



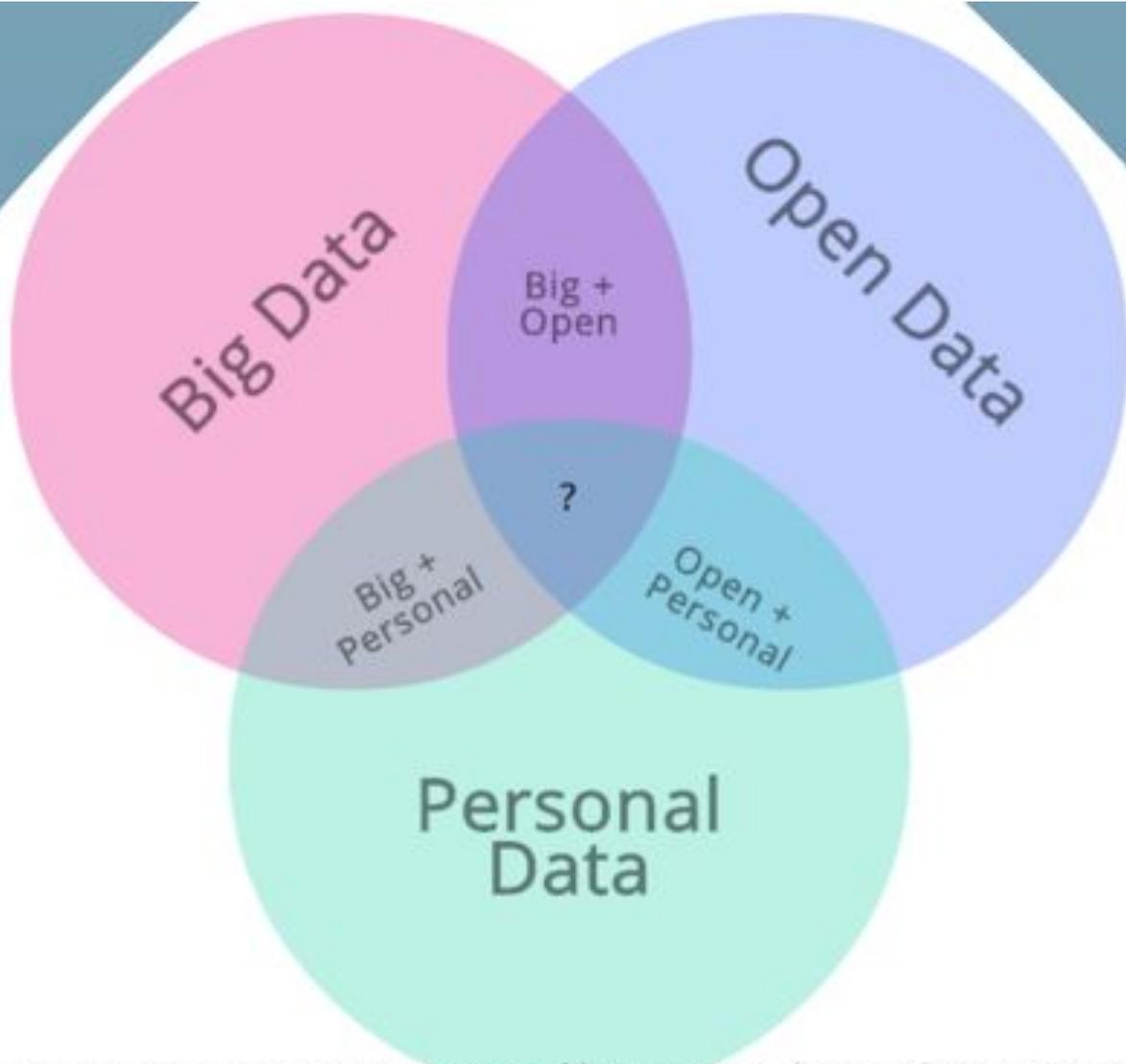
© Kalenderson (CC BY-NC)

Merci !
Geoffrey Delcroix
Prospective & Innovation

gdelcroix@cnil.fr

 [geoffdelc](https://twitter.com/geoffdelc)





Source : Open Data Institute <https://github.com/theodi/data-definitions>



- Alex Sandy Pentland, computational social scientist @MIT
- "This data is a new asset, you want it to be liquid and to be used."
- Conscience pour le régulateur du climat économique
- La concurrence peut être vertueuse pour la protection des données



Les data, pétrole du XXIème siècle ?

Le rêve "big data" :
 "Finding the needle in the haystack"
 (Big data privacy review, John Podesta)



OF THE 15 OR COPIES ON YOUR DESK
 EQUALS ONE CIGARETTE
A ZETTABYTE
 THE SAME VOLUME AS
THE GREAT WALL OF CHINA





- Alex Sandy Pentland, computational social scientist @MIT
 "This data is a new asset, you want it to be liquid and to be used."
- Conscience pour le régulateur du climat économique
- La concurrence peut être vertueuse pour la protection des données



Les data, pétrole du XXIème siècle ?

Le rêve "big data" :
 "Finding the needle in the haystack"
 (Big data privacy review, John Podesta)





- Alex Sandy Pentland, computational social scientist @MIT
"This data is a new asset, you want it to be liquid and to be used."
- Conscience pour le régulateur du climat économique
- La concurrence peut être vertueuse pour la protection des données



Les data, pétro

Les data, petroie du XXIème siècle



Le rêve "big data" :
"Finding the needle in the haystack"
(Big data privacy review, John Podesta)



But how much is that really?

