

Section 4: L'écosystème des sources de données publiques

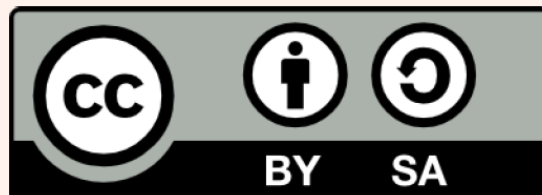
Culture générale des données

Dataactivist, 2018-2019

Ces slides en ligne : <http://dataactivist.coop/SPoSGL/>

Sources : <https://github.com/dataactivist/SPoSGL/>

Les productions de Dataactivist sont librement réutilisables selon les termes de la licence [Creative Commons 4.0 BY-SA](#).



Plan du cours

Podcast introductif section 4, par Timothée Gidoin

1. Les multiples sources et producteurs de données publiques

Regarder la vidéo #DATAGUEULE "Gouvernés par les nombres ? Échecs et maths"

2. Comment bien distinguer les sources ?

Ecouter le podcast "Section 4 - la jungle d'acronymes"

3. L'importance de croiser les sources

Lire l'article d'Alternatives Economiques "Quel est le « bon » seuil de pauvreté ?"

Bibliographie

Quizz section 4

Introduction

Ecoutez le **podcast introductif** de la section 4

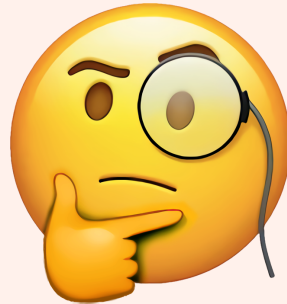


© Xavier Gorce

1. Les multiples sources et producteurs de données publiques

Qui produit des données publiques ?

- A quels acteurs pensez-vous naturellement lorsqu'on parle de données ou de statistiques publiques ?



Qui produit des données publiques ?

- A quels acteurs pensez-vous naturellement lorsqu'on parle de données ou de statistiques publiques ?

Généralement, à des producteurs de statistiques ou d'études comme l'Insee, Eurostat ou l'OCDE



Qui produit des données publiques ?

- Les organismes officiels produisant des **statistiques publiques** sont loin d'être les seuls à produire des données publiques
- **La plupart des administrations ou collectivités produisent des données** dont beaucoup d'entre elles peuvent être ouvertes et réutilisées par tous (cf section 6 et 8 autour de l'*Open data*) et faire l'objet de statistiques.

Exemples:

- Ville de Paris, Région Bretagne, Ministère des armées, Assemblée Nationale, CAF, Pole Emploi, douanes...

Point définition : les statistiques visent d'abord à fournir des renseignements utiles à l'aide de chiffres. La statistique pourrait donc être définie comme de **l'information qu'on peut obtenir à partir d'opérations mathématiques sur des données numériques**. Les statistiques peuvent servir de base à la **prise de décisions**

Commune de Malbos Ardèche Département de l'Ardèche				Liste des Enfants dévorerés par les loups			
Noms et prénoms		âgés	Nom du père et de la mère		Date et observation		
Augiol (Joseph)		14 ans	Père Augiol et Marie Lardillou		le 13. 8 ^{bre} 1812 enterré le même jour		
Augiol (Pierre)		7 ans	Frère du précédent		le 15. 9 ^{bre} 1812 Blessé par les loups, dévoré par Pierre Malbos		
Gours (St. Claude)		6 ans	Matthieu Gours et Marie Mourin		29. 8 ^{bre} 1812 enterré le même jour		
Gours (Rose)		5 ans	Joseph Gours et Rose Guilhard		31. 10 ^{bre} 1812 enterré le même jour		
Gours (Rose)		11 ans	Joseph Gours Marie Meunier		24 Janvier 1813 mourut par les loups 8 mois avant et morte par suite de ses blessures		
					le 4. 9 ^{bre} 1814 date enterré à Malbos la tête et la jambe d'un enfant dévorerés par les loups, on n'a pu découvrir à quel âge il se trouvait, il paraissait être du sexe masculin âgé d'environ 10 ans		
Coste (Rose)		7 ans	Pierre Coste Marie Andarou		10. Mai 1811 enterré le même jour		

Liste d'enfants dévorés par les loups sur le territoire de la commune de Malbos en Ardèche au début du XIXème siècle

Qui produit des données publiques ?

