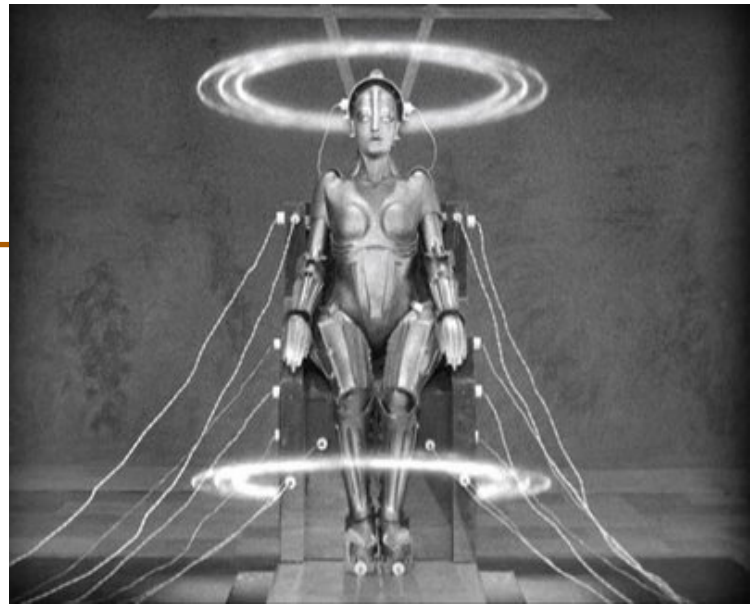
Observatoire
de la vie littéraire



Des *Big Data* aux *Big Brothers*

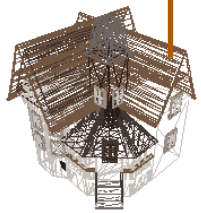
Jean-Gabriel Ganascia

Equipe ACASA – LIP6 – Université Pierre and Marie Curie

Labex OBVIL – PRES "Sorbonne Université"

4, place Jussieu, 75252 Paris Cedex 05, FRANCE

Jean-Gabriel.Ganascia@lip6.fr

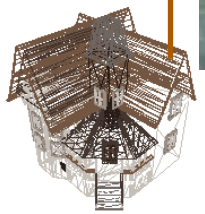


Synoptique

- *Big Data*
  - *Déluge informationnel*
  - *Les 3 V*
  - *La logique des Big Data*
- *« Big Brother »*
  - *Données personnelles*
  - *Protection?*
- *Qui sont les « Big Brothers »?*
  - *Utilisation des Big Data*
  - *Nouvelle économie*
- *Conséquences politiques*

Big Data – masses de données
Déluge informationnel

- Livres
- Gazouillis (tweets, ...)
- Traces
  - Caméras de surveillances
  - Empreintes digitales
  - Requêtes google
- Archives individuelles
- Science
  - Génome, séquence ADN
  - Astronomie
- ...

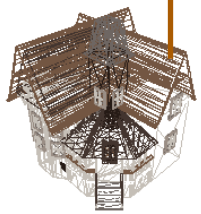


UPMC SORBONNE UNIVERSITÉS

Unité de mémoire: le nœud de mouchoir

1 nœud de mouchoir = 1 « bit »

1 caractère typographique = 8 mouchoirs

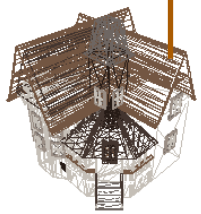


UPMC SORBONNE UNIVERSITÉS

Accroissement des quantités
Masses de données - « Big Data »

Quantité d'information

- *1 livre, 1 million de caractères*
  → 1000 Ko = 1Mo (10^6 octets) = 8 millions de mouchoirs
- *Catalogue des livres et imprimés de la BNF:*
  *14 millions d'ouvrages = **14To***
  1 Go = 10^3Mo = 10^9o (1 milliard)
  1 To = 10^{12}o (1000 milliards)

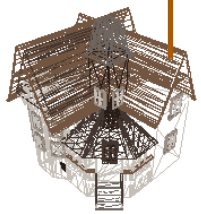


UPMC SORBONNE UNIVERSITÉS

Production des données en nombre de BNF

- **Twitter** produit 7 To/jour c'est-à-dire ½ BNF!
- **Facebook**: 500 To/jour
- Radio télescope *Murchison Widefield Array* (Australie):
  - **données brutes** 7000 To/ minutes, c'est-à-dire 500 BNF/minute
  - **données analysées**: 50To par jour (4 BNF/jour)
- ***Volume total du web en 2015***:
  7 Zeta-octets = 7 10^{21}o