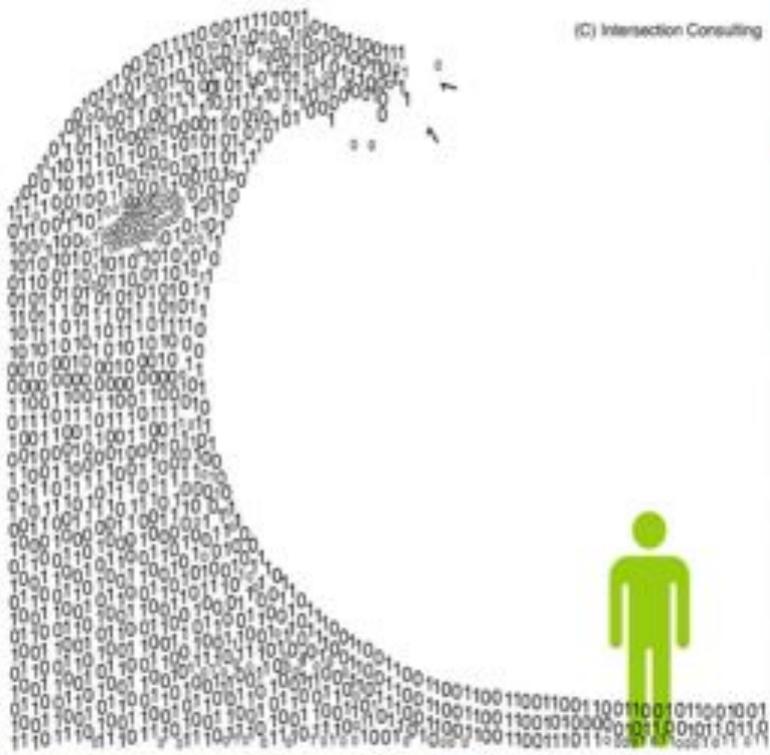
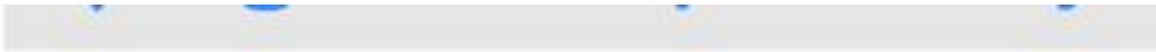
IF THE 11 OZ COFFEE ON YOUR DESK EQUALS ONE GIGABYTE

A ZETTABYTE

would have

THE SAME VOLUME AS THE GREAT WALL OF CHINA

Source: Cisco Visual Networking Index, 4th Year Update

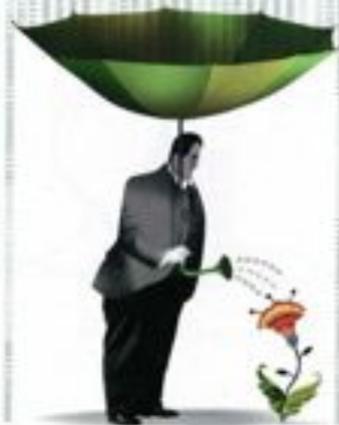


The Economist

- Obama the warrior
- Misgoverning Argentina
- The economic shift from West to East
- Genetically modified crops blossom
- The right to eat cats and dogs

The data deluge

AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT





But how much is that really?

IF THE 11 OZ COFFEE ON YOUR DESK EQUALS ONE GIGABYTE

A ZETTABYTE

would have

THE SAME VOLUME AS THE GREAT WALL OF CHINA

Source : Cisco Visual Networking Index, via Henri Verdier



big data VS. privacy ?

"on a le big data mais pas le mode d'emploi"

danah boyd et Kate Crawford "Six Provocations for Big Data" / sept. 2011
 "Big Data tempts some researchers to believe that they can see everything at a 30,000-foot view. It is the kind of data that encourages the practice of apophenic seeing patterns where none actually exist, simply because massive quantities of data can offer connections that radiate in all directions."

1. Automating Research Changes the Definition of Knowledge.
Do numbers speak for themselves? The answer, we think, is a resounding 'no'.
2. Claims to Objectivity and Accuracy are Misleading
Big Data offers the humanistic disciplines a new way to claim the status of quantitative science and objective method.
3. Bigger Data are Not Always Better Data
4. Not All Data Are Equivalent
5. Just Because it is Accessible Doesn't Make it Ethical
6. Limited Access to Big Data Creates New Digital Divides



big data VS. privacy ?

"on a le big data mais pas le mode d'emploi"

danah boyd et Kate Crawford "Six Provocations for Big Data" / sept. 2011

"Big Data tempts some researchers to believe that they can see everything at a 30,000-foot view. It is the kind of data that encourages the practice of apophenia: seeing patterns where none actually exist, simply because massive quantities of data can offer connections that radiate in all directions."

1. Automating Research Changes the Definition of Knowledge.

Do numbers speak for themselves? The answer, we think, is a resounding 'no'.

2. Claims to Objectivity and Accuracy are Misleading

Big Data offers the humanistic disciplines a new way to claim the status of quantitative science and objective method.

3. Bigger Data are Not Always Better Data

4. Not All Data Are Equivalent

5. Just Because it is Accessible Doesn't Make it Ethical

6. Limited Access to Big Data Creates New Digital Divides

BIG DATA LAW #1

"Droit d'accès" ?

The faster you analyze your data, the greater its predictive value.

Companies are moving away from batch processing to real-time to gain competitive advantage.

BIG DATA LAW #3

"Croisements" ?

Use more diverse data, not just more data. More diverse data leads to greater insights. Combining multiple data sources can lead to the most interesting insights of all.

BIG DATA LAW #4

"Traitement" ?
"Finalité" ?
"Conservation" ?

Data has value far beyond what you originally anticipate. Don't throw it away.

BIG DATA LAW #5

Plan for exponential growth. The number of photos, emails, and IMs while large, is limited by the number of people. Networked "sensor" data from mobile phones, GPS, and other devices is much larger.

BIG DATA LAW #8

Big Data is transforming business the same way IT did. Those that fail to leverage the numerous internal and external data sources available will be leapfrogged by new entrants.

"Collecte" ?

*Source : Big Data Trends
Forbes, David Feinleib, 2012*

Big Data is transforming business the same way IT did. Those that fail to leverage the numerous internal and external data sources available will be leapfrogged by new entrants.

