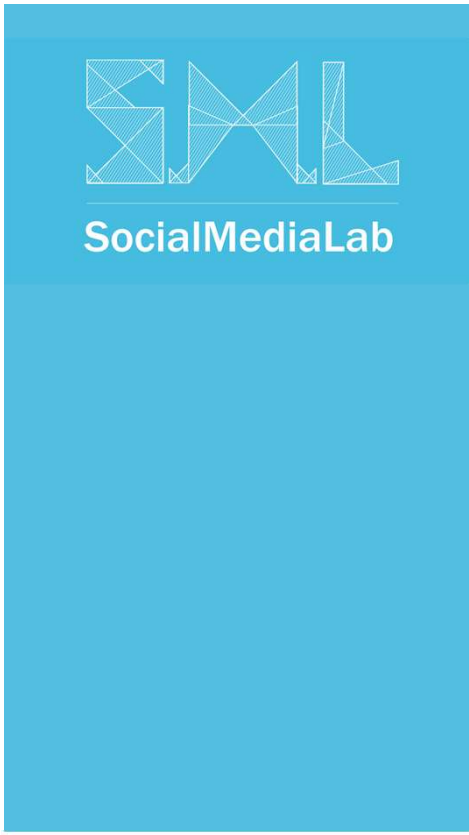


SocialMediaLab

Pour construire **votre avenir**

L'**Europe** et les **autorités publiques** investissent dans plus de **1.000 projets**.

Le Social Media Lab est une fabrique d'intelligibilité des données numériques en contexte professionnel

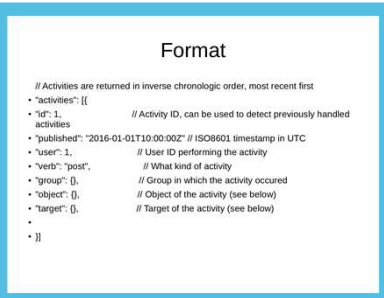


UCLouvain

Pour construire **votre avenir**

L'**Europe** et les **autorités publiques** investissent dans plus de **1.000 projets**.

Social Media Analytics
From data creation to decision making

	UX	INTERFACE	MACHINE
DATA	 Usage	 Production	 Extraction
TREATMENT	 Sensemaking	 Visualization	 Treatment

Our Mission

SocialMediaLab

UCLouvain

Pour construire **votre avenir**

MEWS

L'**Europe** et les **autorités publiques** investissent dans plus de **1.000 projets**.

La fabrique des données en santé: Le cas des registres du cancer en France

Prof. François Lambotte

UCLouvain

Social Media Lab

Francois.lambotte@uclouvain.be

Mise en contexte

- Création du Health Data Hub en 2019

Avec l'allongement de la durée de vie et les progrès de la médecine, les contextes cliniques des patients se complexifient. Par exemple, les personnes âgées souffrent de plus en plus d'une pluralité de pathologies et sont amenées à être suivies par différents professionnels de santé et structures médicales. Le praticien doit alors arbitrer pour offrir le meilleur traitement tout en prenant garde à la qualité de vie du patient, notamment lorsque les traitements sont lourds et admettent d'importants effets secondaires.

Analyser d'importants jeux de données en mobilisant les approches dites d'« intelligence artificielle » peut alors éclairer sur la stratégie thérapeutique à adopter, patient par patient…

(Mission de préfiguration Health Data Hub, p.9)

Ethnographie dans la fabrique des registres du cancer

- Les Registres des Cancers sont des organisations ayant pour mission de réaliser « un recueil continu et exhaustif de données nominatives intéressant un ou plusieurs événements de santé dans une population géographiquement définie, à des fins de recherche et de santé publique, par une équipe ayant les compétences appropriées. » (arrêté du 6/11/1995 relatif au Comité National des Registres).
- Les organisations en charge des Registres ont une double mission :
  - d'une part, elles assurent une veille sanitaire concernant les cancers, en relevant l'incidence et la mortalité liées des cancers sur un territoire donné ;
  - d'autre part, elles effectuent des études et des recherches à partir de leurs données, afin, notamment d'évaluer l'organisation des parcours de soin, les campagnes de prévention ou encore l'influence des inégalités sociales sur l'incidence et/ou la survie.
- Les Registres sont ainsi évalués sur la qualité des données, mais aussi et surtout sur les publications effectuées en leur sein.

