

# Transcription de la Conférence débat : L'Open Data au service du pouvoir citoyen

16 avril 2025 - Bibliothèque Municipale de Lyon – La Part-Dieu

Conférenciers :

VALENTYNA DYMYTROVA : maître de conférences en sciences de l'information et de la communication à l'université Lyon 3

HERVE RIVANO : professeur en informatique à l'INSA de Lyon

Modérateur :

YVES GODDE : Chef de projet informatique documentaire, Bml

## 00:00:00 Introduction

YVES GODDE

Bonjour à tous, merci d'être venus. Nous allons aujourd'hui vous présenter 2 intervenants, Valentina Dimitrova qui est maître de conférences en sciences de l'information et de la communication à l'université Lyon 3 et HERVE RIVANO, professeur en informatique à l'INSA de Lyon qui vont nous parler de l'Open Data au service du pouvoir citoyen. Je leur laisse la parole.

## 00:00:35 Définition de l'Open Data

VALENTYNA DYMYTROVA

Bonjour à toutes et à tous, merci d'être là. L'Open Data peut être tout d'abord pour définir. Cette notion qui peut paraître très abstraite pour beaucoup entre nous, alors qu'on parle beaucoup des data, des Big data, d'Open Data. L'Open Data c'est à la fois le processus d'ouverture des données publiques et le résultat de ce processus. Et pour commencer notre échange avec vous qui était pensé comme une conférence débat. on a décidé de commencer par rendre cette notion de données plus concrète à travers la question des portails Open Data qui matérialisent à la fois le processus d'ouverture des données publiques et qui permettent de saisir d'une manière plus concrète ce que c'est une donnée. Vous voyez actuellement une capture d'écran du portail Open Data de la métropole de Rennes, qui était le pionnier en France parmi les collectivités territoriales sur l'ouverture des données publiques.

Ce que ce portail nous montre, c'est d'une part les jeux de données de diverses natures, et c'est là qu'on voit que l'Open Data peut toucher à des différents domaines et des différentes thématiques qui concernent aussi bien les politiques publiques que la vie des citoyens, des habitants d'une ville, d'une métropole, des citoyens, d'une collectivité. vous voyez des données, des transports, des données liées à l'école ? L'information de différents jeux de données. On voit également le nombre et ce nombre renvoie aussi à quelque chose qui est très propre à l'Open Data et à l'idée des données en général. C'est l'idée de la quantification, c'est à dire que derrière le processus de l'ouverture des données et derrière la production d'une donnée, il y a la mesure de quelque chose.

Ces mesures-là, de différentes manières se retrouvent, se matérialisent dans ces portails qui s'inscrivent dans des stratégies d'ouverture des données. Pourquoi des stratégies d'ouverture des données à l'échelle territoriale ? D'une part, bien sûr, il peut y avoir des volontés politiques, comme c'était le cas de la métropole de Rennes tout au début, avant les injonctions législatives car depuis 2010, on a effectivement des injonctions législatives qui obligent des collectivités d'une taille supérieure à 3500 habitants à ouvrir leurs données, soit en créant leur propre portail, ce que peuvent se permettre des grandes collectivités, soit en les déposant sur un autre portail, un méga portail, on va dire data.gouv.fr le portail national. C'est un dispositif socio technique qui articule de nombreux acteurs. Pourquoi de nombreux acteurs ? Parce que derrière, il faut avoir des producteurs de données, des agents de la ville par exemple, ou des entreprises qui, dans le cadre de la délégation des services publics, produisent différents types de données. Par exemple, dans le domaine des transports, on pourrait citer Syra, ce sont des producteurs qui vont être en lien avec ce portail pour le nourrir de données.

Mais il y aura effectivement dans cette fabrique de la donnée des personnes qui vont aussi les agréger, les nettoyer, les rendre propres. Des gestionnaires de la ville qui vont aussi décider quel jeu de données publier et lesquels ne le seront pas. Il y a beaucoup de travail sur la donnée avant qu'elle ne soit disponible ici. Ça fait un écosystème de réutilisateurs, de gens qui vont proposer de faire quelque chose avec ces données. Parce que la donnée en soi ne sert pas à grand-chose. Et derrière la métaphore de l'or noir qu'on associe souvent à des données, il y a un gros travail qui consiste à sortir cette valeur de la donnée. Et pour sortir cette valeur de la donnée, il faut les réutiliser et la réutilisation ne va pas de soi. Mais au final, peu de citoyens s'en emparent. En tout cas peu de citoyens sans compétences techniques peuvent s'en emparer. Parce que ça nécessite une acculturation à la donnée, ça nécessite de comprendre ce qu'est la donnée, ce qu'on peut en faire. Sur cette capture d'écran vous voyez une forme de réutilisation qui est proposée, c'est la ligne de bus du réseau des transports en commun lyonnais. c'est un jeu de données éditorialisé par le portail pour montrer l'utilité de la donnée. Mais c'est à partir de jeux de données, qu'on peut rendre l'une des formes que ça peut prendre. C'est une cartographie qui nous rendra visible, par exemple la manière dont le Grand Lyon est desservi en transport en commun. Pour devenir utile à des citoyens comme vous et moi, la donnée a besoin des médiations et des réutilisations.

## 00:06:16 Historique et idéologie

YVES GODDE

Pouvez-vous nous faire un historique de l'Open Data en général et puis nous expliquer quelle forme territoriale peut prendre l'Open Data.