Source : Big Data Trends
Forbes, David Forstall, 2012

big data VS. privacy

"on a le big data mais pas le mode d'emploi"

danah boyd et Kate Crawford "Six Provocations for Big Data" / sept. 2011

"Big Data tempts some researchers to believe that they can see everything at a 30,000-foot view. It is the kind of data that encourages the practice of apophenia: seeing patterns where none actually exist, simply because massive quantities of data can offer connections that radiate in all directions."

1. Automating Research Changes the Definition of Knowledge.

Do numbers speak for themselves? The answer, we think, is a resounding 'no'.

2. Claims to Objectivity and Accuracy are Misleading

Big Data offers the humanistic disciplines a new way to claim the status of quantitative science and objective method.

3. Bigger Data are Not Always Better Data

4. Not All Data Are Equivalent

5. Just Because it is Accessible Doesn't Make it Ethical

6. Limited Access to Big Data Creates New Digital Divides

big data mais pas le mode d'emp

danah boyd et Kate Crawford "Six Provocations for Big Data" / sept. 2011

"Big Data tempts some researchers to believe that they can see everything at a 30,000-foot view. It is the kind of data that encourages the practice of apophenia: seeing patterns where none actually exist, simply because massive quantities of data can offer connections that radiate in all directions."

1. Automating Research Changes the Definition of Knowledge.

Do numbers speak for themselves? The answer, we think, is a resounding 'no'.

2. Claims to Objectivity and Accuracy are Misleading

Big Data offers the humanistic disciplines a new way to claim the status of quantitative science and objective method.

3. Bigger Data are Not Always Better Data

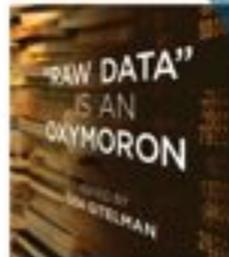
4. Not All Data Are Equivalent

5. Just Because it is Accessible Doesn't Make it Ethical

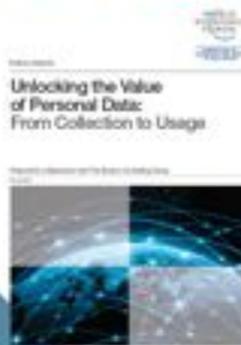
6. Limited Access to Big Data Creates New Digital Divides

Big data is like teenage sex:
everyone talks about it, nobody
really knows how to do it, everyone
thinks everyone else is doing it, so
everyone claims they are doing it ...

- "Le chant du cygne d'une informatique productiviste, centralisatrice, centrée sur les grandes organisations." Daniel Kaplan (FING)
- "Les Big Data se fondent entièrement sur la statistique : dégager des schémas et des tendances à partir de grandes collections de données. (...) Le défi est celui de la logistique de l'information : comment transmettre exactement la bonne information à, et depuis, les bonnes personnes, au bon format, au bon moment. Une affaire de 'Toutes Petites Données' (Very Small Data)." Alan Mitchell, ctrl-shift
- "Big data" is dead. What's next? John De Goes :
Smart Data / Data Science / NewSQL / Predictive Analytics



et si le big data n'existait pas?



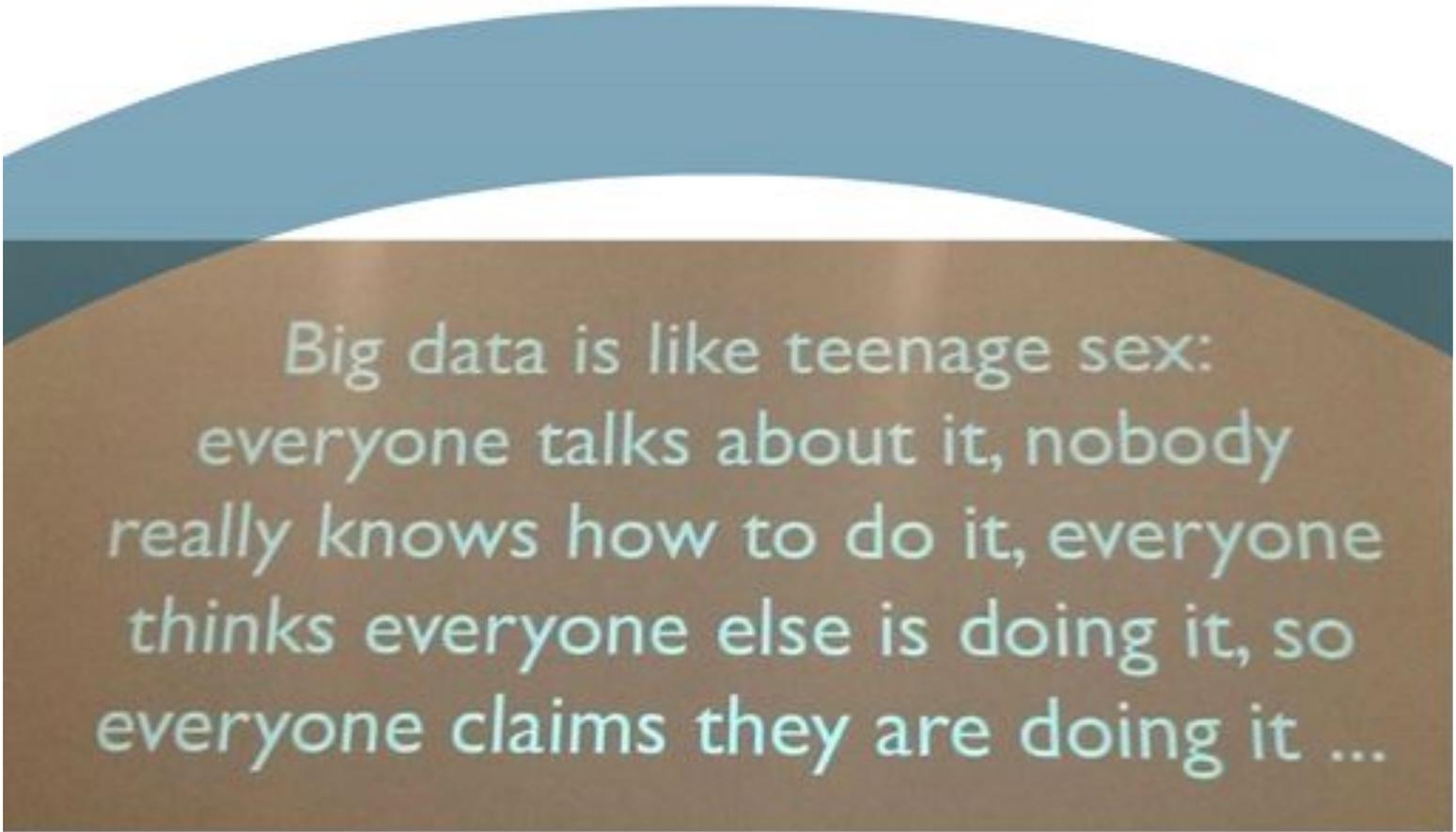
"Big Data and a Renewed Debate Over Privacy" - Steve Lohr