½ milliard de BNF

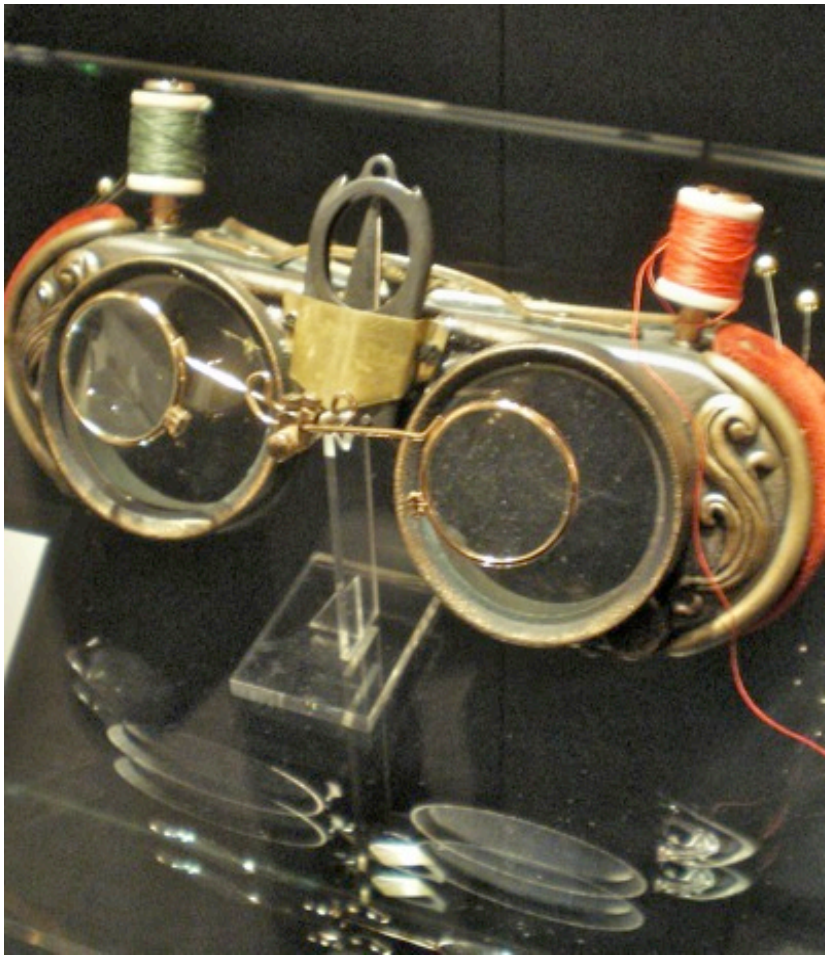


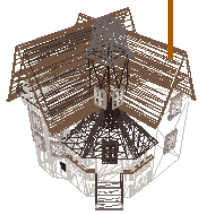
UPMC SORBONNE UNIVERSITÉS

Big Data

Les 3 V

- **Volume**: To – Po (de 10^{12}o à 10^{15}o)
- **Vélocité**: données changeantes
- **Variabilité**: non organisées, toutes les langues, mélange image, son, …



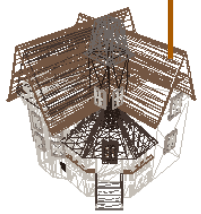


UPMC SORBONNE UNIVERSITÉS

La logique des masses de données

Trois caractéristiques:

- **Les 3V**
  - Volume
  - Variété
  - Vélocité
- **Acquisition continue**
  - « Crowdsourcing »
  - requêtes, ...
- **Absence d'échantillonnage**
  - Recherche de d'anomalies



UPMC SORBONNE UNIVERSITÉS

Synoptique

- *Big Data*
  - *Déluge informationnel*
  - *Les 3 V*
  - *La logique des Big Data*
- *« Big Brother »*
  - *Données personnelles*
  - *Protection?*
- *Qui sont les « Big Brothers »?*
  - *Utilisation des Big Data*
  - *Nouvelle économie*
- *Conséquences politiques*

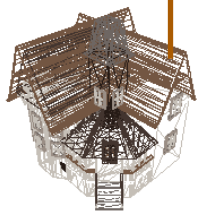
1984 est il devant ou derrière nous?

- Progrès des dispositif de prise d'information personnelle
- Nous sommes suivis à la trace

UPMC
SORBONNE UNIVERSITÉS

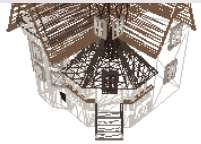
Ces traces que nous laissons!