DES DISPARITÉS DÉPARTEMENTALES

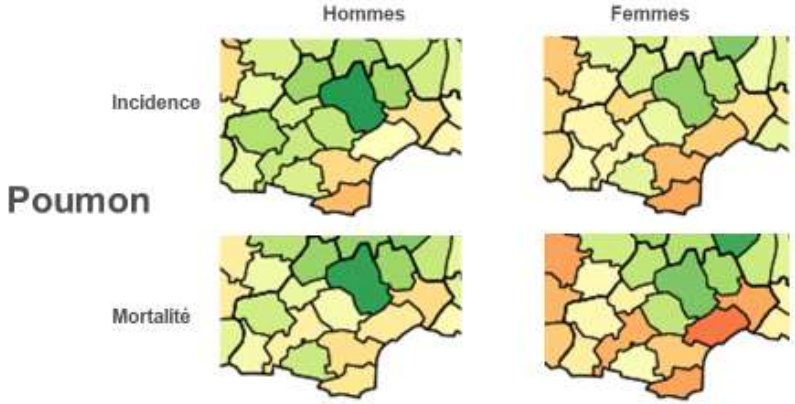
Santé publique France

- **SITUATION FAVORABLE DANS LES DÉPARTEMENTS DU NORD-OUEST**

  Aveyron, Lot, Tarn, Tarn-et-Garonne : notamment pour les cancers liés à l'alcool ou le tabac

- **SUR-INCIDENCE SUR LE POURTOUR MÉDITERRANÉEN OÙ LES FACTEURS DE RISQUE SONT FORTEMENT PRÉVALENTS (TABAC ET DÉFAVORISATION)**

Poumon



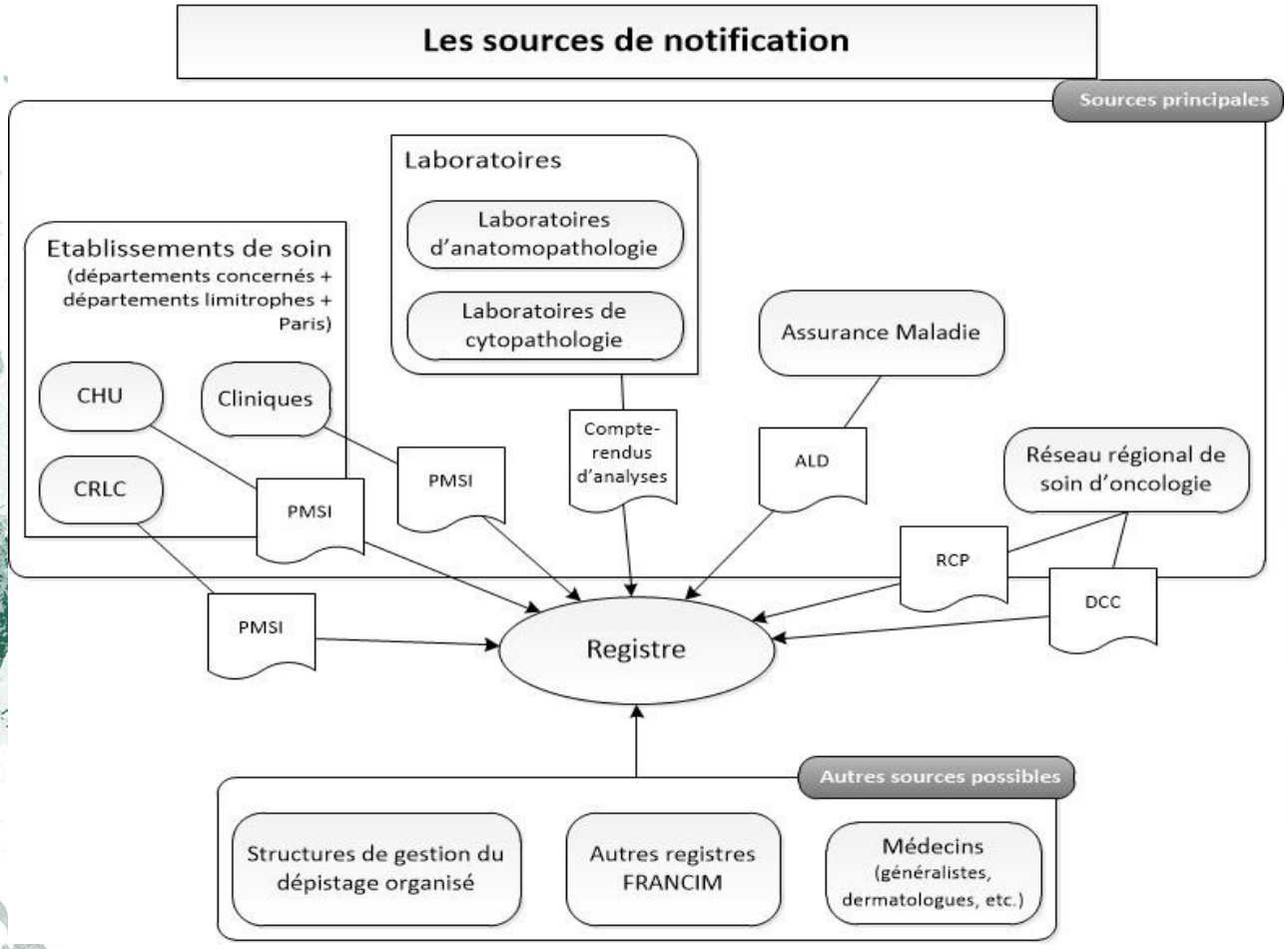
L'idée que les données attendent patiemment sur les étagères d'être identifiées pour être « ouvertes » relève du fantasme (Denis, 2018)