## 00:06:16

VALENTYNA DYMYTROVA

L'historique va être long. Je vais essayer d'être rapide et faire en sorte que vous me compreniez. On parle de l'Open Data à partir des années 2008, 2007 à l'époque mais faut pas oublier que la notion en tant que telle vient de la sociologie des sciences. C'est dans le domaine de la science que la question s'est posée : est-ce qu'on rend disponible des recherches scientifiques aux citoyens ? Cette question émerge dans les années 40, grâce à Robert Merton, grand sociologue des sciences. Mais on revient vers cette notion progressivement. Et même si on ne parle pas de l'Open Data, il y a cet esprit d'ouverture, de transparence dans les lois qui concernent par exemple la France. A la fin des années

70, on ne parle pas de données à l'époque, mais de documents administratifs qu'il faut rendre publics. les lois CADA. Notamment, la création de la Commission CADA. Cette question de l'Open Data n'est pas toujours associée sémantiquement à cette expression, mais est dans l'air du temps. Certains peuvent faire remonter cette idée de transparence à la Révolution française. Dans le contexte actuel, pourquoi on peut parler d'idéologie ? Parce que derrière, il y a l'idée que les gouvernements de nos sociétés démocratiques doit être redevables aux citoyens qui les ont élus, et donc de leur rendre des comptes sur des informations qui concernent la vie de la cité, les administrations, la vie de l'administration et les décisions qui sont prises. Et dans le cadre de cette transparence démocratique, on va avoir ce mouvement-là. Mais de l'autre côté, il y a aussi la notion différente qui arrive avec l'idée qui se développe en parallèle du libéralisme informationnel. C'est à dire l'idée qu'on vit dans une société ouverte où tout circule et où on n'a pas des données stockées dans un endroit, mais qu'on a des flux, des données, des flux d'informations et en plus qui doivent circuler sans frontières d'une manière libre. C'est un autre courant idéologique qui va se retrouver aussi dans cette idée d'Open Data. Les politiques de l'Open Data peuvent aussi correspondre à ces attentes-là, au croisement de ces idéaux de transparence, de ces demandes qui vont dans le sens de libéralisme informationnel. On retrouve les politiques qui se mettent en place progressivement, qui commencent aux États-Unis. Barack Obama, a joué un rôle dans la popularisation de ce mouvement en incluant dans son programme politique, en politisant cette question. On retrouve effectivement la mise en place des politiques Open Data, des sommets de l'Open Data qui vont structurer le paysage des pays démocratiques, des index d'Open Data parce qu'on va mesurer en fait l'ouverture des données. Ça se retrouve au niveau territorial plus près de nous parce que les villes aussi vont s'en emparer dans l'idée double, d'une part de la transparence démocratique, mais d'autre part dans l'idée d'amener de l'économie, de faire vivre leur territoire par la création des services numériques. On a cette approche servicielle de la ville qui en résulte accompagnée de l'idéologie des Smartcities, à savoir des villes qui sont gouvernées grâce à des données managées. Open North Foundation a élaboré en 2005 des critères d'ouverture des données. Qu'est ce qui est ouvert ? Qu'est ce qui n'est pas ouvert ?

C'est très simple : une donnée publique dans un document comme un PDF, ce n'est pas une donnée qui cocherait les cases de l'Open Data, parce que pas exploitable directement par des formats qu'on appelle lisibles par des machines. Il faudrait des données renseignées dans des tableaux, des tableurs type excel. Et le risque dans tout ça, c'est de tomber dans le piège de prendre la donnée comme un fait ultime. On justifie beaucoup de décisions à partir des données. Mais cette quantification permettait toujours de prendre des décisions, y compris politiques. La difficulté peut être le danger derrière pour les démocraties, y compris pour les démocraties, c'est de penser la donnée comme quelque chose d'objectif, alors que « les données ne sont jamais données mais sont des obtenues ». On les obtient par des processus techniques qui sont complexes et peuvent comporter des biais et des choix. Parce qu'on ne peut pas tout mesurer. Il y a aussi la question des instruments de mesure, des critères de mesure, qu'est-ce qu'on calcule, qu'est-ce qu'on prend en compte ? L'interface du portail peut vous présenter une donnée comme étant ultime le nombre de lignes, le nombre de bus, le nombre de places disponibles sur un parking.

Il y a des choses qui rendent cette donnée plus complexe qu'elle ne le paraît et qui nécessite de réfléchir à comment cette donnée a été obtenue, produite, mais aussi quelles sont les médiations qui étaient mises en place pour arriver à cette donnée et surtout dans les réutilisations. Parce qu'au final, qui parmi vous est allé voir le portail Open Data de Lyon ou de Rennes ou [data.gouv.fr](http://data.gouv.fr) ? Peut-être des collègues, des spécialistes mais autrement pas grand monde alors qu'on nous présente comme objectif sur les applications mobiles qu'on a dans nos poches et qui nous disent que tu as tant de temps pour prendre ton tram par exemple. Il y a des médiations qui sont mises en place et ces médiations c'est ça qu'il faut questionner.

00:13:20 Médiation des données

YVES GODDE

Quel type de médiation de données on met en place et surtout quels sont les acteurs qui interviennent dans cette médiation ?

00:13:40

VALENTYNA DYMYTROVA

Il y a une multiplicité d'acteurs. Vous avez des schémas qui viennent d'un ancien projet ANR où j'ai participé avec l'école de mines de Saint Étienne. Moi j'assurais le côté plutôt sciences humaines et sociales. Dans ce projet, on s'est intéressé à quels sont les acteurs de cet écosystème de réutilisation des données ? On a identifié des développeurs, des data scientists, des Data analysts et des data journalists. Au final, pas du tout des citoyens, même si on peut retrouver parfois des applications citoyennes mais qui sont très limitées. Ces acteurs ont leurs pratiques professionnelles avec les formats des données dont ils ont l'habitude, avec leur finalité des données, qu'est-ce qu'ils cherchent dans la donnée ? Parce qu'un journaliste ne va pas chercher dans la donnée la même chose qu'un développeur. Un développeur va se soucier surtout de l'efficacité de son service.

Si on est dans le cadre du développement d'une application mobile par exemple. L'idée pour lui ce sera de trouver des jeux de données déjà disponibles et qui correspondent à son objectif. Par exemple, fluidifier le trafic routier dans une ville alors qu'un data journalist va prendre la donnée va aussi chercher la donnée, identifier, nettoyer, homogénéiser, croiser avec d'autres jeux de données, mais va y chercher peut être autre chose, pour répondre à une enquête, pour répondre à une question ou alors aussi pour explorer. Parce qu'on a aussi ce 4e paradigme dans la science qui arrive avec ces explorations des données. Ça m'est arrivé d'avoir des journalistes qui ne savent pas exactement ce qu'ils cherchent mais qui examinent des données, des jeux de données, des tableurs, en pensant y trouver quelque chose. Il y a aussi cette manière de travailler. L'idée derrière, c'est que ça illustre la notion de médiation qui est derrière la production de la donnée, mais aussi de la production d'un service basé sur des données.