- Conscientes
  - Réseaux sociaux
  - Blogs, écrits, …



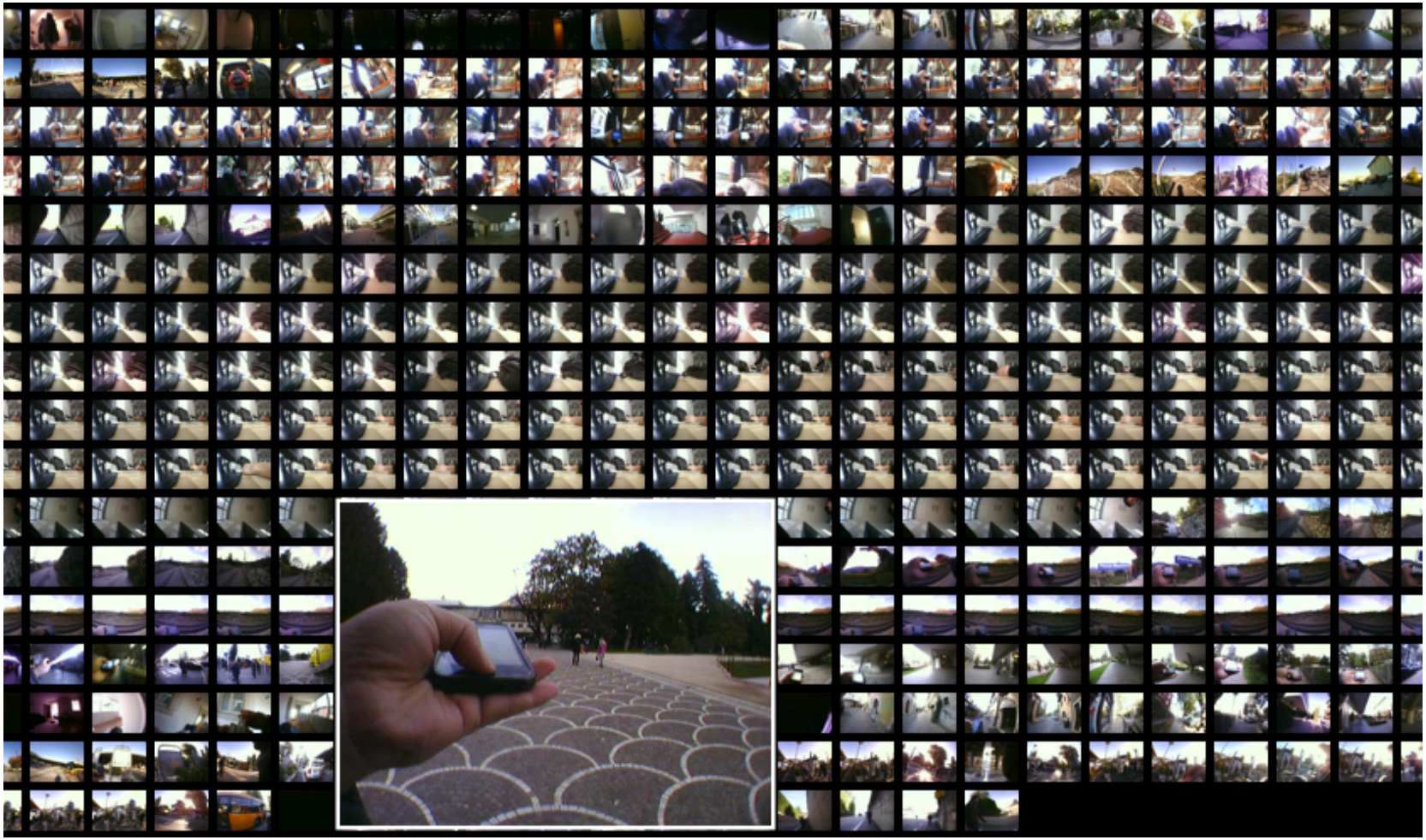
UPMC
SORBONNE UNIVERSITÉS

Archives individuelles



UPMC
SORBONNE UNIVERSITÉS

« Lifelog » – mémoire de vie



SORBONNE UNIVERSITÉS

Kiwi: traces physiologiques

kiwi wearable tech

Home | Product | Developers | About Us | Contact

PRODUCT INFO

Temperature | Heart Rate | Blood Pressure | Location (GPS) | Activity/Motion | Time

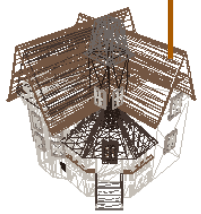
Body Vitals + Context

The Kiwi accurately and continously measures your body vitals using the latest in sensor technology. Using real time data from different sensors the Kiwi is able to derive blood pressure, stress, and a host of other signals from your body

Combining rich sensor data with context is where the real magic happens. Without user intervention the Kiwi is able to distinguish between sleep and a variety of different modes (running, walking, sitting). With this data, the number of potential applications is endless!

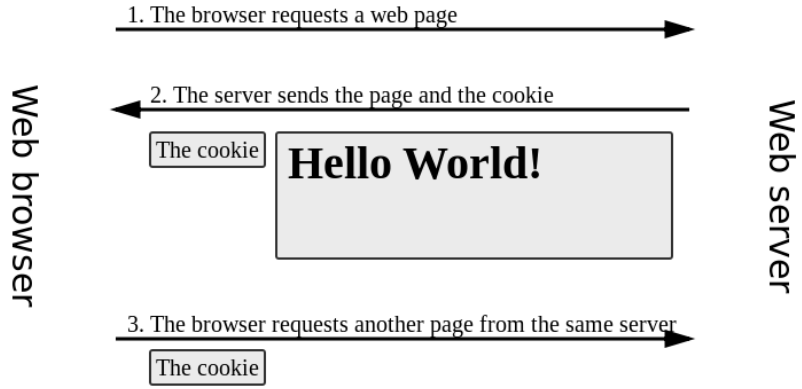
Ces traces que nous laissons!

