

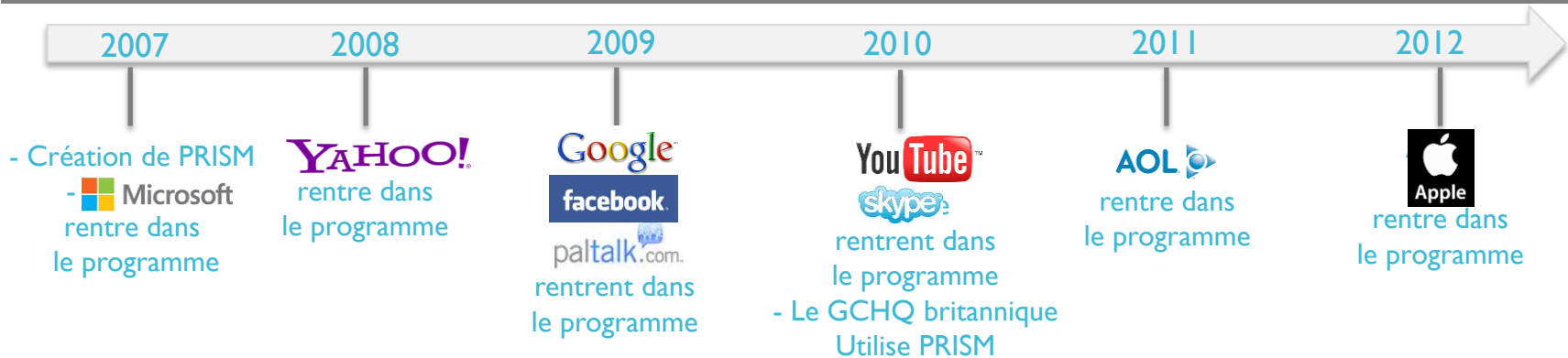


Big Data, Big Brother

Séminaire ASPROM Big Data

3 juillet 2013

Programme américain qui permettrait aux NSA et FBI d'accéder aux données d'internautes non américains sur certains réseaux sociaux et services de communication :



Types de données collectées :

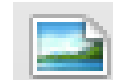
Seraient collectées semi-automatiquement
Par le Data Intercept Technology Unit du FBI
Fournirait 2000 rapports/mois
Budget : \$20M/an



E-mail



Vidéo



Photos



VOIP



Transfert de fichiers



Activités de la cible



Données stockées dans le Cloud



Conversations : Vidéo, Audio Vidéoconférences



Informations sur les réseaux sociaux



Edward Snowden

Ancien consultant employé par le NSA qui a révélé l'existence du programme PRISM



Ce que l'on ne sait pas encore

- **Détails techniques du processus de collecte et sélection des données :**
- **Quel est le point de collecte? :**
 - Directement chez les serveurs des entreprises (Google, Microsoft,...) ou
 - En amont chez les fournisseurs d'accès à internet ou sur les CDN, les data center?
- **Nous savons que :**
 - Le projet « dépendant à 100% du provisioning chez les fournisseurs d'accès à internet »
 - Les analystes ont accès aux données stockées sur les serveurs des partenaires et l'activité de la cible en temps réel
- **L'implication de :**
- **Dropbox**
- **Pays de l'Union Européenne :**
 - UK, Allemagne, France, Pays-Bas, Danemark, Italie, Espagne

Les questions que PRISM soulève

- **PRISM a-t-il tué le Cloud?**
- **Cela va-t-il accélérer l'émergence d'alternatives européennes à Facebook, Google, Microsoft, ...?**
- **N'oublions pas que PRISM n'est qu'un élément d'un ensemble progressivement mis en place depuis des années, y compris en France : PNIJ (Plate-forme Nationale des Interceptions Judiciaires)**
- **Peut-on tout faire au nom de la sécurité ?**