- Il existe des données sur **l'ensemble des sujets de l'action publique**, ces dernières étant consubstantielles à la prise de décisions et à la mise en oeuvre de politiques publiques
- **Ceux qui les produisent** (exemple: une plainte enregistrée par les services de police et de gendarmerie) **ne sont pas toujours ceux qui les exploitent**
- Nous allons nous intéresser dans les prochaines parties à **l'écosystème de ceux qui étudient, analysent et retraitent ces données publiques**, sous la forme de statistiques, d'études, de rapports ou autres publications visant à informer ou influencer le débat public

Du grain à moudre : Regardez [la vidéo #DATAGUEULE "Gouvernés par les nombres ? Échecs et maths" obligatoire](#)

2. L'écosystème des sources : bienvenue dans la jungle des acronymes

La jungle des acronymes

Clever Acronyms: the Holy Grail of Academia

Step 1 Use the loose definition of the word "acronym"

Step 2 Is it coherent? Does it makes sense? What matters is that it *sounds* cool.

Step 3 Missing a letter? Pull out an obscure buzzword that fits!

Step 4 Desperate? Just pick letters from the middle. I'm sure no one will notice.

Step 5 Ignore words that don't contribute. Kind of like your part in the project.

Actually Random Onomastic initials You Make (up)

Types of Acronyms:

- Folksy Names: a cheery name will distract people from the fact your project cost millions	→	A.L.I.C.E., B.O.B., D.A.V.E. ✓	A.D.O.L.F., Z.I.P.P.O., S.I.G.M.U.N.D. ✗
- Aggressive verb/predatory animal: a requirement for getting military funding	→	K.I.L.L., S.H.A.R.K., W.O.L.F. ✓	O.B.L.I.T.E.R.A.T.E. (too many words!), B.U.N.N.Y. ✗
- Greek names: nothing says "Sci-Fi" like a good greek name	→	O.M.E.G.A., A.L.P.H.A., S.I.R.I.U.S. ✓	T.O.G.A., P.I.T.A., T.Z.A.T.Z.I.K.I. ✗

Remember:
Acronyms cleverly reveal one's nimble youthful mastery abbreviating construed rigidly opted nomenclature, yielding monetary awards contracting research overtures not yet manifested!

Bonus points: make your acronym recursive!
recursive
recursive
recursive

Jorge Cham © 2008

WWW.PHDCOMICS.COM

© Jorge Cham, 2008

La jungle des acronymes

Beaucoup des producteurs de statistiques ou d'études publiques ont des noms "*jargoneux*" ou des acronymes peu connus du grand public. Par exemple, connaissez-vous les organismes suivants :



Distinguer les sources

Il n'existe pas de **classification officielle de ces sources**. Il est cependant possible d'esquisser une certaine typologie des acteurs :

- **Service statistique public (SSP)** : l'Insee et les services statistiques ministériels. Les membres du SSP ont la particularité de produire des études en toute indépendance de leur ministère de rattachement
 - Ex: Insee, Dares, Drees, Depp...
- **Organismes publics** : Certains sont rattachés à des ministères ou au premier ministre (ex: CGET, France Stratégie, Direction Générale du Trésor...)

D'autres en sont indépendants (Ex: Défenseur des droits, Ofpra...)

- **Les thinks tanks** : Ce sont des groupes de réflexion privés qui produisent des études sur des thèmes de société
 - Ex: Fondation Jean Jaurès, Terra Nova, Institut Delors...

Distinguer les sources

- **Les organisations internationales**
 - Ex: OCDE, FMI, Banque Mondiale, HCR...
- **Les centres de recherche**
- **Les instituts de sondage ou d'études**
 - Ex: Ifop, Ipsos, Kantar, BVA...

Et d'autres acteurs (**ONG, cabinets de conseil, lobbys**) qui produisent régulièrement des publications pour alimenter ou peser sur le débat public

NB : cette **nomenclature proposée par Datagora**, connaît de nombreuses limites, certains acteurs se considérant comme l'un et l'autre ou redéfinissant leur identité et missions. Par exemple l'OFCE, observatoire affilié à Sciences Po, produit des études de conjoncture économique. Il se définissait précédemment comme un think tank et désormais comme un centre de recherche

Distinguer les sources

Au-delà de cette typologie approximative, il y a des **critères objectifs** sur lesquels il peut être intéressant de s'attarder quand on cherche à décrypter une source :

- Est-elle **indépendante du politique** dans la publication de ses résultats ?
- Est-elle **financée par des fonds privés ou publics** ?
- Quelle type de **gouvernance** ?
 - *Exemple de Rexecode et de l'Ires* : ce sont deux organismes produisant des analyses de conjoncture économique. Cependant, l'un (Rexecode) a dans son **conseil d'administration** plusieurs personnalités issues d'organisations patronales comme le Medef ou l'UIMM. L'autre (Ires) a dans son **Assemblée générale** de nombreux syndicalistes
- A-t-elle une **orientation idéologique ou défend-elle des intérêts** ?
- Est-elle constituée de **véritables chercheurs** ?