- Conscientes
  - Réseaux sociaux
  - Blogs, écrits, …
- Par devers nous
  - Paiements bancaires
  - Cartes de santé
  - Passe navigo, …
- Inconscientes
  - Coups de fils, mails…
  - Cookies

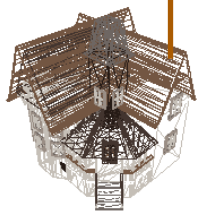


UPMC SORBONNE UNIVERSITÉS

Cookies

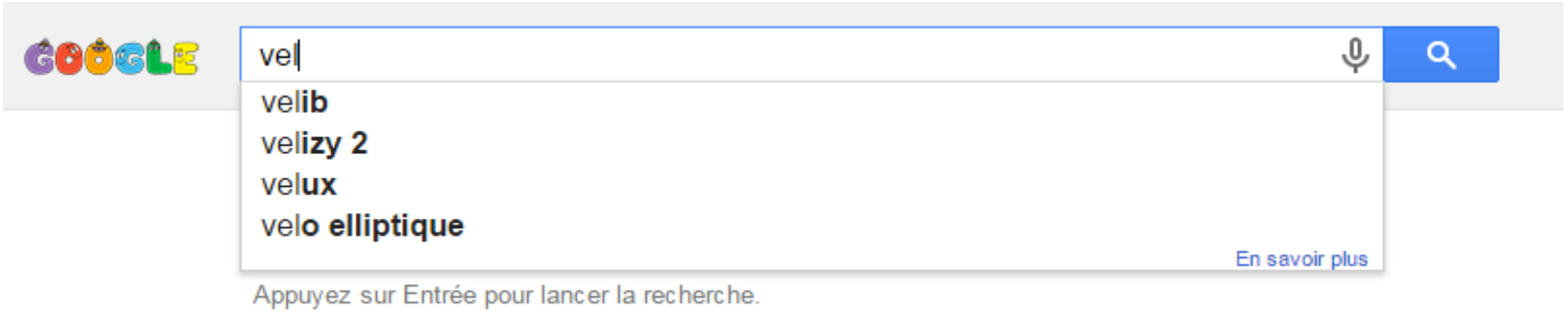



```
HTTP/1.1 200 OK
Cache-Control: private
Content-Type: text/html
Set-Cookie: PREF=ID=5e66ffd215b4c5e6:
TM=1147099841:LM=1147099841:S=Of69MpW
Bs23xeSv0; expires=Sun, 17-Jan-2038 1
9:14:07 GMT; path=/; domain=.google.c
om
```

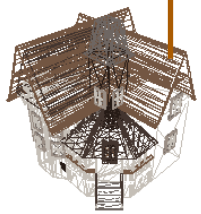


UPMC SORBONNE UNIVERSITÉS

Requêtes sur moteurs de recherche



- Auto-complétion des requêtes
- Suivi et profilage
  - Publicité ciblée
  - Sécurité…
- Exploitation massive de ces requêtes – *Big Data*

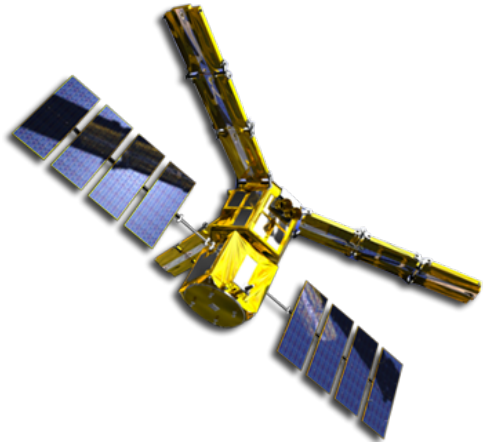


UPMC SORBONNE UNIVERSITÉS

Echapper?

- Peut-on l'espérer en se contentant de cultiver son jardin?

- Pas pour longtemps…

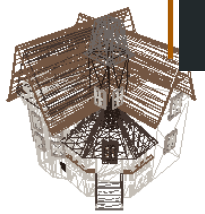


UPMC
SORBONNE UNIVERSITÉS

Synoptique

- *Big Data*
  - *Déluge informationnel*
  - *Les 3 V*
  - *La logique des Big Data*
- *« Big Brother »*
  - *Données personnelles*
  - *Protection?*
- *Qui sont les « Big Brothers »?*
  - *Utilisation des Big Data*
  - *Nouvelle économie*
- *Conséquences politiques*

Changer de moteur de recherche



UPMC SORBONNE UNIVERSITÉS

Principe de finalité de la CNIL

- **Big Data**: recueil systématique, sans hypothèse *a priori*
- Comment respecter ce principe dans le contexte des *Big Data*?

La finalité des traitements

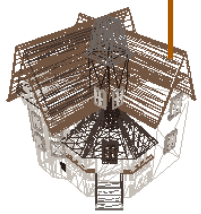
Un fichier doit avoir **un objectif précis**.

Les informations exploitées dans un fichier doivent être **cohérentes par rapport à son objectif**.

Les informations **ne peuvent pas être réutilisées de manière incompatible avec la finalité** pour laquelle elles ont été collectées.

Tout détournement de finalité est passible de 5 ans d'emprisonnement et de 300 000 € d'amende.

art. 226.21 du code pénal



UPMC SORBONNE UNIVERSITÉS

Principe de proportionnalité de la CNIL

- **Big Data**: conservation, partage et accumulation de toutes les données
- Comment respecter ce principe dans le contexte des *Big Data*?

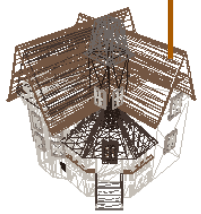
La durée de conservation des informations

Les données personnelles **ont une date de péremption**.