The report, "Unlocking the Value of Personal Data: From Collection to Usage," recommends a major shift in the focus of regulation toward restricting the use of data. Curbs on the use of personal data, combined with new technological systems, can give individuals control of their own information, according to the report, while permitting important data users to flow relatively freely.

"There's no bad data, only bad uses of data," says Craig Mundie, a senior adviser at Microsoft, who worked on the position paper.

The forum report suggests a future in which all collected data would be tagged with software code that included an individual's preferences for how his or her data is used. All uses of data would have to be registered, and there would be penalties for violators. For example, one violation might be a smartphone application that stored more data than is necessary for a registered service like a smartphone game or a restaurant finder.

"We're on the cusp of a golden age of medical science and care delivery," Mr. Huberman (whose permanent) says. "But a privacy backlash could cripple progress."



Big data is like teenage sex:
everyone talks about it, nobody
really knows how to do it, everyone
thinks everyone else is doing it, so
everyone claims they are doing it ...

e informatique productiviste, centralisatrice, centrée sur les
Daniel Kaplan (FING)





*really knows how to do it, everyone
thinks everyone else is doing it, so
everyone claims they are doing it ...*

- "Le chant du cygne d'une informatique productiviste, centralisatrice, centrée sur les grandes organisations." Daniel Kaplan (FING)
- "Les Big Data se fondent entièrement sur la statistique : dégager des schémas et des tendances à partir de grandes collections de données. (...) Le défi est celui de la logistique de l'information : comment transmettre exactement la bonne information à, et depuis, les bonnes personnes, au bon format, au bon moment. Une affaire de 'Toutes Petites Données' (Very Small Data)." Alan Mitchell, ctrl-shift
- 'Big data' is dead. What's next?" John De Goes :