Distinguer les sources

Distinguer les sources "primaires" des sources "secondaires"

- Il est fréquent de voir dans les media des chiffres qui sont attribués à des sources qui calculent, analysent, comparent, croisent des données. Ces sources "secondaires" (ex: Cour des comptes) peuvent produire de nouveaux résultats ou des analyses pertinentes mais il faut bien garder à l'esprit qu'elles ne sont pas à l'origine des données.
 - Exemple : Etude "Les dynamiques de population" du CGET dont la plupart des données analysées proviennent de l'Insee ou d'Eurostat
- Lorsque vous cherchez des données "*brutes*", il est préférable d'aller directement sur les sites des sources primaires. Exemples :
 - Site de la CAF pour les données sur les allocataires des APL ;
 - Site de Pole Emploi pour les données sur les demandeurs d'emploi

Distinguer les sources

Ecoutez le **podcast "La jungle d'acronymes"** de la section 4



© Xavier Gorce 2018

3. L'importance de croiser les sources de données

Croiser les sources

Même provenant de sources sérieuses et primaires, il n'est pas inutile de croiser les données trouvées avec d'autres sources qui peuvent également avoir des données sur les mêmes sujets mais avec des chiffres... différents. *Comment expliquer ces écarts ?*

- Les **différences de méthodologie**
 - Exemple 1 : chiffres du chômage de Pole Emploi vs ceux de l'Insee
 - Exemple 2 : Taux de pauvreté, OCDE vs Eurostat
- L'incertitude liée aux **estimations**
 - Certaines données sont issues d'estimations. Elles sont donc souvent temporaires et en attente de révision ou de confirmation
- **L'interprétation des données**
 - Souvent les données sont accompagnées de contexte et d'analyse
 - Ces éléments peuvent différer d'une source à une autre

Les différences de méthodologie

Certains producteurs de données ou d'études n'utilisent pas les mêmes conventions statistiques ou méthodologies pour calculer un même indicateur.

Un exemple connu est celui du **taux de chômage** : même si les tendances sont souvent proches, les chiffres peuvent connaître d'importants écarts entre ceux de l'Insee et de Pole Emploi, pour 2 raisons principalement :

1) La définition du chômeur est différente entre celle de l'Insee (qui reprend la définition du BIT) et celle de Pole Emploi (qui comptabilise les personnes inscrites sur ses listes)

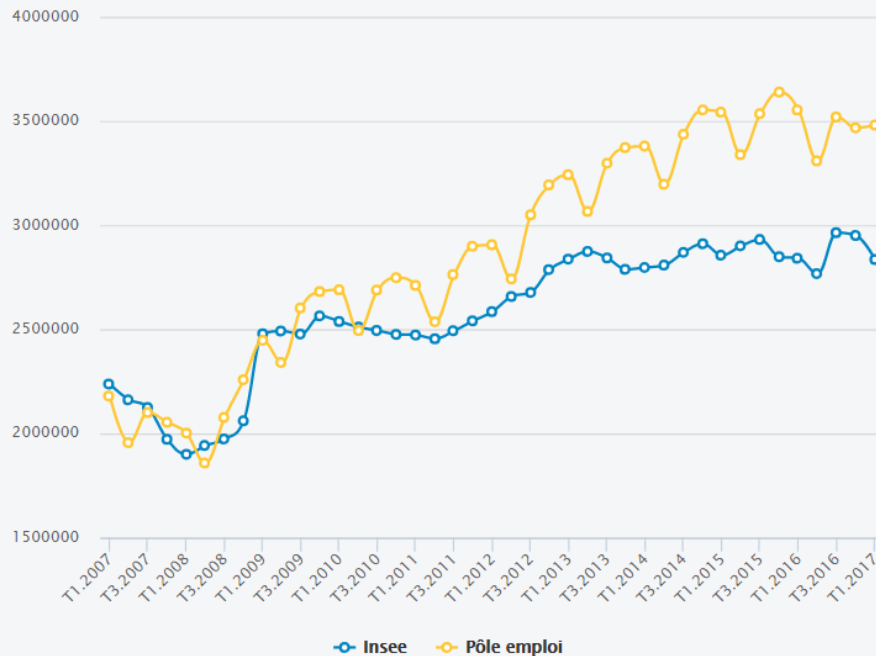
Un chômeur, au sens du BIT, n'est donc pas forcément inscrit à Pôle emploi, et l'inverse est aussi vrai.