Le responsable d'un fichier fixe **une durée de conservation raisonnable** en fonction de l'objectif du fichier.

Le code pénal sanctionne la conservation des données pour une durée supérieure à celle qui a été déclarée de 5 ans d'emprisonnement et de 300 000 € d'amende.

art. 226-20 du code pénal



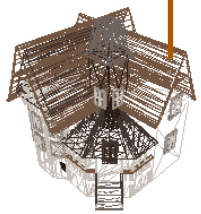
UPMC SORBONNE UNIVERSITÉS

Peut-on oublier sur Internet?

- Fonctionnement moteurs de recherche:
  - Aspiration du web avec des bots, puis analyse
  - Indexation des sites:
    - Principe: dictionnaire inversé:
      De « collectionner »
      et « monnaies »,
      retrouver
      « numismate » et
      exclure
      « philatéliste » parce
      qu'il y a « monnaies »

philatéliste
n. m. et f.
Personne qui collectionne les timbres-poste.

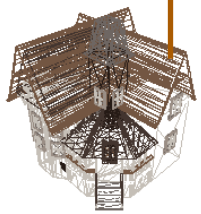
numismate
n. m. et f.
Personne qui collectionne les monnaies anciennes et des médailles.



UPMC SORBONNE UNIVERSITÉS

Peut-on oublier sur Internet?

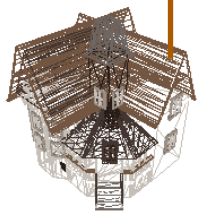
- Fonctionnement moteurs de recherche:
  - Aspiration du web avec des bots, puis analyse
  - Indexation des sites:
    - Principe: dictionnaire inversé
  - Ordonnancement des sites:
    - Élimination des mots sans intérêt (TFIDF)
      - Mots trop fréquents
      - Mots trop rares (hapax)
    - Algorithme de préférences (ex. Google)



UPMC SORBONNE UNIVERSITÉS

Indexation moteurs de recherche

- **Mot_1**: $url_{1,1}$, $url_{1,2}$, $url_{1,3}$, ... $url_{1,n}$, $url_{1,n+1}$, ...
- **Mot_2**: $url_{2,1}$, $url_{2,2}$, $url_{2,3}$, ... $url_{2,m}$, $url_{2,m+1}$, ...
- ...
- **Mot_p**: $url_{p,1}$, $url_{p,2}$, $url_{p,3}$, ... $url_{p,l}$, $url_{p,l+1}$, ...
- ...



UPMC SORBONNE UNIVERSITÉS

Oubli sur Internet Mot_P $url_{p,1}$, $url_{p,2}$

Sociétés commerciales de déréférencement:

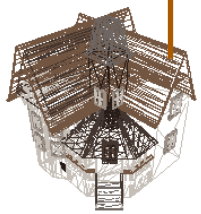
- **Mot_1**: $url_{1,1}$, $url_{1,2}$, $url_{1,3}$, ... $url_{1,n}$, $url_{1,n+1}$, ...
- **Mot_2**: $url_{2,1}$, $url_{2,2}$, $url_{2,3}$, ... $url_{2,m}$, $url_{2,m+1}$, ...
- ...
- **Mot_p**: $url_{p,q}$, $url_{p,q+1}$, $url_{p,q+2}$, ... $Url_{p,r}$,...

$url_{p,1}$, $url_{p,2}$, $url_{p,3}$, ... $url_{p,l}$, $url_{p,l+1}$, ...

- ...

Action: ajout information neutre

But: enfouir l'information gênante



UPMC SORBONNE UNIVERSITÉS

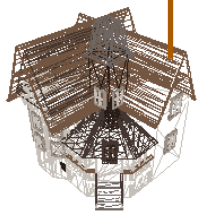
Oubli sur Internet Mot_P $url_{p,1}$, $url_{p,2}$

« Droit à l'oubli » Google

- **Mot_1**: $url_{1,1}$, $url_{1,2}$, $url_{1,3}$, ... $url_{1,n}$, $url_{1,n+1}$, ...
- **Mot_2**: $url_{2,1}$, $url_{2,2}$, $url_{2,3}$, ... $url_{2,m}$, $url_{2,m+1}$, ...
- ...
- **Mot_p**: $url_{p,1}$, $url_{p,2}$, $url_{p,3}$, ... $url_{p,l}$, $url_{p,l+1}$, ...
- ...

Suppression Mot_p: $url_{p,1}$, ... $url_{p,l}$, $url_{p,l+1}$,

Les URLs $url_{p,1}$, $url_{p,2}$, restent présentes, mais associées à d'autres noms...