Smart Data / Data Science / NewSQL / Predictive Analytics

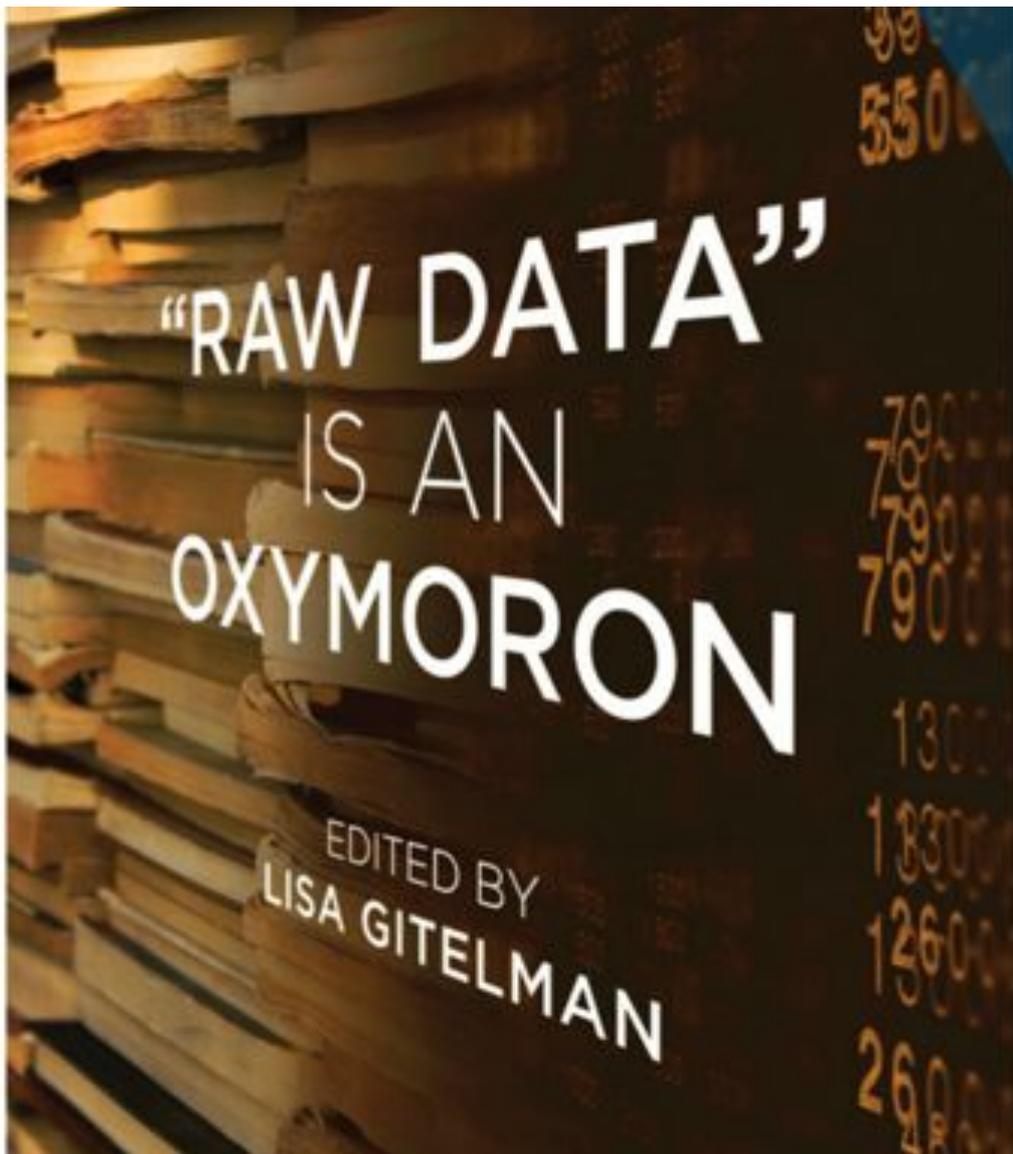
et si le big data n'exist

WORLD
ECONOMIC
FORUM

"Big Data and a Renewed Debate Over Privacy" - Stev

The report "Rethinking the Value of Personal Data: From

r les
des
ation
de



Si le big data n'existait



Industry Agenda

Unlocking the Value of Personal Data: From Collection to Usage

Prepared in collaboration with The Boston Consulting Group



"Big Data and a Renewed Debate Over Privacy" - Steve Lohr

The report, "Unlocking the Value of Personal Data: From Collection to Usage," recommends a major shift in the focus of regulation toward restricting the use of data. Curbs on the use of personal data, combined with new technological options, can give individuals control of their own information, according to the report, while permitting important data assets to flow relatively freely.

"There's no bad data, only bad uses of data," says Craig Mundie, a senior adviser at Microsoft, who worked on the position paper.

The forum report suggests a future in which all collected data would be tagged with software code that included an individual's preferences for how his or her data is used. All uses of data would have to be registered, and there would be penalties for violators. For example, one violation might be a smartphone application that stored more data than is necessary for a registered service like a smartphone game or a restaurant finder. "We're on the cusp of a golden age of medical science and care delivery," Mr. Halvorson (kaiser permanente) says. "But a privacy backlash could cripple progress."

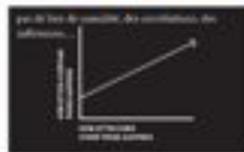
Tout est données personnelles / Rien n'est données personnelles

Fin de la dichotomie?
Continuum en fonction de l'énergie à mettre dans l'identification
 FTC (Protecting Consumer Privacy in an Era of Rapid Change) prohibits strategies of re-identification + obligations légales anonymisation + obligations contractuelles en cas de non pas re-identifier. According to the FTC: "As long as (1) a given data set is not reasonably identifiable, (2) the company publicly commits not to re-identify it, and (3) the company requires any downstream users of the data to keep it in de-identified form, that data will fall outside the scope of the framework."
 → voir GDPR sur les techniques d'anonymisation. Opinion 04/2004 on "Anonymisation Techniques over the web" distingue 3 critères:
 1) is it still possible to single out an individual, 2) is it still possible to link records relating to an individual, and 3) are information be inferred concerning an individual?"



On fait quoi ?

Euphémisation des discriminations



Risque de refaire de l'individu un "sujet" de données

The benefits and consequences of perfect personalization
 Furthermore, "perfect personalization" also brings some benefits and not so subtle forms of discrimination in pricing, services, and opportunities. For example, one study found web searches including their identifying factors (e.g., "browser") also return "bias-by-credibility" also with the word "score" for those that search with other identifying factors (e.g., "Google")."

Après tout, les gains d'efficacité et d'innovation sont élevés. Mais il y a un revers à cette médaille. Les données personnelles sont une mine d'or pour les entreprises, mais elles sont aussi une mine d'or pour les discriminateurs.

Il y a un autre aspect à prendre en compte : la personnalisation des services peut aussi entraîner une discrimination. Par exemple, les entreprises peuvent offrir des services personnalisés à certains clients et pas à d'autres. Cela peut entraîner une discrimination basée sur le genre, l'âge, le statut marital, etc.

Enfin, la personnalisation des services peut aussi entraîner une discrimination basée sur le statut social. Les entreprises peuvent offrir des services personnalisés à certains clients et pas à d'autres. Cela peut entraîner une discrimination basée sur le statut social.



"Capacitation" + PETs

L'un des outils de gestion des données, permettant d'automatiser les vérifications. PETs, acronym anglais.

Le gouvernement algérien a un rôle de gouvernement dans le développement de données utiles. Il y a un aspect de responsabilité de la part des entreprises qui doivent être conscientes de leur rôle dans la protection des données. Il y a aussi un aspect de responsabilité de la part des individus qui doivent être conscients de leur rôle dans la protection des données.

Le gouvernement algérien a un rôle de gouvernement dans le développement de données utiles. Il y a un aspect de responsabilité de la part des entreprises qui doivent être conscientes de leur rôle dans la protection des données. Il y a aussi un aspect de responsabilité de la part des individus qui doivent être conscients de leur rôle dans la protection des données.

Autoblog - Blogger

Tout est données personnelles / Rien n'est données personnelles

Fin de la dichotomie?