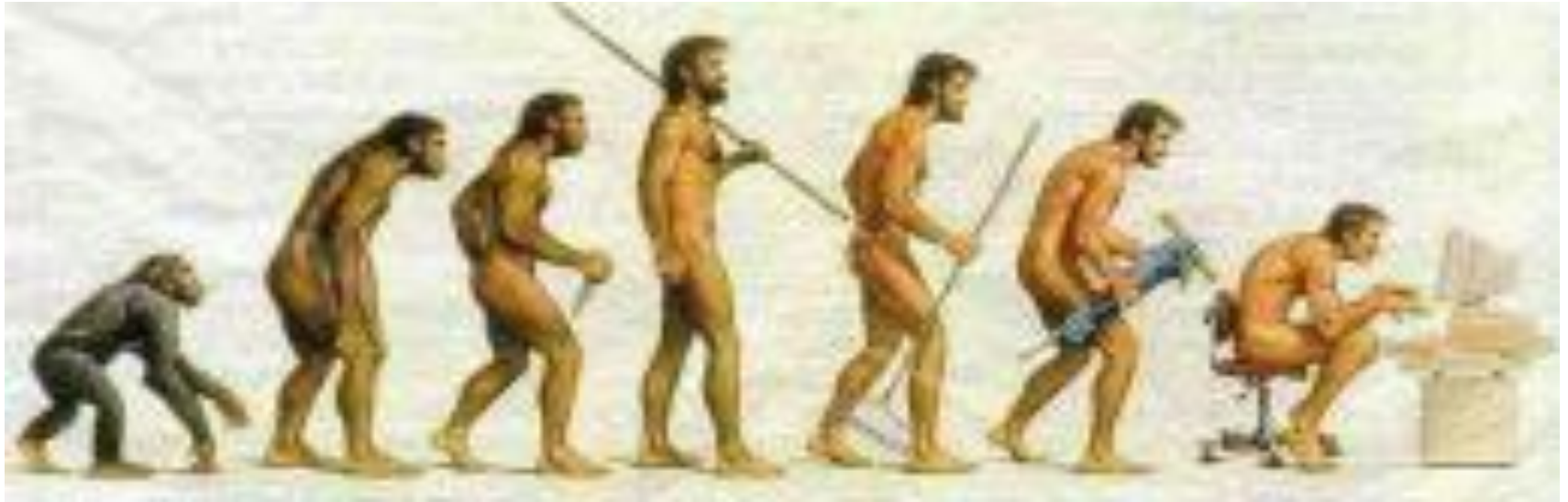


- Les criminels sont aujourd'hui principalement identifiés par les objets communicants



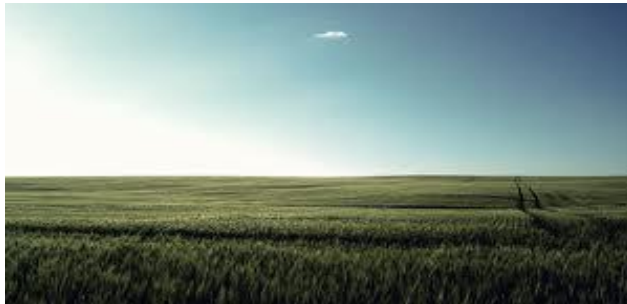
Dans la lutte contre le terrorisme, ce qui se passe « derrière le pare-feu » est tout aussi important que ce qui se passe sur le scène du crime

- 1 ■ **Moins de 24h après l'explosion, le FBI compile 10To de données**
- 2 ■ **Il analyse ces données pour repérer rapidement les suspects :**
 - Des milliers d'appels de téléphones portables, messages, données des réseaux sociaux, photos, séquences de vidéo surveillance sont analysées, décortiquées
 - Ils utilisent également un logiciel de reconnaissance faciale
- 3 ■ **Les recherches révèlent des références à la bombe supprimées sur les comptes Twitter des 2 suspects**



Mais est-ce normal de jouer avec les ordinateurs en étant nu et vulnérable ?

