

La science politique face aux enjeux du « big data » et de la protection des données personnelles sur internet

Julien Boyadjian

Sciences Po Lille – CERAPS

12/01/2017

Plan de la communication

1. L'intérêt scientifique des données numériques pour la science politique
2. Les enjeux éthiques soulevés par l'analyse des données numériques
3. Une tentative de conciliation des enjeux scientifiques et éthiques

1. L'intérêt scientifique des données numériques

- Données numériques : verbatims d'internautes (tweets, posts, billets de blog, etc.) + données sur les auteurs des messages (sexe, âge, nombre d'amis, de followers, etc.)
- Données numériques : des indices précieux concernant les représentations, pratiques et opinions des internautes
- Données numériques : un matériau d'enquête pour les SHS

1. L'intérêt scientifique des données numériques



KI @ContactKi · 26 min

Aucun respect sur le bancs des Ministres !!! Ça se marre bien entre eux : **Valls**, **Macron**, **Cazeneuve**, **Pellerin**... #AN image déplorable !!



JulBzz @JulBzz · 1 h

Mr Jacob, le privment **à la source** est une bonne mesure pr les classes moyennes qui n'auront plus **à payer** les variations **à retardement** #Impot



1. L'intérêt scientifique des données numériques

Trois « avantages comparatifs » des données numériques pour la science politique :

1. Des données produites « hors laboratoire »
2. Des données justiciables d'une analyse quantitative et qualitative
3. Des données traçables et générées en temps réel

1. L'intérêt scientifique des données numériques

Trois « avantages comparatifs » des données numériques pour la science politique :

1. Des données produites « hors laboratoire »

Des données générées indépendamment de la situation d'enquête

Leur analyse permet d'éviter les biais inhérents associés aux enquêtes déclaratives : imposition de problématique, défaillance de la mémoire, mensonge, auto-censure, etc.

1. L'intérêt scientifique des données numériques

Trois « avantages comparatifs » des données numériques pour la science politique :

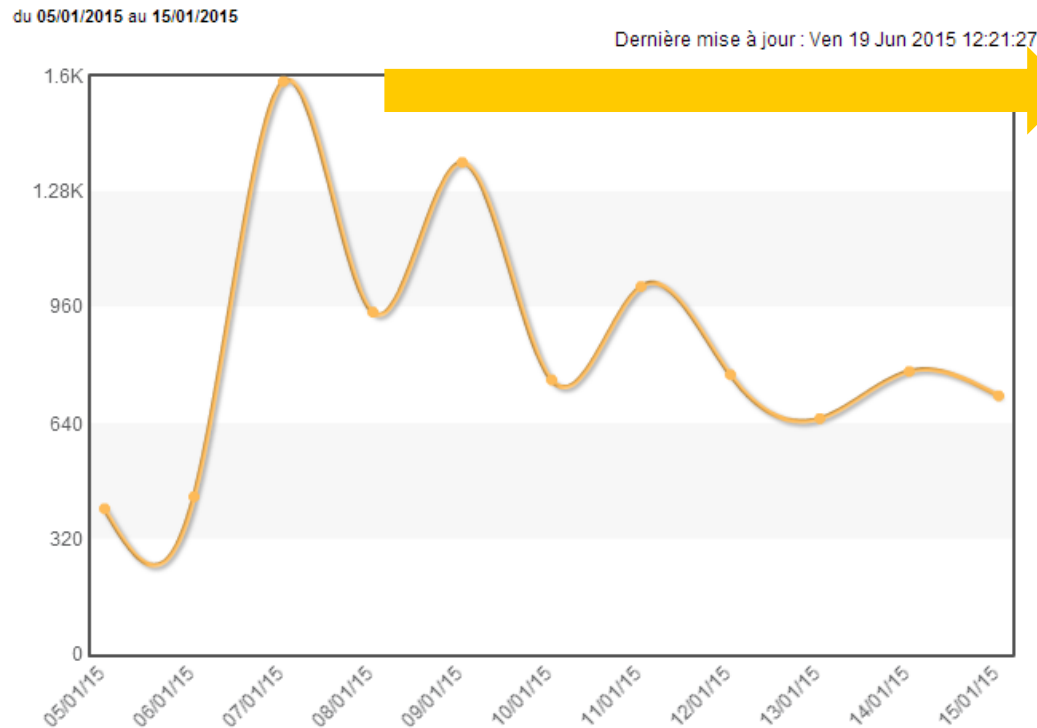
2. Des données justiciables d'une analyse quantitative et qualitative


Des verbatims...

...aisément quantifiables (logiciels de tracking)

=> Outils de volumétrie, lexicométrie, etc.


1. L'intérêt scientifique des données numériques



 **Marc-Antoine Daneau**
@DaneauMa Suivre

Et en passant, ça serait une bonne idée que l'on sorte de l'Âge du bronze...
[#Religion](#) [#CharlieHebdo](#) [#JeSuisCharlie](#)

15:47 - 7 janv. 2015

 **Laurie Gandon**
@Laurie_Gandon Abonné

Horreur et dévastation ! Mes pensées et mon soutien aux victimes et familles
[#JeSuisCharlie](#)

15:22 - 7 janv. 2015

1. L'intérêt scientifique des données numériques

Trois « avantages comparatifs » des données numériques pour la science politique :

3. Des données traçables et générées en temps réel

Permet d'étudier la réaction des internautes sur un événement au moment-même où s'est produit l'événement (et non *a posteriori* comme avec les méthodes déclaratives)

Données horodatées : permet d'étudier les dynamiques de production et circulation des messages

2. Les enjeux éthiques autour des données numériques

Un régime de publicité indéterminé

Données accessibles = données publiques ?

Les internautes maîtrisent-ils les paramètres de confidentialité de leur compte ?

2. Les enjeux éthiques autour des données numériques

L'obtention du consentement des enquêtés

Le chercheur peut-il agir incognito ?

Comment obtenir le consentement explicite de millions d'internautes ?

2. Les enjeux éthiques autour des données numériques

Dilemme posé par le « big data » :

- Opter pour une analyse de type « big data » (et profiter de ses avantages méthodologiques) mais se passer du consentement des enquêtés
- Rechercher le consentement des enquêtés mais travailler alors sur des corpus quantitativement et qualitativement plus limités

3. Une tentative de conciliation des enjeux scientifiques et éthiques

Construire un dispositif d'enquête panélisté d'un échantillon représentatif d'internautes

Construire un échantillon représentatif et raisonné de comptes d'internautes

Leur présenter l'enquête et rechercher leur accord

Ne plus interagir avec eux par la suite et se contenter de collecter leurs données à l'aide d'un logiciel de tracking

3. Une tentative de conciliation des enjeux scientifiques et éthiques

Construire un dispositif d'enquête panélisté d'un échantillon représentatif d'internautes

Limites du dispositif :

Réintroduit une logique de représentativité (et les biais qui lui sont associés)

Un dispositif particulièrement coûteux