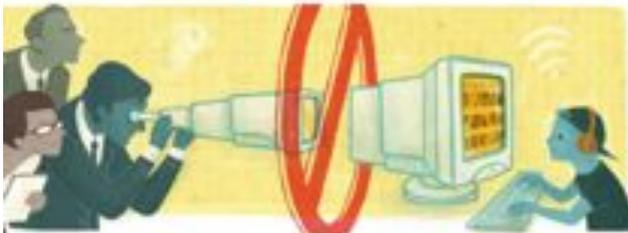
Continuum en fonction de l'énergie à mettre dans l'identification

FTC "Protecting Consumer Privacy in an Era of Rapid Change" probabilité statistique de ré-identification + obligations légales sanctionnables + obligations contractuelles en aval de ne pas ré-identifier. According to the FTC: "As long as (1) a given data set is not reasonably identifiable, (2) the company publicly commits not to re-identify it, and (3) the company requires any downstream users of the data to keep it in de-identified form, that data will fall outside the scope of the framework. + avis G29 sur les techniques d'anonymisation (Opinion 05/2014 on "Anonymisation Techniques onto the web") distingue 3 criteres :
1/ is it still possible to single out an individual, 2/ is it still possible to link records relating to an individual, and 3/ can information be inferred concerning an individual?"

#de-identification
#re-identification
#anonymisation



Matrice ou outils d'évaluation du risque de réidentification ?



On fait quoi

Tout est données person Rien n'est données persor

Fin de la dichotomie?

Continuum en fonction de l'énergie à mettre dans l'identification

FTC "Protecting Consumer Privacy in an Era of Rapid Change" probabilité statistique de ré-identification + obligations légales sanctionnables + obligations contractuelles en aval de ne pas ré-identifier. According to the FTC: "As long as

(1) a given data set is not reasonably identifiable, (2) the company publicly commits not to re-identify it, and (3) the company requires any downstream users of the data to keep it in de-identified form, that data will fall outside the scope of the framework.

+ avis G29 sur les techniques d'anonymisation (Opinion 05/2014 on "Anonymisation Techniques onto the web") distingue 3 criteres :

1/ is it still possible to single out an individual, 2/ is it still possible to link records relating to an individual, and 3/ can information be inferred concerning an individual?



Ma
du



On fa

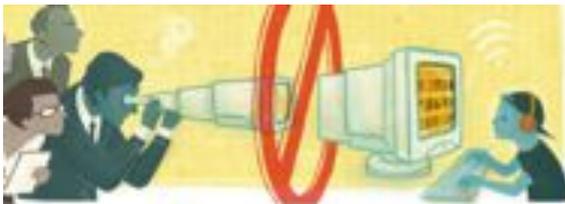


que
s en
à to
on
ial?



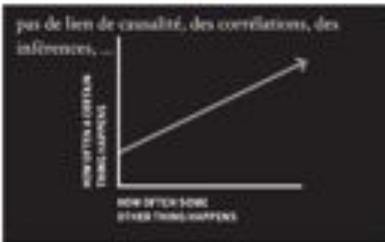
#de-identification
#re-identification
#anonymisation

Matrice ou outils d'évaluation
du risque de réidentification ?



On fait quoi ?

Euphémisation des discriminations



Risque de refaire de l'individu un "sujet" de données

The benefits and consequences of perfect personalization

Unfortunately, "perfect personalization" also leaves room for subtle and not-so-subtle forms of discrimination in pricing, services, and opportunities. For example, one study found web searches involving black-identifying names (e.g., "Jermaine") were more likely to display ads with the word "arrest" in them than searches with white-identifying names (e.g., "Geoffrey").

Loi n° 78-17 de 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés Article 10

L'informatique doit être au service de chaque citoyen. Son développement doit s'insérer dans le cadre de la coopération internationale. Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou politiques.

Article 10

Aucune décision de justice impliquant une application ou le comportement d'une personne ne peut être prise fondamment sur le traitement automatisé de données à caractère personnel destiné à traiter certains aspects de sa personnalité.

Aucune autre décision produisant des effets juridiques à l'égard d'une personne ne peut être prise sur le seul fondement d'un traitement automatisé de données destiné à évaluer le profil de l'intéressé ou à traiter certains aspects de sa personnalité.

Ne sont pas regardées comme prises sur le seul fondement d'un traitement automatisé les décisions prises dans le cadre de la conclusion ou de l'exécution d'un contrat et pour lesquelles la personne concernée a été mise à même de présenter ses observations, et celles soulignant les données de la personne concernée.



"Capacitation" + PETs

Créer des outils de gestion des données, permettant d'instrumenter les individus : PDS, agents logiciels,

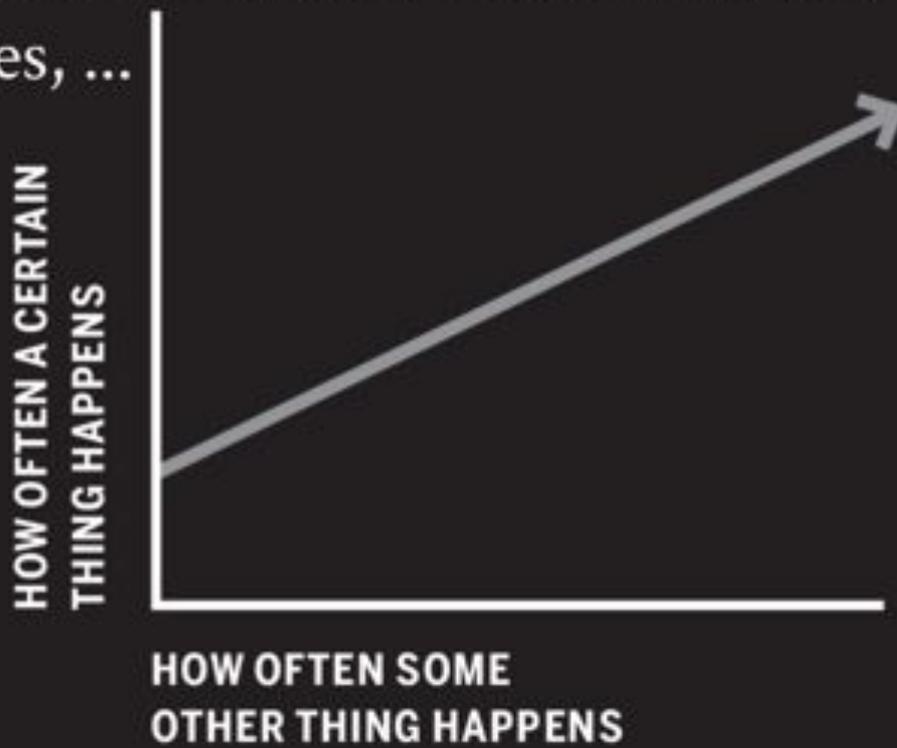
"Le gouvernement algorithmique est un mode de gouvernement nourri essentiellement de données brutes, signaux infra-personnels et a-signifiants mais quantifiables, opérant par configuration anticipative des possibles plutôt que par réglementation des conduites, et ne s'adressant aux individus que par voie d'alertes provoquant des réflexes plutôt qu'en s'appuyant sur leurs capacités d'entendement et de volonté. [...]"