- **Mais à qui appartiennent les données ?**
- Données personnelles
- Photos
- Trafic
- Localisation
- Traces
-

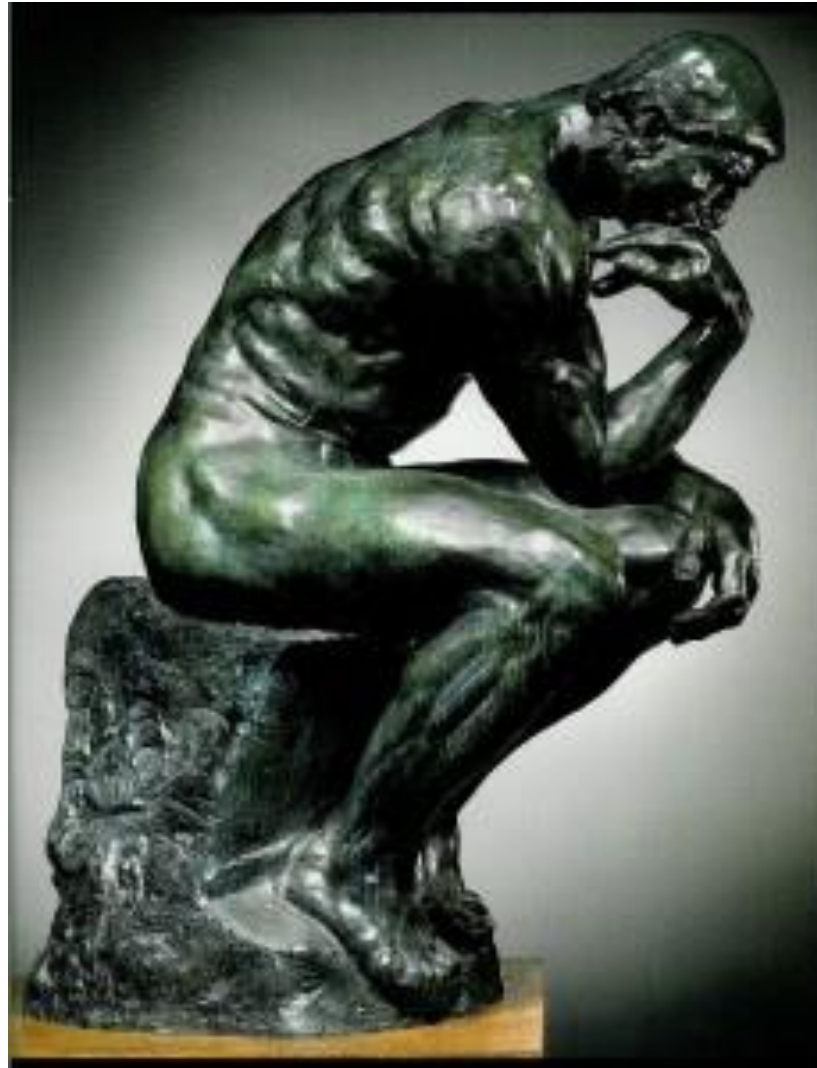


Phase 1 : Le corbeau et le renard



- On parle de Big Data
- Mais ces objets sont avant tout des objets personnels
- La valeur de ces objets sont intimement liés à l'usage et aux personnes







\$41

1 Milliard d'Utilisateurs qui partagent leur vie privée

Valorisation boursière : \$41 Milliards

Google™

\$64

3,5 Milliards d'utilisateurs

Valorisation boursière : \$224 Milliards

Fondé par Larry Page et Sergey Brin en 2010, ce complexe basé à Mountain View a pour objectif de développer les technologies de demain

Google Glass



Google Loon



Google Car



Réseau Neuronal

- Réseau neuronal à 9 couches qui permet de détecter les visages sur des images sans classer ces dernières au préalable.
- Le réseau, composé de 1 000 machines pour 16 000 cœurs et 1 milliard de connexions, a fonctionné pendant 3 jours pour analyser les données issues de 10 millions d'images (en format 200 × 200 pixels) issues de vidéos YouTube choisies aléatoirement.
- Il en a extrait des concepts déterminés sans caractéristiques initiales à rechercher, et notamment « l'idée » de visage de chat, ou de silhouettes humaine l8.
- À partir de ces données, le réseau parvient à reconnaître 20 000 catégories d'objets avec un taux de précision de 15,8 % (soit une amélioration de 70 % par rapport à ce qui se faisait de mieux dans le domaine auparavant).