UPMC SORBONNE UNIVERSITÉS

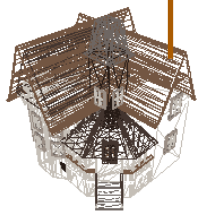
Synoptique

- *Big Data*
  - *Déluge informationnel*
  - *Les 3 V*
  - *La logique des Big Data*
- *« Big Brother »*
  - *Données personnelles*
  - *Protection?*
- *Qui sont les « Big Brothers »?*
  - *Utilisation des Big Data*
  - *Nouvelle économie*
- *Conséquences politiques*

Nouvelle révolution scientifique

- Science moderne
  expérimentation *in vivo* et *in vitro* Intervention sur le monde extérieur

- Science *in silico*
  Expérimentation sur données



UPMC
SORBONNE UNIVERSITÉS

« Au bonheur des dames » *Emile Zola*

SORBONNE UNIVERSITÉS

Marchandisation des identités

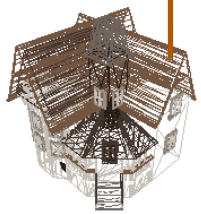
De la

« femme sans tête »

à la

« femme 100 têtes »

Importance du profilage…

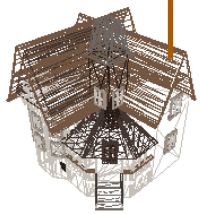


UPMC SORBONNE UNIVERSITÉS

Qui sont les Big Brothers?

- Gouvernements
  - PRISM (Etats-Unis)
  - …

BIG BROTHER IS WATCHING YOU

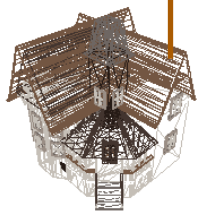


UPMC SORBONNE UNIVERSITÉS

PRISM

Differents programmes de surveillance

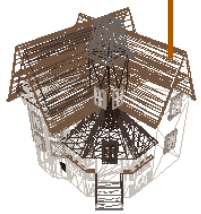
- « UPSTREAM »: Aspiration données sur cables sous-marins
  - Mails
  - Téléphones
- Requêtes auprès des acteurs de l'internet
  - AOL, Skype, Microsoft, Facebook, Yahoo, Apple, Google, YouTube, …



UPMC SORBONNE UNIVERSITÉS

Qui sont les Big Brothers?

- Gouvernements
  - PRISM (Etats-Unis)
  - …
- Sociétés privées
  - GAFA (*Goole, Apple, Amazon, Facebook*)
  - Nouvelle économie fondée sur le profilage

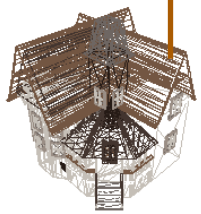


UPMC
SORBONNE UNIVERSITÉS

La prise de pouvoir des GAFA

Prérogatives régaliennes des Etats

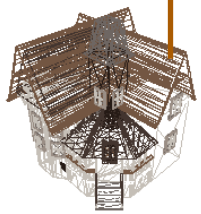
- Battre monaie
- Organiser la santé
- Gérer l'état civil
- S'occuper de la formation
- Dresser le cadastre
- …



UPMC
SORBONNE UNIVERSITÉS

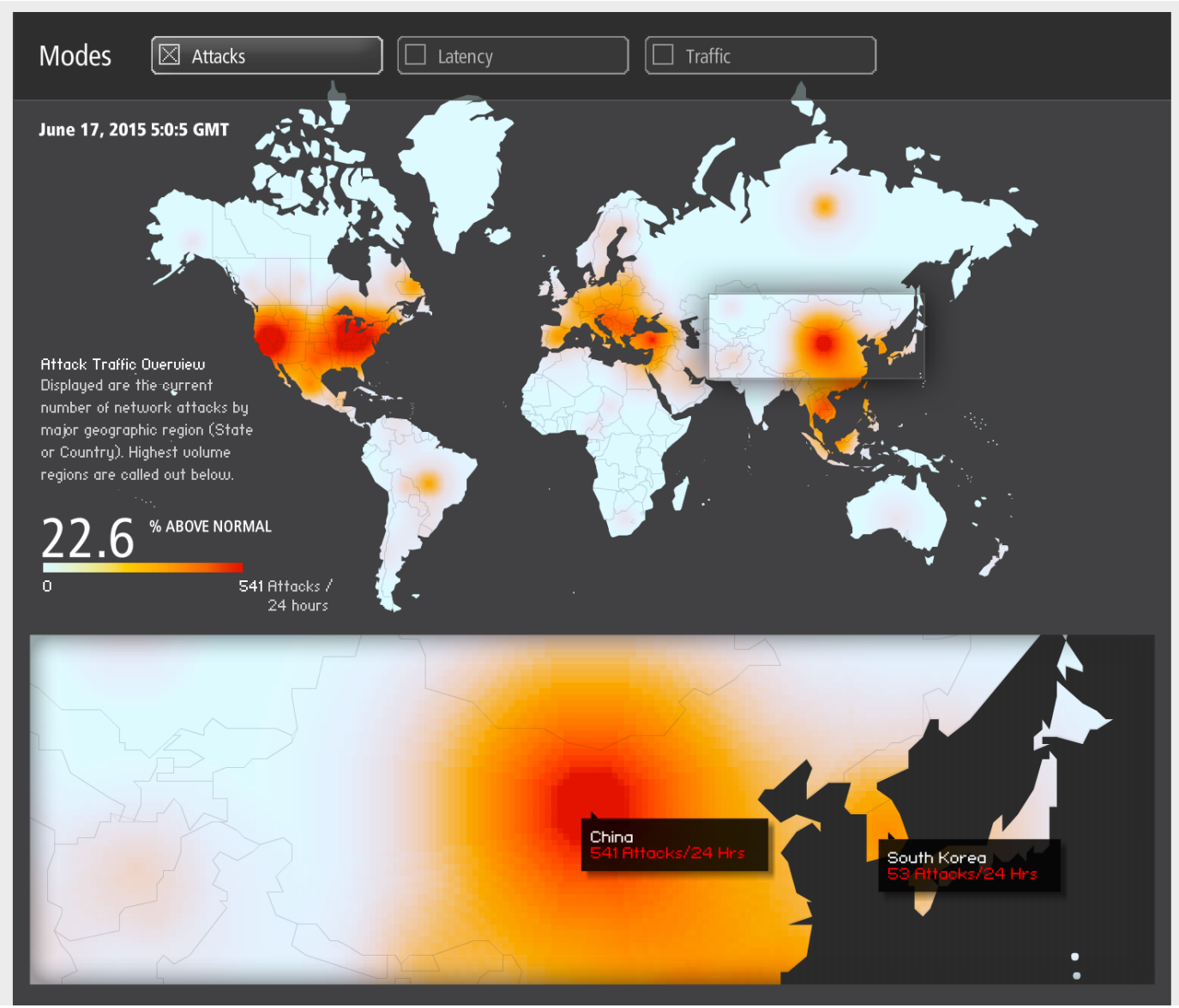
Qui sont les Big Brothers?