Instaurer les données

Si les « données » des Registres ne sont pas déjà constituées en tant que données, il faut donc que les agents des Registres les *instaurent:*

- en construisant et reconstruisant des collaborations avec des interlocuteurs de différentes organisations, en explorant différents documents et fichiers → **Accéder aux données**
- en construisant des continuités entre des indices pour identifier un «cas » → **Réparer les données**
- en rassemblant des cas toujours singuliers en autant de catégories établies par des définitions internationales → **Standardiser et harmoniser**

Accéder aux données



Un accès négocié à plusieurs niveaux…

« En fait, c'est de l'argent qui va à deux organismes. A l'INSEE, donc derrière, c'est Bercy, le ministère des Finances, qui n'est pas facile à faire bouger. Et il y a des gens qui prennent la moitié de la somme, c'est l'INSERM. Ils prennent la moitié de la somme simplement car ils servent de tiers de confiance.

C'est-à-dire que quand on envoie nos données, nous… moi, j'envoie les données, je m'appelle registre des cancers du Tarn. Donc ça veut dire que toute la liste des gens que j'envoie, ils ont un cancer. Je peux pas envoyer ça à l'INSEE puisque l'INSEE aurait un diagnostic. Donc je l'envoie à l'INSERM, l'INSERM « anonymise », c'est-à-dire enlève… met un numéro à la place de « registre des cancers du Tarn », et transmet.

L'INSEE fait tourner sa moulinette, redescend sur l'INSERM, l'INSERM prend le numéro, remet « registre des cancers du Tarn » et me le renvoie. 150000 euros. C'est bien payé. Et 150000 pour dix minutes de temps machine pour l'INSEE parce que c'est pas long. »


By focusing ethnographic attention towards demystifying how digital data is constituted in situ, we argue that to better understand data futures, we need to account more fully forthe incomplete, contingent and fractured character of digital data.
Pink et al. 2018 p.1

Réparer les données

Alors nous, ça part et ça revient. Quand les enquêtrices ramènent les cas, elles les donnent à la personne qui code. Ça peut être l'une ou l'autre, vu qu'on croise les choses. Et donc la personne qui code, elle va dire « je n'ai pas assez pour coder ». L'enquêtrice ne le sait pas nécessairement qu'il n'y a pas assez pour coder puisqu'elle, elle va collecter toute l'information qu'elle trouve dans l'endroit où elle est. Et elle ne sait pas si ailleurs, on n'aura pas d'autres infos qui seront suffisantes pour compléter. Donc elle arrive avec tout ce qu'elle a trouvé et elle le donne.