Historique

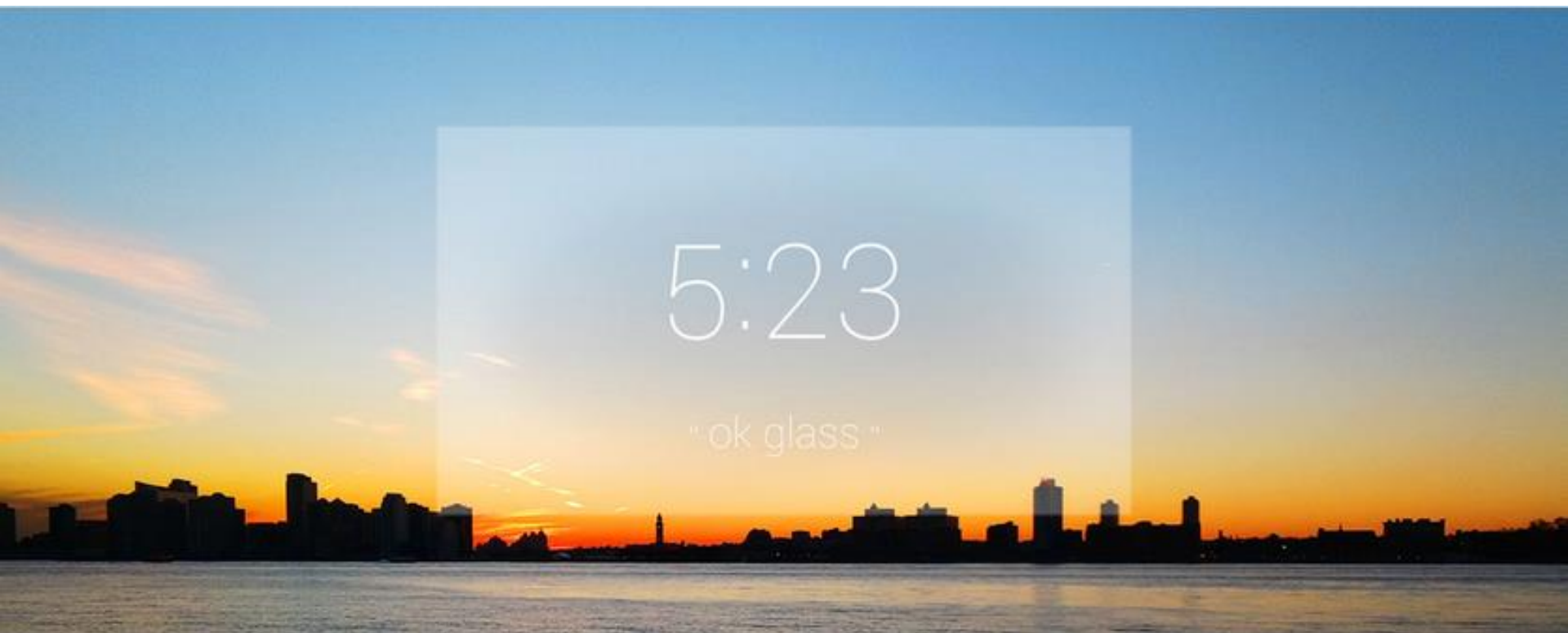
- Les Google Glass sont des lunettes que Google pense **commercialiser en 2014**, affichant des informations issues d'Internet en surimpression de la vue classique (principe de la réalité augmentée)
- Les Google Glass entrent en phase de **test en avril 2012**
- Le **20 février 2013**, la firme lance un site afin de **recruter des volontaires pour tester** ces lunettes dont le prototype sera mis en vente à 1 499 \$.
- Il faut noter que le concept date de 2001. En effet la société MINOLTA - bien connue dans le monde de la photo - a créé des lunettes proposant ce genre de vision. Elles ne possédaient pas cependant l'interactivité de ce que propose Google.