00:15:30

YVES GODDE

On a vu la question des acteurs mais en coulisses, il y a des traitements qu'on applique à ces données ? Vous avez parlé tout à l'heure de nettoyage...quels sont les acteurs qui interviennent dans ce processus ?

00:16:08

VALENTYNA DYMYTROVA

Les acteurs peuvent être divers et variés. Ça peut être des data ingénieurs où le cas échéant, des data journalists qui font le nettoyage, il y a des outils aussi techniques. Derrière les interfaces qu'on trouve souvent faciles et pratiques, il y a ce travail invisible sur la donnée. Nous on voit le propre ; on ne voit

pas, les brouillons et du travail de coulisses, mais ce travail de coulisses de l'Open Data est une notion qui a été beaucoup travaillé par Samuel Goëta de Dataactivist.

## 00:16:55 Travail invisible sur les données

Question du public

N'oublie-t-on pas les annotateurs exploités dans les pays pourvoyeurs de main d'oeuvre bon marché ?

## 00:17:09

VALENTYNA DYMYTROVA

Oui, il peut avoir des annotateurs sur la donnée bien sûr. Effectivement, se pose la question du travail sur les données. Qui dit données dit aussi métadonnées. Produire des métadonnées, ça veut dire annoter les données, donner du sens à des données. Dans des processus de création de valeur à partir d'une donnée, il va y avoir aussi le travail invisible des petites mains. Des petites mains qui ont toujours existé dans l'histoire de la science. Des petites mains existaient à l'époque de Robert Merton qui propulsait l'idée d'Open science. Mais aujourd'hui, on a d'autres petites mains qui existent, qui sont d'ailleurs souvent dans des pays autres que la France, et qui sont exploités. Ce sont de nouvelles formes d'exploitation aussi basées sur des chaînes techniques.

## 00:18:07 Cartographie et représentations

HERVE RIVANO

Sur la question du prolétariat numérique, on y reviendra à la fin sur les modèles économiques qui arrivent. Vous voyez une carte et en fait on vous a beaucoup montré 2 cartes. Parce que dans la façon de faire de la médiation, de rendre la donnée accessible à tout le monde, la cartographie c'est très puissant parce qu'on se représente assez facilement notre environnement sur une carte. Là il y a 2 exemples, vous en connaissez forcément un des 2, peut être que vous connaissez les 2, celui de gauche, c'est OpenStreetMap, donc c'est de la cartographie, comme Google Maps, mais ouverte, open source.

C'est une représentation assez naturelle, mais un petit peu différente. On les comparera tout à l'heure, et ça, c'est COVID tracker qu'on a tous regardé assidûment, il y a quelques années. J'ai commencé avec COVID tracker parce que c'est un exemple intéressant. La loi sur les données ouvertes impose au ministère de la santé de mettre plein de données en ligne. Ils mettent plein de données en ligne, mais elles sont strictement illisibles. C'est un travail d'expert d'aller fouiller dedans et pendant le COVID il y a une personne qui s'est dit que ce n'était pas possible d'avoir ces données illisibles et qui a fait ces visualisations-là. Il y a un travail derrière sur la donnée, c'est à dire que la donnée telle que partagée par le ministère ne ressemble pas du tout à ça. Il y a une façon de représenter, il y a un choix de représentation des données, il y a un calcul sur les données pour arriver à ces courbes-là. Il n'est pas très profond au sens où il n'y a pas beaucoup de choix dans la façon de faire, il y a un choix de représentation mais il n'y a pas une manipulation des données. En tout cas le chemin entre la donnée et ça est assez direct mais pas faisable par tout le monde. Par contre lire ça c'est faisable par tout le monde. Et en donnant à voir ces informations-là, il a fait des choix éditoriaux. Mais en donnant à voir

ça, l'air de rien, il a changé totalement la façon dont le gouvernement a piloté la lutte contre la pandémie. Il y a un avant COVID tracker et un après COVID tracker, juste dans la relation politique des citoyens aux politiques publiques de gestion de la pandémie.

00:20:54

Ce travail-là de médiation consiste à prendre des chiffres complètement abscons et à en faire des courbes rouge et bleu. Ce travail-là est un travail démocratique énorme qui lui a valu la légion d'honneur. Par ailleurs, certains affichaient un sourire crispé quand ils lui ont donné, parce que ça les a bousculés, mais ça a été le moment de redonner du pouvoir aux citoyens. Openstreetmap c'est un autre sujet parce que c'est un commun. Les associations comme Wikipédia, ce sont des gens qui mettent de l'argent en commun. C'est une association mondiale qui se donne des moyens d'avoir des serveurs, d'avoir ces services-là. C'est matériel en fait, il y a des ordinateurs qui tournent, il y a du code qui est développé pour faire de la cartographie alors qu'on pourrait se dire, pas besoin, il y a Google qui le fait. Ça sert justement à ne pas être sous la férule de Google et donc de donner les moyens à tout un chacun.

00:21:51

Ça consiste à éditer cette carte et à rajouter de l'information dessus. C'est une plateforme pour que chaque citoyen puisse transmettre de l'information et en réalité les données d'OpenStreetMap sont bonnes parce qu'il y a beaucoup de gens qui viennent les corriger quand il y a des erreurs, des citoyens, des collectivités qui se disent que finalement c'est bien de mettre de la donnée fiable là-dedans. On voit la bibliothèque municipale de Lyon, on voit la gare SNCF. Les choix qui sont faits de représentation de l'information sont les choix des utilisateurs et des contributeurs d'OpenStreetMap. C'est un choix collectif à défaut d'être démocratique à minima. C'est aussi le choix des communs. Les deux utilités, de la donnée ouverte, c'est avoir un commun de cette donnée avec des choix de représentation collectif et surtout le travail invisible, c'est tout un chacun qui peut le faire.