un gouvernement algorithmique qui façonne l'avenir, qui affecte sur le mode de l'alerte et du réflexe, mais n'éprouve ni n'est éprouvé par aucun sujet et de quoi faire thème - ne fût-ce que dans la mesure où il ne se laisse plus provoquer par la liberté humaine."

Antoinette Rouvroy



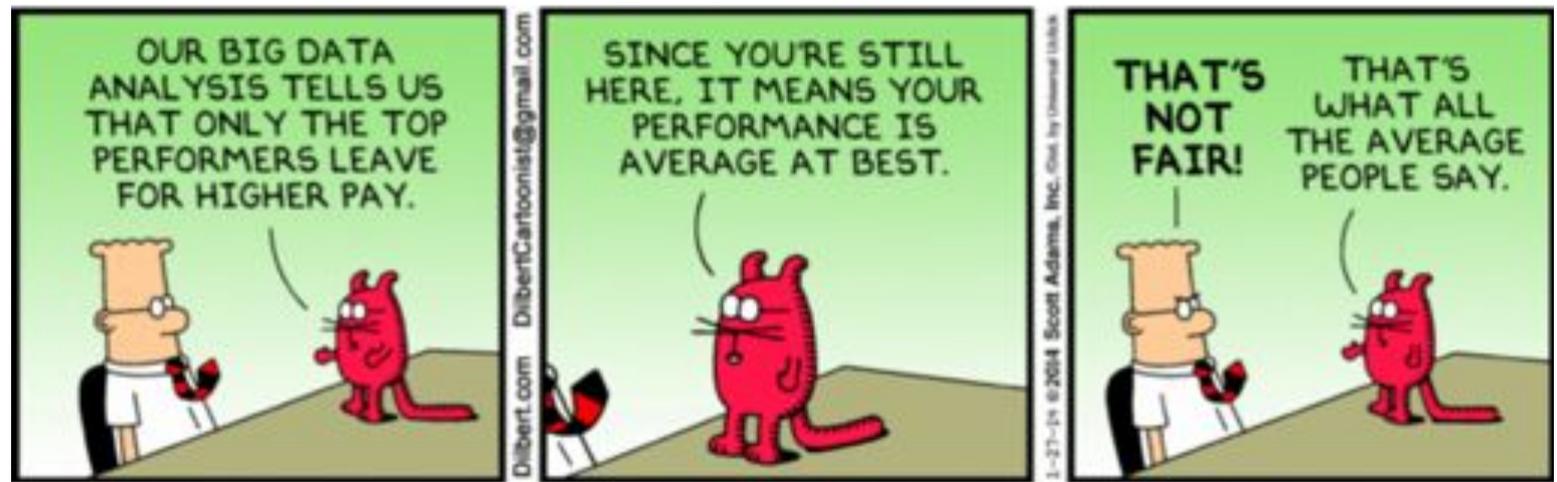
pas de lien de causalité, des corrélations, des inférences, ...



THIN

HOW OFTEN SOME OTHER THING HAPPENS

searches involving black-identif more like-ly to display ads with searches with white-identifying



"Capacitation" + P

Créer des outils de gestion des données

Risque de refaire de l'ind

The benefits and consequences of perfect personalization

Unfortunately, "perfect personalization" also leaves room for subtle and not-so-subtle forms of discrimination in pricing, services, and opportunities. For example, one study found web searches involving black-identifying names (e.g., "Jermaine") were more likely to display ads with the word "arrest" in them than searches with white-identifying names (e.g., "Geoffrey").

Loi
Artic
L'ind
cadr
droit

Artic
Auct
peut
éval
Auct
sur l
l'inté
Ne s
décis
pers
dem



"Le gouvernement algor
nourri essen

L'individu un "sujet" de données

Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés

Article 1er

L'informatique doit être au service de chaque citoyen. Son développement doit s'opérer dans le cadre de la coopération internationale. Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques.

Article 10

Aucune décision de justice impliquant une appréciation sur le comportement d'une personne ne peut avoir pour fondement un traitement automatisé de données à caractère personnel destiné à évaluer certains aspects de sa personnalité.

Aucune autre décision produisant des effets juridiques à l'égard d'une personne ne peut être prise sur le seul fondement d'un traitement automatisé de données destiné à définir le profil de l'intéressé ou à évaluer certains aspects de sa personnalité.

Ne sont pas regardées comme prises sur le seul fondement d'un traitement automatisé les décisions prises dans le cadre de la conclusion ou de l'exécution d'un contrat et pour lesquelles la personne concernée a été mise à même de présenter ses observations, ni celles satisfaisant les demandes de la personne concernée.