Fonctionnalités

- Concrètement, **pour utiliser** ces lunettes, vous n'avez rien à faire, ou bien **simplement leur parler !**
- dites "**prendre une photo**" pour qu'elles prennent immédiatement une photo
- dans le même ordre d'idées, vous pouvez **filmer ce que vous êtes en train de voir**
- ces photos ou vidéos peuvent **être partagées avec vos amis / contacts**
- obtenez un **guidage GPS** de votre itinéraire directement en surimpression de ce que vous voyez (principe de la **réalité augmentée**)
- enregistrez un message audio à envoyer
- demandez une **information de type encyclopédique**, par exemple l'année de construction du monument qui est devant vous, vous aurez la réponse immédiatement sous les yeux
- obtenez une **traduction instantanée** d'une phrase que vous dictez
- obtenez des informations sans même les demander, comme dans le très bon **Google Now** (dans la photo de démonstration, on voit une personne qui marche dans un aéroport : ses lunettes lui indiquent quel vol elle doit prendre, lui confirment qu'il est à l'heure et précisent dans combien de temps elle doit embarquer, sans qu'elle ait eu besoin de le demander)

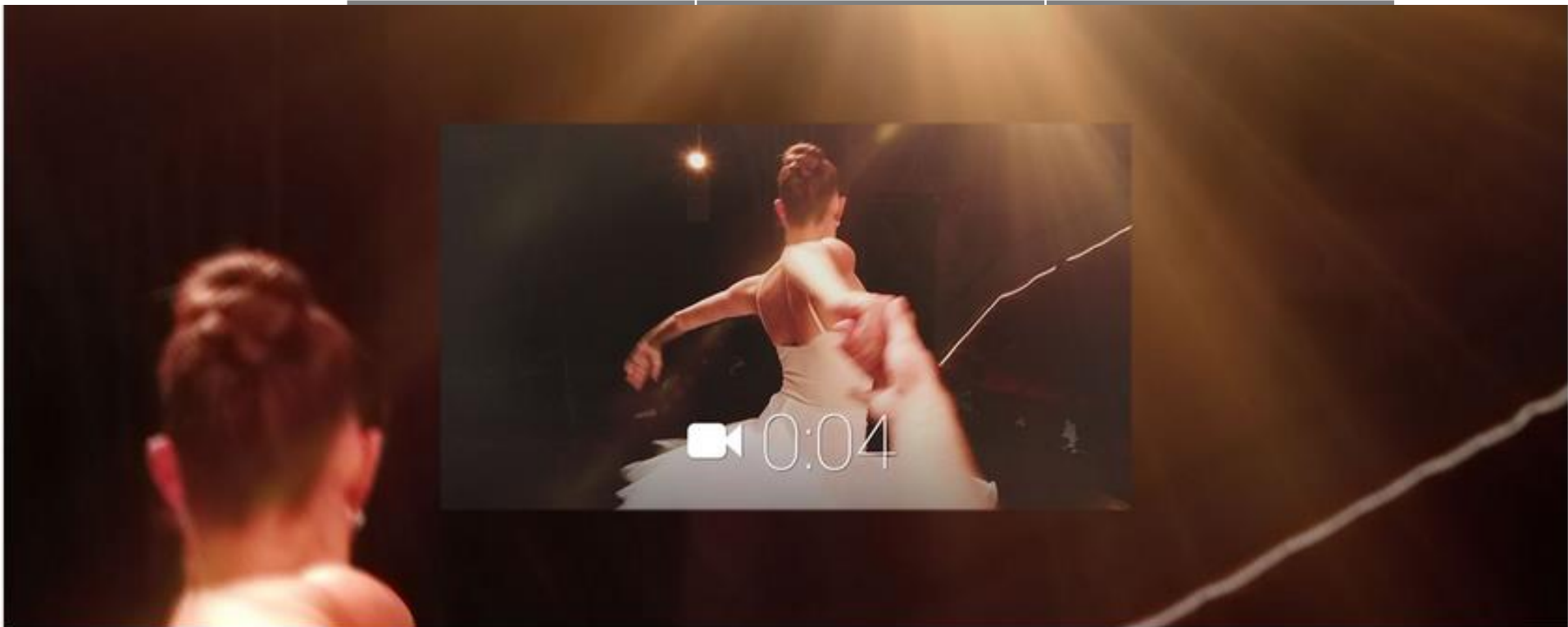
<http://www.google.com/glass/start/how-it-feels/>



It's surprisingly simple.