00:22:47

C'est un travail qui demande à être un minimum spécialiste mais qui a permis à tout un chacun d'avoir un regard critique sur une politique publique et d'influer sur la façon dont la politique publique se mène. Et il y a un 3e usage de l'Open Data. C'est France stratégie, un service public qui a été renommé récemment. C'est l'État qui donne à voir des cartes. Dans cet exemple on a des répartitions de niveau de revenus ou de patrimoine. C'est le dernier décile de revenus et le premier décile. Ces cartographies-là permettent de se rendre compte de comment les politiques publiques ont un impact sur la répartition des richesses dans le territoire. Sur ce sujet-là il y en a plein d'autres. Cette utilisation-là, c'est une façon qu'a choisi l'État de donner à voir ces données qui sont disponibles en brut sur data.gouv.fr, de les re cartographier de le remédier pour faire comprendre ces politiques publiques aux citoyens et faire un compte rendu démocratique. Dans les 3 exemples il y a 3 circuits de reprise du pouvoir par le citoyen via la médiation de la donnée qui ont des objectifs et des directions différentes. De l'État vers le citoyen, du citoyen vers l'État et puis du commun qui est plus horizontal.

00:24:41

YVES GODDE

Toodego c'est un guichet qui apporte quoi dans l'organisation des services urbains ?

## 00:24:42 Services numériques citoyens

VALENTYNA DYMYTROVA

Ce guichet est intéressant à double titre. Tout à l'heure on a parlé de l'histoire de l'Open Data et sur le territoire lyonnais. Il y a une histoire aussi plus circonscrite. L'arrivée de ce service-là permettait d'impliquer les citoyens dans la production de la donnée, mais aussi de leur proposer un service très concret basé sur les données. Ce guichet numérique, a pour objectif les données effectivement. Celles qui sont produites ne sont pas admises en Open Data mais il y a le lien entre le portail Open Data. Cette plateforme nourrit des informations ; apporte par les jeux de données des informations très concrètes à la disposition des citoyens. Ça illustre aussi les politiques publiques. gérer la ville aussi, montrer comment les citoyens s'en emparent, ce qu'ils peuvent en faire, le potentiel démocratique. On l'a choisi pour illustrer le côté serviciel. Sauf que les données peuvent aussi être utiles parce qu'elles peuvent nourrir des services numériques à destination des citoyens.

## 00:26:10

HERVE RIVANO

Je vais mettre un regard critique dessus. Toodego a une utilité de marketing territorial qui est de dire la ville se donne les moyens de communiquer pour dire regardez, nous sommes moderne... il y a un vrai sujet compliqué, c'est comment je capte de l'information sur mon territoire. Soit je mets de l'argent pour avoir une infrastructure de mesure, je paye des agents pour aller vérifier l'état des routes ou je demande aux citoyens de dire s'il y a un nid de poule à tel endroit. C'est une tarte à la crème des élections de dire je vais mettre en place un système Smart City où tout le monde pourra dire quand il y a un tag ou quand il y a un nid de poule Pour moi le marqueur de l'inutilité de ce truc-là, c'était à Meaux, c'était ridicule. Toodego, ne fait pas que ça mais fait ça aussi. Je pense que ça ne sert pas à grand-chose. Il faudrait vérifier, mais je pense qu'il n'y a pas beaucoup d'informations pertinentes qui remontent. Par contre il y a une autre volonté dans Toodego qui est de d'afficher les politiques publiques, d'afficher les services et de donner la possibilité aux citoyens de venir interroger la vision qu'a la ville de son territoire, notamment si dans Toodego je peux demander un itinéraire et l'itinéraire qui va être calculé par Toodego va favoriser le transport décarboné, l'utilisation de vélo, l'utilisation des TCL par rapport à la voiture ou la marche. Parce que il y a une politique publique de déplacement qu'elle essaye de mettre en œuvre et on verra un tout petit peu plus tard pourquoi c'est un échec, parce que elle se retrouve en compétition directe avec des services similaires produits par Google, avec une puissance de tir incommensurable et à la fin, ça n'est plus qu'un outil de marketing.

## 00:28:10

VALENTYNA DYMYTROVA

Par un outil de marketing, on peut critiquer aussi la question. Ce sont des arguments classiques, quand on a par rapport à l'inégalité numérique que ça exclut certains. On peut retrouver des arguments pour et contre en revanche sur ce côté déplacement parce que les données de la mobilité sont parmi les premières à être ouvertes, parmi les premières à être réutilisées, le nombre de hackathons qui ont été faits autour de ces données, le nombre d'applis qui apparaissent et qui meurent avant 3 ans, faute de moyens, à partir de ces données. Je pense que les données de mobilité c'est un exemple intéressant, mais y a pas que des données de mobilité sur Toodego sur ce guichet. Le guichet unique je pense que

c'est important que les pouvoirs publics s'en emparent aussi, créent aussi leurs propres services. Et même si pour l'instant ça ne prend pas ou ça connaît moins de succès, pour moi ce sont aussi des solutions importantes parce que justement contrôlées par les pouvoirs publics, parce que prenant en compte les spécificités du territoire, des difficultés du territoire, des inégalités socio-économiques d'un territoire. Je suis plus favorable à ce type de guichet qu'à des guichets créés. Parce qu'on a aussi des plateformes qui essaient aujourd'hui de s'emparer des plateformes privées. Je veux dire qu'ils essaient devenir ce guichet unique de notre accès au territoire alors que ce sont des plateformes numériques qui sont aux États-Unis ou en Chine. On avait une cartographie participative qui a été mise en place par une association Bellebouffe avec un ancien de tuba d'ailleurs qui a beaucoup contribué. C'était au moment du COVID où on était confiné, où il y avait beaucoup de commerces qui étaient fermés, où on était limité aussi dans nos déplacements. L'idée était c'est de renseigner sur une carte des endroits où on pouvait se procurer de la nourriture, de l'alimentation. Ça répondait aussi aux besoins de résilience d'un territoire qui est aussi quelque chose de très important et qu'on découvre dans d'autres contextes des guerres ou des crises majeures dans d'autres pays.