**Le management algorithmique est un mode de gouvernement
nourri essentiellement de données brutes,
personnelles et a-signifiants mais quantifiables, opérant
sur configuration anticipative des possibles
à l'instar de ce que par réglementation des conduites**

personne concernée a été mise à même de présenter ses observations et demandes de la personne concernée.

"Le gouvernement algorithmique est un mode de gouvernement nourri essentiellement de données brutes, signaux infra-personnels et a-signifiants mais quantifiables, opérant par configuration anticipative des possibles plutôt que par réglementation des conduites, et ne s'adressant aux individus que par voie d'alertes provoquant des réflexes plutôt qu'en s'appuyant sur leurs capacités d'entendement et de volonté. (...)

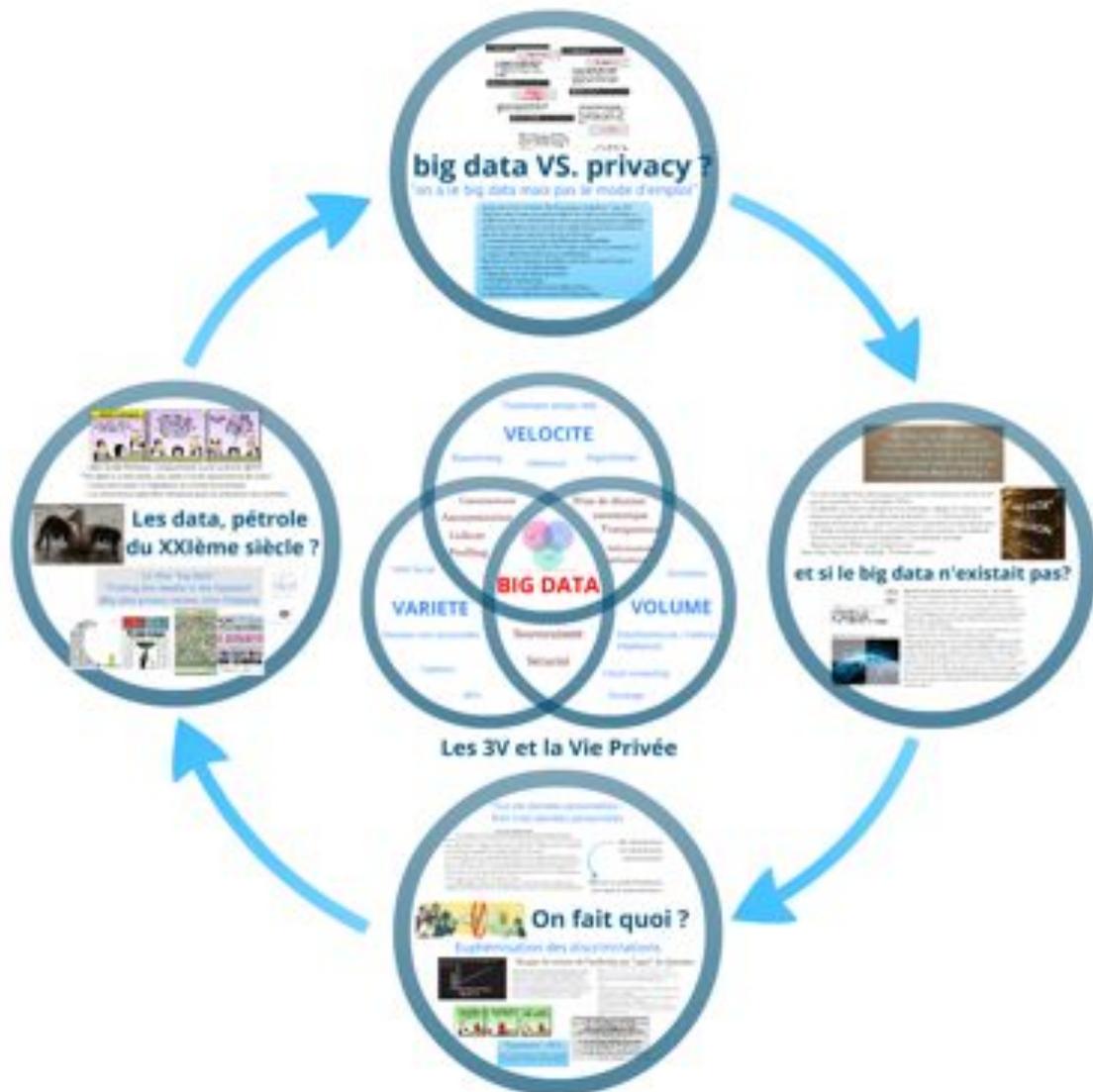
Un gouvernement algorithmique qui façonne l'advenir, qui affecte sur le mode de l'alerte et du réflexe, mais n'éprouve ni n'est éprouvé par aucun sujet a de quoi faire frémir - ne fût-ce que dans la mesure où il ne se laisse plus provoquer par la liberté humaine."

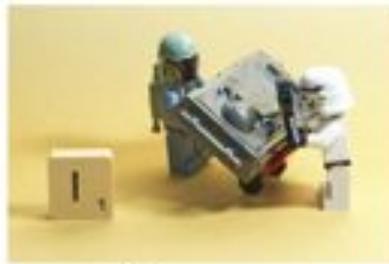
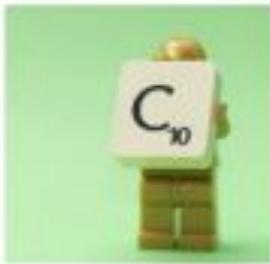
Antoinette Rouvroy



"Capacitation" + PETs

Créer des outils de gestion des données, permettant d'instrumenter les individus : PDS, agents logiciels,





© Kalexanderson (CC-by-nc-sa)

Merci !

Geoffrey Delcroix

Prospective & Innovation

gdelcroix@cnil.fr

 [geoffdelc](https://twitter.com/geoffdelc)