2) Les techniques de calcul diffèrent également (technique du sondage pour l'Insee vs nombre des demandeurs d'inscrits dans les Pole Emploi)

Les différences de méthodologie

Le chômage selon l'Insee et Pôle emploi

Nombre de demandeurs d'emploi sans aucune activité (catégorie A) à la fin de chaque trimestre, chiffres de Pôle emploi corrigés des variations saisonnières, chômage au sens du BIT pour l'Insee.

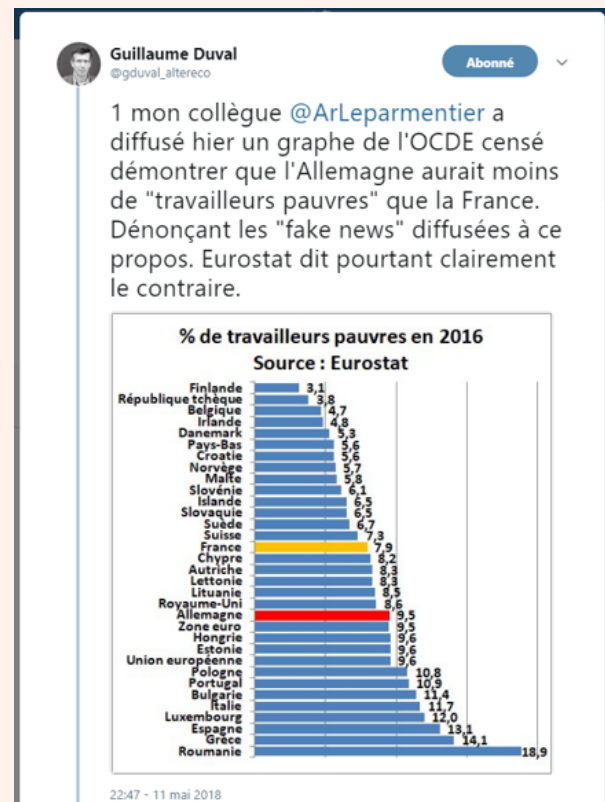
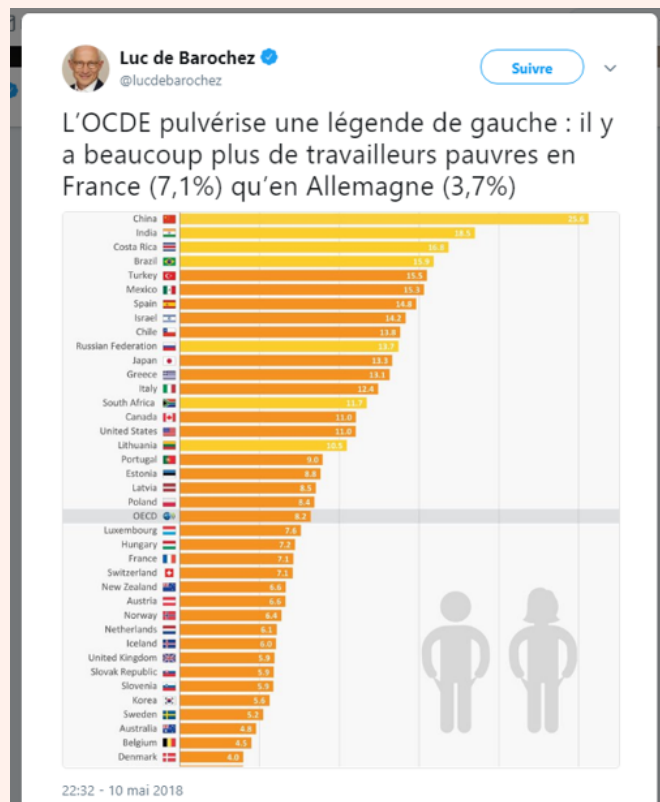


source

Source : Insee et Pôle emploi

Les différences de méthodologie

Nombre de travailleurs pauvres, Allemagne vs France... Qui a raison ?



Les différences de méthodologie

Nombre de travailleurs pauvres, Allemagne vs France... Qui a raison ?