C'est un exemple qui est réussi et qui d'ailleurs a été récupéré en partie par les pouvoirs publics parce que ça montre aussi le lien entre le citoyen, comment le citoyen peut nourrir le collectif, peut être le commun. une question très importante pour tous. La question « qui gouverne la ville ? Est-ce celui qui produit les données qui gouverne la ville ? Est-ce que c'est celui qui a la maîtrise des données du territoire ? Ou est-ce des réutilisateurs privés qui peuvent s'emparer de ces données d'une manière gratuite ? Parce que la loi pour la République numérique oblige de produire les données. Mais la question de redevance est très compliquée dans cette affaire. Comment maintenir le pouvoir de la ville sur ses propres données, la maîtrise de ses propres données et des services basés sur ces données ? Dans le domaine de la mobilité on le voit d'une manière criante à Lyon, il y avait un gros projet européen Optimod qui visait la création d'une plateforme de mobilité et en regroupant des différents acteurs du territoire, des producteurs de la donnée, mais aussi des réutilisateurs parce que on peut être à la fois des producteurs et utilisateurs et qui aujourd'hui n'en parlent plus. C'est devenu une agence de mobilité qui prend d'autres formes. La question de l'usage par les citoyens. Les citoyens se sont tournés pour des solutions plus privées en choisissant Waze et non pas l'application de la ville. Il y a aussi la question de l'usage qui est très importante. On a d'un côté des acteurs privés et d'autre part des acteurs publics qui cherchent à proposer leurs services pour la ville. Mais derrière les services il y a aussi la vision de la ville. Parce que Waze ne va pas se soucier où il va passer et faire passer les voitures. Est-ce que c'est face à une école ? Il va juste essayer de fluidifier le trafic à partir des données dont il dispose et que d'ailleurs remonte tous les utilisateurs de Waze du moment où vous l'utilisez. Vous aussi vous remontez les données concernant la vitesse avec laquelle vous avancez dans une rue. Toutes ces questions-là se posent et font que cette question est très épineuse.

## 00:33:40 Enjeux politiques et démocratiques

HERVE RIVANO

Le premier point c'est le gratuit qui pose un vrai souci parce que c'est l'idée que la loi impose aux collectivités de rendre leurs données accessibles. Et sans rétribution financière pour y avoir accès. Mais ça n'est pas gratuit parce que pour mettre la donnée en ligne, il faut la stocker quelque part. Il faut payer des serveurs, il faut payer la facture d'électricité et l'impact environnemental de toute l'infrastructure numérique qui n'est pas dans le cloud. C'est du cuivre, des matériaux rares et de l'énergie. Mais c'est pire que ça. C'est que pour que les données de mobilité soient prises en compte par Google, il faut qu'elles soient dans un format particulier et Google impose que ça soit le fournisseur de données qui paye le coût de la transformation des données pour les mettre à son propre format,



ce qui est génial. C'est un racket total. C'est quoi l'objectif de Google? C'est centré sur Google parce que c'est eux les meilleurs en mobilité. C'est quoi l'objectif de Google quand il fait Waze ou Google Maps ?

C'est quoi l'objectif de la métropole quand elle fait Toodego ou optimod ? La métropole a une délibération publique avec laquelle on est d'accord ou pas d'accord, mais en tout cas à un moment on s'exprime en tant que citoyen par un vote. Il y a une représentation démocratique des citoyens qui dit, on va faire un plan de déplacement urbain, on va favoriser tel type de mode, on va essayer d'éviter que les gens passent à 08h30 devant les écoles quand ils sont pressés avec des SUV de 40 tonnes.

Waze, Google, ont intérêt à ce qu'on passe devant des pubs. C'est ça son modèle économique, c'est d'exposer les utilisateurs à la publicité. Il est financé par les annonceurs qui viennent mettre de la pub chez lui. Ce sont 2 objectifs strictement différents. Et le pouvoir à la fin, va à celui dont on utilise l'application. Le problème c'est que c'est Google qui a gagné, principalement parce qu'il est disponible sur tous les téléphones tandis que Toodego n'est pas disponible sur tous les téléphones. Toodego n'est pas développé par la même armée de développeurs à moitié mal payés au fin fond de l'Inde ; c'est pas les mêmes budgets, on ne peut pas lutter et puis Toodego ça marche à Lyon, et puis il faudrait celle de Bordeaux et puis il faudrait celle de Paris, il faudrait celle de Dijon.

Pour vous montrer l'autre impact de ça sur un autre sujet qui est plus la mobilité, qui est juste la cartographie de l'espace. Cette image c'est OpenStreetMap , c'est le citoyen qui met ses infos, ça c'est la représentation du même territoire par Google. Tout ce qui est bleu c'est du commerce. Tout ce qui est orange c'est du commerce de bouffe. Et en orange, c'est des hôtels. Il y a un petit truc qui dit qu'il y a une gare. Il y a aussi du commerce de bouffe. Il y a la gare, il y a la bibliothèque municipale parce que c'est un service public et les gens ont dit qu'ils voulaient que les services publics apparaissent.

Il y a une zone qui dit attention, il y a des travaux, il y a une voie cyclable provisoire, il y a un accès vélo station, il y a le tunnel vivier Merle qui est indiqué. Il y a beaucoup plus d'informations sur la vie citoyenne à cet endroit-là et c'est normal puisque c'est les citoyens qui ont décidé ce qu'ils voulaient mettre, ce qu'ils voulaient représenter. Si on était dans une ville de gens qui se baladaient tous en bagnole, il y aurait toutes les places de stationnement. On a ce que Google veut nous montrer pour qu'on y aille. C'est ça la différence en fait. Finalement, c'est ce qu'on accepte d'être pris en charge par une entreprise qui a un intérêt économique à nous exposer à certains services ou pas ? Ou est-ce qu'on veut faire l'effort de contribuer à une création commune où on va pouvoir régler, choisir soi-même ce qu'on veut. Il y a de l'information qui est collectivement proposée et collectivement pas forcément débattue parce que ce n'est pas OpenStreetMap. Ce n'est pas un outil démocratique, c'est un outil de commun, c'est un petit peu différent. C'est plus proche de la démocratie que l'autre sur le capitalisme.