- Gouvernements
  - PRISM (Etats-Unis)
  - …
- Sociétés privées
  - GAFA (*Goole, Apple, Amazon, Facebook*)
  - Nouvelle économie fondée sur le profilage
- Nouveaux acteurs
  - Hommes politiques
  - Maffias
  - ONG
  - …

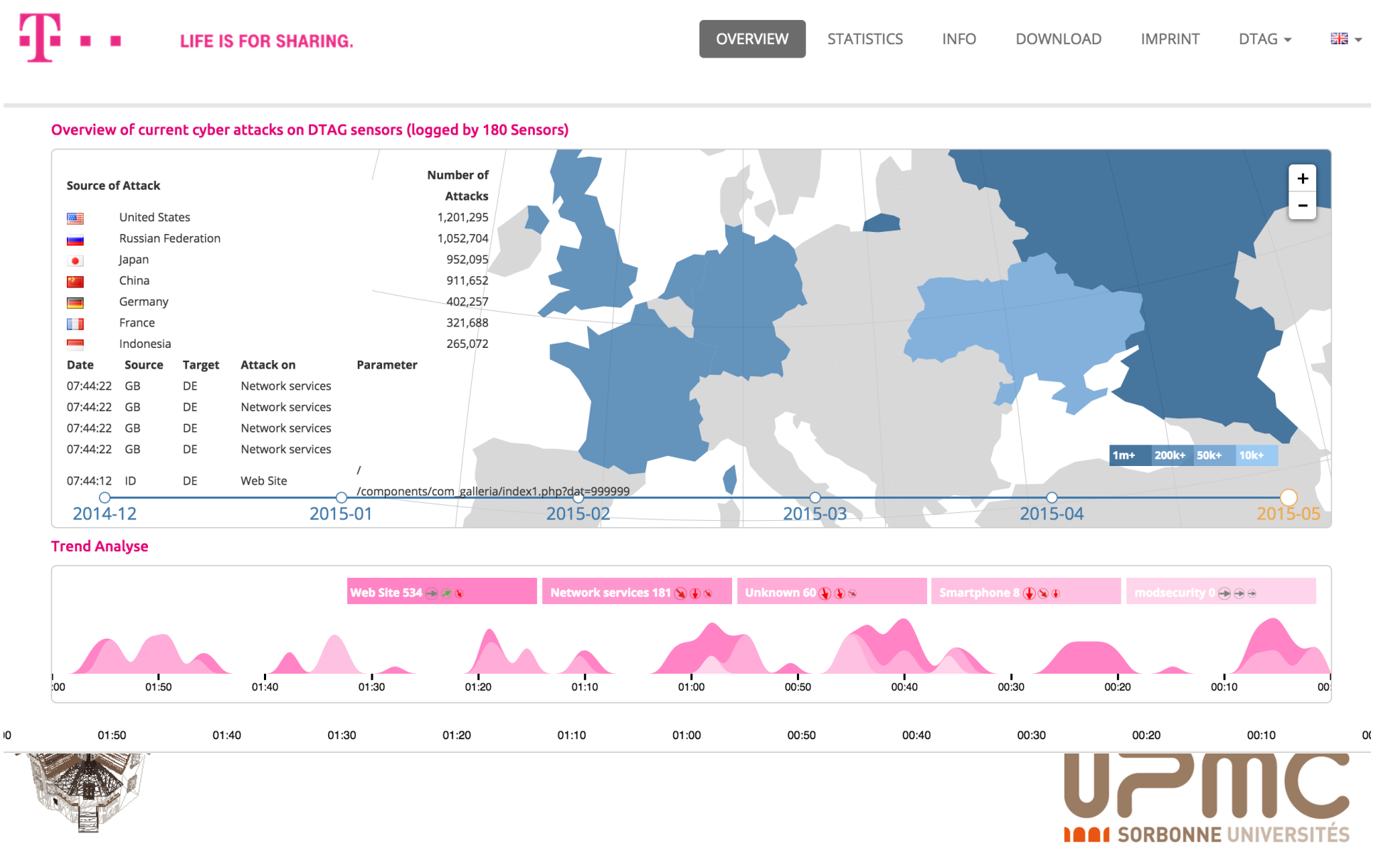


UPMC SORBONNE UNIVERSITÉS

Cyber-attaques

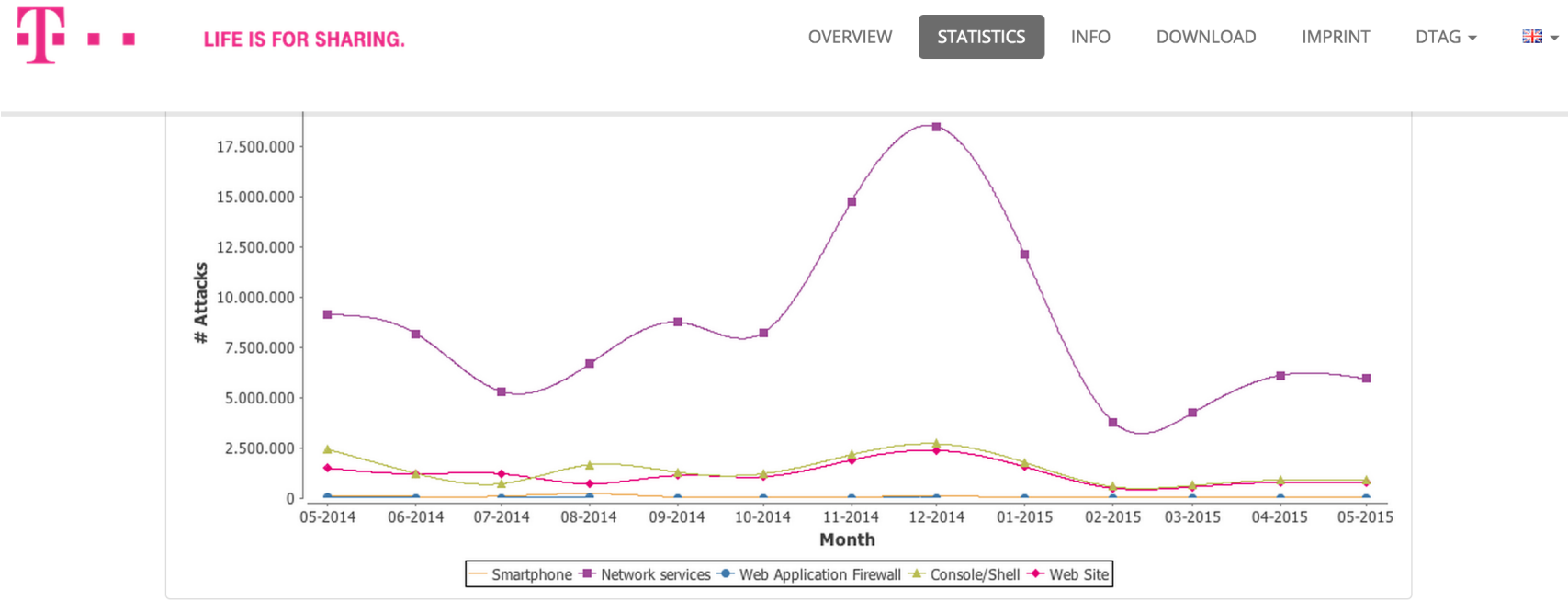


Cyber-attaques

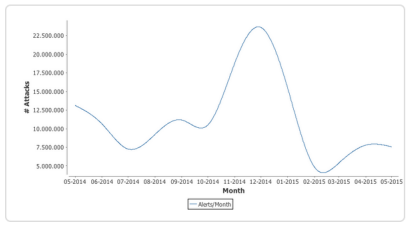


SORBONNE UNIVERSITÉS

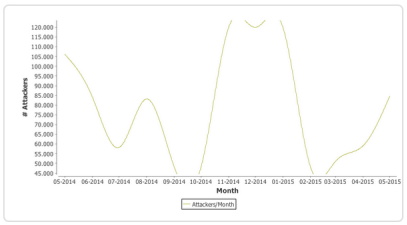
Nombre d'attaques



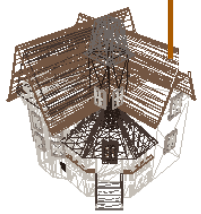
Overall sum of attacks per Month



Overall sum of attackers per Month

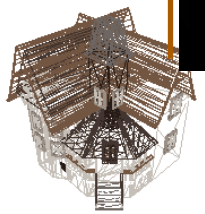


Passwords



SORBONNE UNIVERSITÉS

ONG – Maffias – Etats



UPMC
SORBONNE UNIVERSITÉS

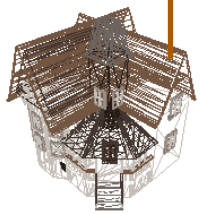
« 1984 » est-il derrière ou devant?

L'année 1984 est en arrière…

Les dispositifs de prise d'information sont en avance

Faut-il craindre la généralisation de la surveillance?

Ou la perte de souveraineté numérique au profit de nouveaux acteurs?



UPMC
SORBONNE UNIVERSITÉS