Les deux, car ces données ne mesurent pas exactement la même chose ! Cela est dû à des **nuances de paramètres statistiques** :

Eurostat regarde la part des travailleurs pauvres pris **individuellement** par rapport à l'ensemble des personnes en activité, tandis que l'OCDE mesure le pourcentage des personnes **membres d'un ménage** vivant sous le seuil de pauvreté, et où au moins l'un des adultes travaille. Ici, la composition du foyer entre donc en compte, avec le nombre des adultes en emploi, et une pondération selon le nombre d'enfants. Et cela explique les écarts

source

Les différences de méthodologie

Tous ces chiffres montrent que la part des travailleurs pauvres est plus élevée en Allemagne qu'en France, lorsqu'ils sont comptabilisés individuellement, mais elle est plus basse si l'on prend en compte la composition du foyer

Stefano Scarpetta, directeur de la division emploi à l'OCDE

Plus largement, on peut aussi parfois se poser la question des seuils retenus pour le calcul de certains indicateurs. Par exemple, est-on pauvre avec moins de 60, 50 ou 40% du revenu médian ?

Lire à ce sujet l'article du blog d'alternatives économiques "[Quel est le « bon » seuil de pauvreté ?](#)" de l'économiste Jean Gadrey (**obligatoire**) en réaction à l'article "[9 millions de pauvres, un chiffre exagéré](#)" de Louis Maurin, directeur de l'Observatoire des inégalités

L'incertitude liée aux estimations

De nombreux indicateurs économiques font l'objet d'estimations de plusieurs acteurs. Les estimations peuvent varier de manière significative. Exemple ici avec les estimations - au 31/07/18 - de croissance du PIB Français pour 2018 :

La croissance de la zone euro s'est établie à 1.7% en 2017, un cran en dessous du haut régime de 2.0% (+2,1 % après +2,6 %) et la France ne retrouverait pas, à court terme, le rythme de croissance de l'an dernier : l'activité y progresserait de +1,7 % en moyenne annuelle, après +2,3 % en 2017. Elle

1.7% (Insee)

Synthèse juin 2018

1.8% (Banque de Fr)

Selon les projections de la Banque de France finalisées fin mai 2018, la croissance du PIB français resterait robuste en 2018, à 1,8%. Elle se

2.1 LA SITUATION EN 2017 ET LES PERSPECTIVES POUR 2018 ET 2019

2% (Fr Govt)

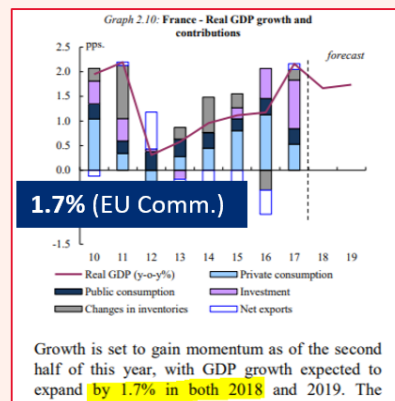
Après avoir atteint +2,0 % en 2017, la croissance s'établirait à nouveau à +2,0 % en 2018 puis +1,9 % en 2019. Après s'être redressée à +1,0 % en 2017, l'inflation atteindrait +1,4 %

1.9% (OECD)

3. DEVELOPMENTS IN INDIVIDUAL OECD AND SELECTED NON-MEMBER ECONOMIES

France: Demand, output and prices

	2014	2015	2016	2017	2018	2019
Current prices EUR billion	2 153.3	1.0	1.1	2.3	1.9	1.9
Percentage changes, volume (2010 prices)						
GDP at market prices						



NB: en décembre 2018, l'Insee a abaissé ses prévisions de croissance pour 2018 à 1,5%

L'interprétation des données

Souvent les données sont accompagnées de **contexte et d'analyse**, qui donnent lieu à une certaine **interprétation**. Ces éléments peuvent différer d'une source à une autre, il est donc utile de les croiser régulièrement.



Ici un exemple classique des divergences d'interprétation des données mensuelles du chômage (Sept 2017). Essayez de remonter directement aux producteurs de données "primaires", ici Pole Emploi, si vous souhaitez approfondir votre analyse en évitant le "*bruit éditorial*"

Bibliographie

Bibliographie

- Vidéo #DATAGUEULE "Gouvernés par les nombres ? Échecs et maths"
- Le Monde, "Chômage : pourquoi les chiffres de l'Insee et de Pôle emploi diffèrent", mis en ligne le 24 août 2017
- Le Monde, "Selon l'OCDE, la part des travailleurs pauvres est deux fois plus élevée en France qu'en Allemagne", mis en ligne le 12 mai 2018
- Jean Gadrey, blog Alternatives Economiques "Quel est le « bon » seuil de pauvreté ?", mis en ligne le 1er mars 2018
- Louis Maurin, blog Alternatives Economiques "9 millions de pauvres, un chiffre exagéré", mis en ligne le 27 février 2018

Quizz section 4 : rdv sur votre espace e-campus !

Merci !

Contact : timothee@dataactivi.st