Une autre utilité très importante de la donnée ouverte et de la remédiation de la donnée, c'est dans le même sens que COVID Tracker, mais en plus agressif. Ça c'est une carte de pollution de la ville de Lyon. La donnée de qualité de l'air, elle est disponible, fournie par une association agréée, un opérateur financé pour des missions réglementaires. C'est une association qui fournit des données, qui est la seule qui a le droit de donner des données de pollution, notamment à la préfecture. Leurs données sont disponibles et il y a Greenpeace qui a pris ces données-là, qui les a mises sur une carte.

Alors ils ont choisi l'échelle de couleur parce qu'il y a une remédiation dessus pour dire que le rouge c'est agressif, ça représente le danger dans beaucoup dans notre culture en tout cas, le vert ça représente les endroits où c'est bien. L'échelle de couleur est faite pour donner un côté inquiétant. Mais ceci dit, la situation est inquiétante et ils ont choisi de prendre un autre jeu de données ouvertes disponible par la métropole qui est l'ensemble des écoles. Et le titre c'est « votre enfant respire-t-il un air trop pollué ? » La carte ne répond pas exactement à la question, mais elle indique bien que la

réponse est probablement non. Quand cette carte est sortie, je vous jure que l'adjoint à l'éducation de la mairie de Lyon, il a perdu 10 kilos de transpiration et en particulier parce qu'il est sorti justement un conseil de d'école qui est à côté du tunnel de la Croix Rousse où il y a un effet de vent qui plaque toute la pollution du tunnel de la Croix Rousse dans l'école et ils ont le choix : soit ils sont à l'intérieur et ils meurent d'un cancer du poumon droit, soit ils sont dans la Cour et ils meurent d'un cancer du poumon gauche. J'exagère beaucoup mais la situation de cette école est catastrophique d'un point de vue qualité de l'air. Il n'y a pas vraiment de bonne solution à part déplacer l'école, ce qui n'est pas possible. Pour l'adjoint à l'éducation, c'était la catastrophe totale. Ce que je veux dire derrière ça, c'est que c'est un acteur politique de la cité, c'est une association militante avec un objectif politique qui prend des données ouvertes, produites par 2 acteurs différents, un associatif mais opérateur d'État et une collectivité territoriale qui fait le choix de les croiser pour donner une information supplémentaire. Ce n'est pas « est ce qu'il y a de la pollution ? » « Où sont les écoles ? » C'est « il y a de la pollution dans l'école de mon enfant » qui fait un choix éditorial très fort sur les couleurs, qui fait un choix de faire de la cartographie parce que ça ramène au sensible, ça ramène à l'intime, c'est chez moi, c'est mon territoire, c'est là où j'habite. Et la cartographie à cette puissance-là. Et ça fait un sujet, ça crée un sujet politique. Parce que avant on savait que c'était pollué Lyon, on savait que dans qu'il y avait des écoles. Puis d'un coup on dit c'est mon enfant qui est là. De la même manière, plus tard, on a fait des mesures de chaleur dans les écoles et tous les parents savaient que leurs enfants avaient chaud au mois de juin et quand il y a eu une canicule et qu'il y a un thermomètre qui mettait 40 dans la salle. C'est devenu un sujet politique parce qu'il y avait un chiffre derrière quoi il y a l'objectivation d'un phénomène environnemental. Ça, ça donne, le pouvoir d'en faire un fait politique et le dernier, un autre exemple d'activisme, dans un autre sens. Clairement, Greenpeace, c'était une attaque frontale aux politiques publiques en disant, il faut que vous changiez tout parce que ça ne va pas. Là c'est la ville à vélo, une association de cyclistes. Evidemment aujourd'hui les associations de cyclistes vis-à-vis de la métropole et de la mairie de Lyon, c'est plutôt copains, copains, ça va ensemble. Mais pour arriver à soutenir une action politique parce qu'une action politique n'est profonde que si elle a un mouvement citoyen qui la, qui la, qui la porte. Parce que pas mal de politiques ont oublié un moment, mais sur cette question-là, la ville a bien compris le lien entre mouvement associatif, mouvement citoyen et mouvement politique. Ce site-là sert à l'association à dire à ses membres ou à ses sympathisants de dire « venez nous dire quand c'est bien, quand ce n'est pas bien. Ça mobilise des gens autour des projets, ça informe qu'il y a des projets, ça permet de mettre la pression sur les maires de petites villes autour qui n'ont pas trop envie que le projet se fasse chez eux ou là il y en a d'autres probablement. C'est un outil de mise sous pression des politiques et de l'information sur les politiques publiques, et ça c'est très intéressant parce que ce n'est plus du tout un outil d'opposition, c'est un outil d'accompagnement critique du pouvoir par les corps intermédiaires, c'est l'autre pendant de la démocratie qui est très fort à cet endroit-là.

00:43:42

YVES GODDE

Merci beaucoup pour cette explication claire de ce qu'est l'Open Data et sa traduction dans la démocratie et le pouvoir citoyen. Moi, ce que j'ai retenu, c'est en effet des aspects techniques de capture de données, de capteur, éventuellement d'annotateurs sous-payés. Dans le flux de données, il y a cette capture. Ensuite il y a le traitement en coulisses, par des informaticiens, et ensuite il y a un traitement par des data analyst, par des data journalistes. Et puis de ces points de vue, en fait, on arrive à des enjeux politiques, que vous nous avez bien montrés, notamment avec cette carte qui croise 2

réalités statistiques qui créent un débat politique. En fait, je reprends l'idée de la carte Greenpeace, des écoles et de la pollution. Tous ces sujets-là ouvrent sur une question politique. Avant de cette intervention nous avons évoqué le point de vue de Y.Varoufakis sur la question de la rente et de l'antagonisme entre les communs et le capitalisme numérique, est ce que vous pouvez nous dire quelque chose là-dessus ?

