cancer-environnement»** en connexion avec l'UMR Espace de l'UNS (soutien cancéropole PACA)
- **Monitoring Urbain Santé Environnement**
 - NCA-IMREDD / à visée professionnel, grand public et décideur
 - identification des sources de données complémentaires (réseaux de capteurs, données de consommation médicale ambulatoire...)
 - techniques de consolidation et niveaux d'agrégation des données (balance bénéfice - risque entre valeur ajoutée et interprétation)
 - méthodes et outils de production d'indicateurs et diffusion



Conclusion

- Santé et conditions environnementales : **thèmes sensibles**
- L'analyse des liens nécessite réflexion & validation
- Retravail des données avant consolidation : **production de nouvelles données**
- Transparence et disponibilité des données signifie-t-elle transparence et disponibilité des analyses ?
- Pour autant, l'exposition des données «publiques» «en silo» est un élément fondateur : apprendre à «utiliser» la donnée