Mais si ailleurs, c'était les mêmes informations et puis qu'il y a un trou, la personne qui code va dire « je n'ai pas assez d'informations, il y a un petit problème, je n'arrive pas à comprendre si c'était quelqu'un qui était métastatique, je ne suis pas sûr de la date de diagnostic. C'est bizarre, on a l'impression que c'est quelqu'un qui avait été déjà diagnostiqué il y a deux, trois ans. »

A ce moment-là, on va renvoyer l'enquêtrice ou une autre enquêtrice sur une autre source d'information possible. En plus, nous, on essaie d'optimiser donc on n'enquête pas partout systématiquement. On va sur la source principale d'abord et si là, il n'y a pas assez, on va relancer les autres sources. On serait une énorme structure, ce serait compliqué, mais là, les enquêtrices se parlent entre elles donc s'il y en a une qui revient, elle sait aussi qu'il n'y a pas assez, elle s'en rend compte. Donc elle dit « non, celle-là, tu vois, j'ai trouvé que ça donc il faut demander à machin d'aller dans tel hôpital parce que je n'ai pas assez ».

Réparer les données

Oui, c'est ça, on va le trouver. Après, pour l'échantillon stade, là où on insiste… parce que le problème des stades, de savoir quel est le stade, ça dépend de l'intensité avec laquelle vous avez cherché l'information. C'est-à-dire que quelqu'un qui a juste une radio, on ne risque pas de lui avoir… enfin, très peu de chance de lui avoir vu des métastases.

Donc ***il y a un niveau de preuve, il y a ce qu'on a dit. Si on a dit « pas de métastase » sauf que si le niveau de preuve, c'est un ou zéro, ça veut dire qu'il peut très bien aussi être métastatique. Parce qu'on n'a pas beaucoup cherché.***

Donc on est obligés d'avoir ces deux informations en même temps. C'est-à-dire que si on fournit le stade, on est obligés de fournir la fiabilité du stade. Et puis si on fournit le stade, on essaie de fournir aussi des informations sur le traitement parce que tant qu'à faire, ça ne coûte pas beaucoup plus cher d'avoir l'info sur… parce qu'elle est un peu au même endroit donc on prend aussi les informations sur le traitement.

Standardiser et Harmoniser

Donc on essaye d'avoir des règles minimales et alors, là, nous on est passé à un peu plus complet que l'obligatoire. C'est-à-dire que nous on prend, effectivement, s'il y a deux seins, droite et gauche, on prend les deux. Et on ne prend un in situ que s'il n'y a pas d'invasif sur le même côté. Et ça, c'est à partir des cas de… en plus, il y a des débuts de règle d'enregistrement… enfin, si à partir d'un certain moment, on le fait, on peut pas le faire avant.

Parce que si tous les gens se mettent à le faire pour des cas antérieurs, on sait plus à partir de quel moment on va considérer qu'on le fait… il faut qu'on ait des dates précises et en général, les dates ne portent pas sur la période où on travaille, mais sur la date d'incidence du cancer, ce qui est logique. Après, il y a les cas particuliers… avec des exemples. En fait, c'est fait pour aider les gens qui ne connaissent pas trop non plus. En même temps, il faut quand même connaître un petit peu. Avoir un minimum de culture sur le sujet.

En conclusion

- Les données en santé (en France) sont fragmentées
  - Pour des raisons politiques et organisationnelles
- Travail de négociation
- Travail d'enquête itératif
- Travail d'harmonisation → nécessité de créer une biographie des bases de données

Bibliographie

- Savy, N., Mayère, A., Martin-Scholz, A., & Lambotte, Francois. (2020). La fabrique des données à l'épreuve des programmes de big data. *Actualité et dossier en santé publique*, *Sous presse*.
- Martin-Scholz, A., Mayere, A., & Lambotte, F. (2020). *Instauration et assemblage des données en santé : Le cas des Registres des cancers*. Colloque de la SFSIC, Grenoble.
- Denis, J., & Goëta, S. (2016). *"Brutification" et instauration des données. La fabrique attentionnée de l'open data*. https://hal-mines-paristech.archives-ouvertes.fr/hal-01347301
- Pink, S., Ruckenstein, M., Willim, R., & Duque, M. (2018). Broken data : Conceptualising data in an emerging world. *Big Data & Society*, *5*(1), 205395171775322 8. https://doi.org/10.1177/205395171775322 8

SocialMediaLab

UCLouvain

Pour construire **votre avenir**

L'**Europe** et les **autorités publiques** investissent dans plus de **1.000 projets**.