## 00:45:31 Technoféodalisme et alternatives

HERVE RIVANO

Alors merci parce que ça permet de rebondir, de continuer sur la question des annotateurs et du modèle économique qu'il y a derrière. Je souhaite montrer qu'il y avait plein de dimensions différentes dans cette relation entre données, données ouvertes, politiques, capitalisme. Avec les lois d'ouverture des données, on s'est collectivement contraints à offrir nos données à nos frais, à un système capitaliste qui est celui des GAFAM, dont Google, Amazon et autres. Ce qui est intéressant dans le propos de Y.Varoufakis, le ministre de l'économie de la Grèce pendant la grosse crise grecque, par ailleurs économiste marxiste chevronné qui a sorti un bouquin qui parle de techno-féodalisme. C'est super intéressant. Il y a des podcasts sur son analyse sur France Culture, sur Libération qui sont vraiment très intéressants. Le capitalisme numérique est en train de transformer le capitalisme en quelque chose qui n'est plus du capitalisme qui est du techno-féodalisme. Il y a toute une partie du travail qui est gratuit ou qui relève de l'esclavagisme. On paie des gens une misère, on demande aux citoyens de fournir de la donnée pour que ces entreprises fassent du profit. Et à un moment, le journaliste dans le podcast que j'avais écouté lui disait, c'est quoi l'alternative ? Pour ne pas tomber dans un discours classiquement anticapitaliste, est ce qu'il y a autre chose. Et c'est là où Varoufakis dit : l'alternative c'est les communs. Les pouvoirs publics français ont les moyens d'organiser des communs numériques qui vont à l'encontre de ce que font les GAFAM. La métropole de Lyon faisait ça et après elle a cédé parce que le gouvernement lui a mis une pression d'enfer sur la loi d'ouverture des données de mobilité. Elle résistait jusque-là en disant je ne veux pas donner l'information en temps réel d'où sont les bus et les métros parce que je veux pouvoir garder le contrôle sur mon plan de déplacement urbain que Google ne respecte pas. Ça s'est réglé avec un accord semi à l'amiable où Waze ne fait plus passer les voitures devant les écoles à 08h30. Mais c'est un accord comme ce que serait un arrêt de la guerre entre la Russie et l'Ukraine. Personne n'a gagné et tout le monde a perdu à la fin. Mais l'idée c'est que ToodeGo ça pourrait marcher. On a les moyens de faire un Airbnb fait maison avec un objectif qui n'est plus un objectif capitaliste. Qui n'est plus un objectif pyramidal de concentration de richesse et qui est un objectif politique qui consiste à décider en commun. On a les moyens parce qu'on a le territoire et parce qu'on a les compétences, l'argent. C'est une question politique qui n'émerge pas beaucoup parce qu'elle parle moins aux gens que l'histoire du salaire, de la dépense énergétique et du changement climatique. Ce sera peut-être un sujet puissant intellectuellement, économiquement et politiquement dans 20 ans.. Il y a des alternatives, mais il faut se retrousser les manches à un moment.

## 00:49:30

VALENTYNA DYMYTROVA

Pour compléter dans ce sens-là, sur les communs dans l'histoire du commun, je dirais qu'il n'y a pas que le commun numérique. Le commun ça commence par les biens communs, ce qui concerne les ressources naturelles. Les communs sont difficiles à mettre en place et à maintenir. Il faut se donner des moyens comme ce qui passe par l'éducation par l'acculturation, par la prise de conscience, par

des échanges, des débats, des ateliers d'éco construction, des politiques publiques Il y a différents enjeux autour de ça, mais ça peut être une piste effectivement. Chacun peut apporter une petite pierre à cet édifice par ses choix en tant que consommateur pour choisir une application plutôt qu'une autre, de faire l'usage d'un site web plutôt qu'un autre. Ce sont des choix qui nous paraissent anodins. En effet, il y a la question de praticité, parce qu'effectivement, si j'habite à Marseille et je passe par Lyon, seulement une semaine, pourquoi vais-je télécharger l'application lyonnaise alors que Waze marche partout ? Mais à l'échelle nationale ça peut marcher. Si on propose des solutions numériques à l'échelle du territoire et si on homogénéise des données à l'échelle des différentes métropoles et des différents producteurs. ? Je pense que l'individu à son échelle peut faire des choses. Il peut contribuer à des choses et peut amener des choses pour ce commun numérique et les données.

Mais il faut aussi que ce soit des moyens d'éducation parce qu'on aborde très peu le sujet des données dans nos vies et dans les programmes scolaires. C'est comme la protection de la vie privée, on utilise un service, on sait que ça bouffe nos données et on dit « ce n'est pas grave parce que j'ai besoin de ces services ». C'est ce qu'on appelle le « privacy paradox », paradoxe de la vie privée, parce que moyennant un service, on laisse nos données. On peut aussi contribuer en tant que citoyen.

00:52:12

YVES GODDE

Merci beaucoup pour vos interventions. On va donner la parole au public.

00:52:43

Question du public

Sur ces sites open ou communs, est-ce qu'il y a des informaticiens ou des robots qui vérifient que les données sont pas noyées par des fakes ? Parce que vous parlez de féodalité au Moyen Âge. Il y a eu l'écriture et à un moment donné, il y en avait trop et notamment des fausses, des réattributions et le ménage a commencé quand on a dit « on arrête ». Il y a trop de choses, c'est trop. Alors est-ce qu'il y a un moyen sur les réseaux sociaux....

00:53:43

HERVE RIVANO

Les réseaux sociaux, il faut arrêter car il n'y a juste rien de bien. C'est une vraie très bonne question. A plus petite échelle sur Wikipédia qui est un autre commun numérique, la question de la véracité de l'information se règle par l'arbitrage, par le nombre des contributeurs, par des débats, mais à un moment par des décisions anti démocratiques de verrouillage d'une page parce qu'elle devient trop compliquée. La question est de savoir qui verrouille. C'est un sujet récurrent et complexe dans Wikipédia. J'imagine que les mêmes se posent dans, dans OpenStreetMap, dans les plateformes de données ouvertes. J'imagine que pour l'instant ça va parce qu'il y a un opérateur public qui est le point d'entrée des données. La question actuelle est notamment de la capacité d'intelligence artificielle générative à produire de la donnée hallucinée dans un certain objectif, y compris celui de déstabilisation d'un pouvoir. Il ne faut pas être naïf, ça va arriver. Je sais qu'il y a de la recherche notamment sur la mesure, les capteurs, sur comment on rend les capteurs robustes à des attaques. Dans mes recherches il s'agit de pollution et de comment on peut faire mentir un capteur de pollution

en le chauffant avec un laser. C'est de la science-fiction pour l'instant, mais ça peut devenir un vrai sujet. Comment on résiste à ça ? Je pense que personne n'a la bonne réponse pour l'instant. L'autre sujet c'est est-ce que la donnée est noyée parce qu'il y en a trop. Oui, suffit d'aller sur data Grand Lyon qui est énorme en termes de données. Il faut savoir ce qu'on cherche pour le trouver et avoir le temps. Data.gouv.fr est un labyrinthe parce qu'il y a beaucoup trop de données et ceci étant dit, c'est très bien parce que c'est bien qu'il y ait beaucoup de données. Il y a un vrai sujet sur les médiations qui ne consiste pas simplement à savoir comment je rends accessible une donnée mais comment je donne la capacité à naviguer dans la donnée. Et je sais que tout le monde se pose la question parce que ce n'est pas un objectif de les masquer. C'est un effort à faire que pas grand monde a vraiment bien fait pour l'instant. Pendant un moment ce n'était pas nécessaire parce qu'il y avait 10000 jeux de données. Dans ces conditions, on arrivait à trouver mais quand il commence à y avoir des centaines de milliers de jeux de données... Même la direction du numérique de la métropole, au bout d'un moment ils sont perdus comme moi avec mes mails. Il est question de noyer l'information, pas forcément de manière malicieuse ou sous du faux, mais seulement dans la masse de tout ce qu'il y a. C'est aussi une bonne technique.

00:56:42

VALENTYNA DYMYTROVA

Pour compléter, ce sont des choses très classiques sur la vérification des données avant de les publier. Il y a des acteurs dédiés, il peut y avoir des erreurs, des erreurs involontaires, souvent dans des jeux de données publics publiés et qui sont révélés et signalés, souvent justement par des médiateurs, par les réutilisateurs. C'est ça le travail de data ingénieur ou de développeur ou de data journalist : trouver des erreurs qu'on va signaler. Ça ne concerne pas tous les fichiers de tous les jeux de données. On ne va pas relever toutes les erreurs. Le travail de nettoyage c'est le plus gros travail, le plus coûteux et c'est le travail qui est très important. Par rapport à la vérification, Il y a aussi des aspects liés à l'intelligence artificielle. On peut l'entraîner aussi pour détecter des erreurs. Ça prend aussi du temps et c'est aussi coûteux,

Vous abordez c'est le cycle de la vie, de la donnée. Dans le cycle de la vie de la donnée, il va se poser la question du temps de conservation de la donnée, de sa date limite de conservation et de sa suppression ou conservation. Cette question n'est pas pour l'instant résolue parce que ça pose des questions de retrouver une bonne donnée, mais aussi la question des infrastructures numériques qui doivent être suffisamment robustes pour maintenir un grand nombre de données. On considère la donnée comme un patrimoine informationnel d'un territoire, d'un pays. On peut se dire, « on en aura besoin un jour », comme à la maison quand on met des choses de côté en se disant « on verra un jour » mais au bout d'un moment, on a plus de place. C'est une vraie question qui touche aussi à la question économique, la question écologique parce que les datacenters, les serveurs.... Il y a aussi la question de l'efficacité des services parce que au bout d'un moment la question de la scalabilité va se poser. L'infrastructure ne va plus pouvoir travailler.

01:00:13

HERVE RIVANO

Il y a il y a un site qui s'appelle the Internet archive qui essaie de stocker toute l'historique d'Internet mais qui manque de financement. Une des tartes à la crème des campagnes politiques sur la souveraineté, c'est la pérennité des données. Quand un pouvoir fasciste arrive en place, on voit ce que

ça donne, ça craint. On est en train de perdre tout l'historique de données météorologiques et climatologiques qui étaient maintenus par les États-Unis, où on maintenait le plus gros observatoire de la planète en terme de climato et de météo. La NOAA a fermé son site du jour au lendemain. Ils sont en train de fermer l'agence qui maintient toutes les vulnérabilités informatiques. Les États-Unis depuis toujours maintiennent la base de données des vulnérabilités de tous les systèmes numériques. Le jour où ils éteignent cette information-là est perdue. Ça veut dire que toutes les agences de sécurité ne savent plus quoi faire. On a beaucoup trop fait confiance aux États-Unis sur le stockage des données. On a beaucoup trop fait confiance à des opérateurs privés. On se disait ils sont gros, ils ont une logique commerciale, ce n'est pas grave car c'est le capitalisme et c'est normal. Et en effet il faut bien que les boîtes fassent du pognon, mais à la fin ils rendent un service à toute l'humanité qui est génial. Et puis du jour au lendemain il n'y a plus ce service. On a l'impression que le numérique c'est hyper robuste parce que ce n'est pas réel, ça navigue dans Internet, c'est dans l'air comme ça, mais il y a quelqu'un qui peut tirer la prise et hop il y a plus rien, c'est fini, les données sont perdues ad vitam aeternam. Ça craint vraiment le jour où il y a quelqu'un qui arrive à éteindre Wikipédia. Il y aura beaucoup de journalistes qui n'auront plus de travail. Elon Musk attaque Wikipédia régulièrement sur les réseaux sociaux depuis qu'il est arrivé au gouvernement de Trump. Il ne faut pas se dire que c'est un fantasme dystopique, ou alors on est déjà entré dans une dystopie. Comment l'Europe se dote de moyens d'être robuste ou résiliente à ce genre de choses, sachant que ça peut nous arriver aussi d'être du mauvais côté de la barrière. Il faut être robuste à ça en tant que citoyen aussi.

01:02:09

YVES GODDE

Parfait merci beaucoup l'Open Data et surtout le pouvoir citoyen semblent intéresser tout le monde et on a largement débordé. C'est pourquoi, on va mettre un terme à cette conférence-débat. Je vous remercie pour votre attention et je vous souhaite une bonne journée et à la prochaine fois.