Say "take a picture" to take a picture.



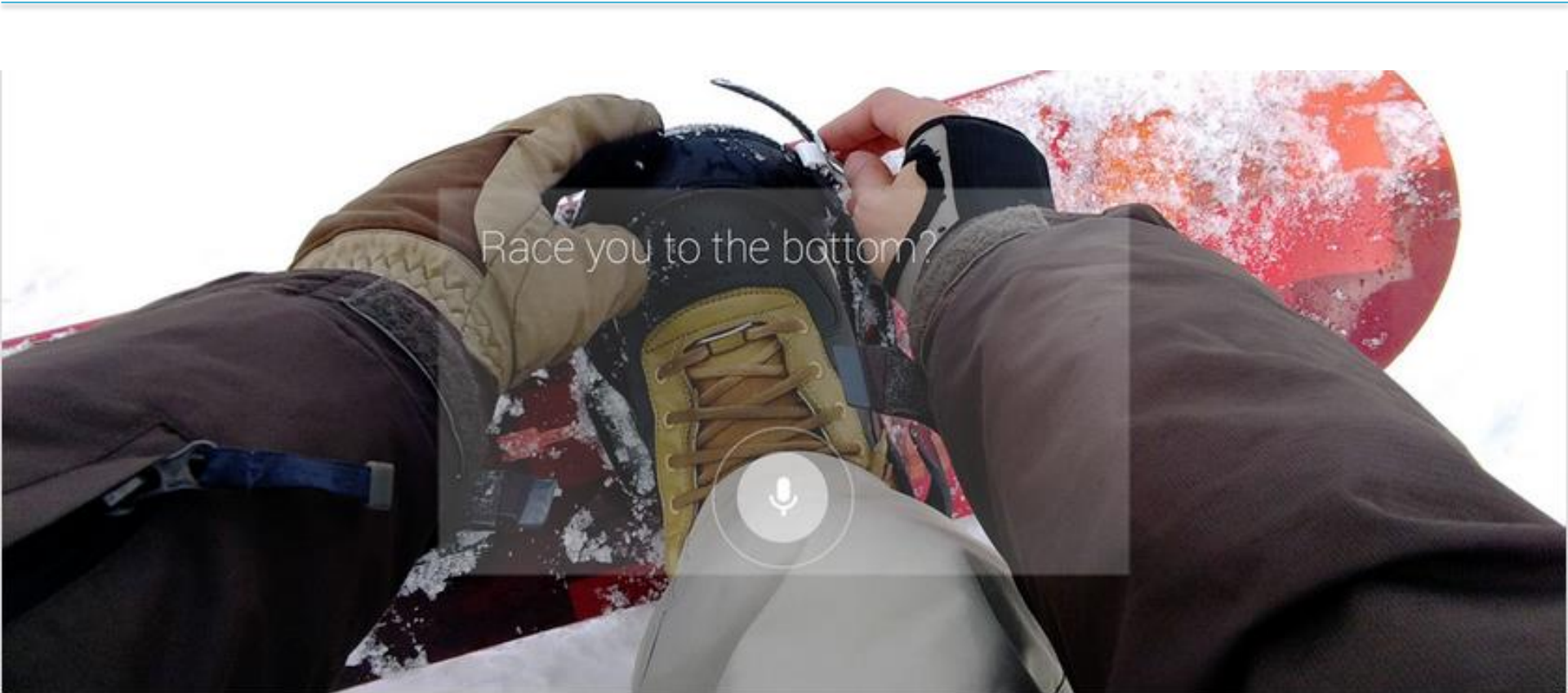
Record what you see. Hands-free.



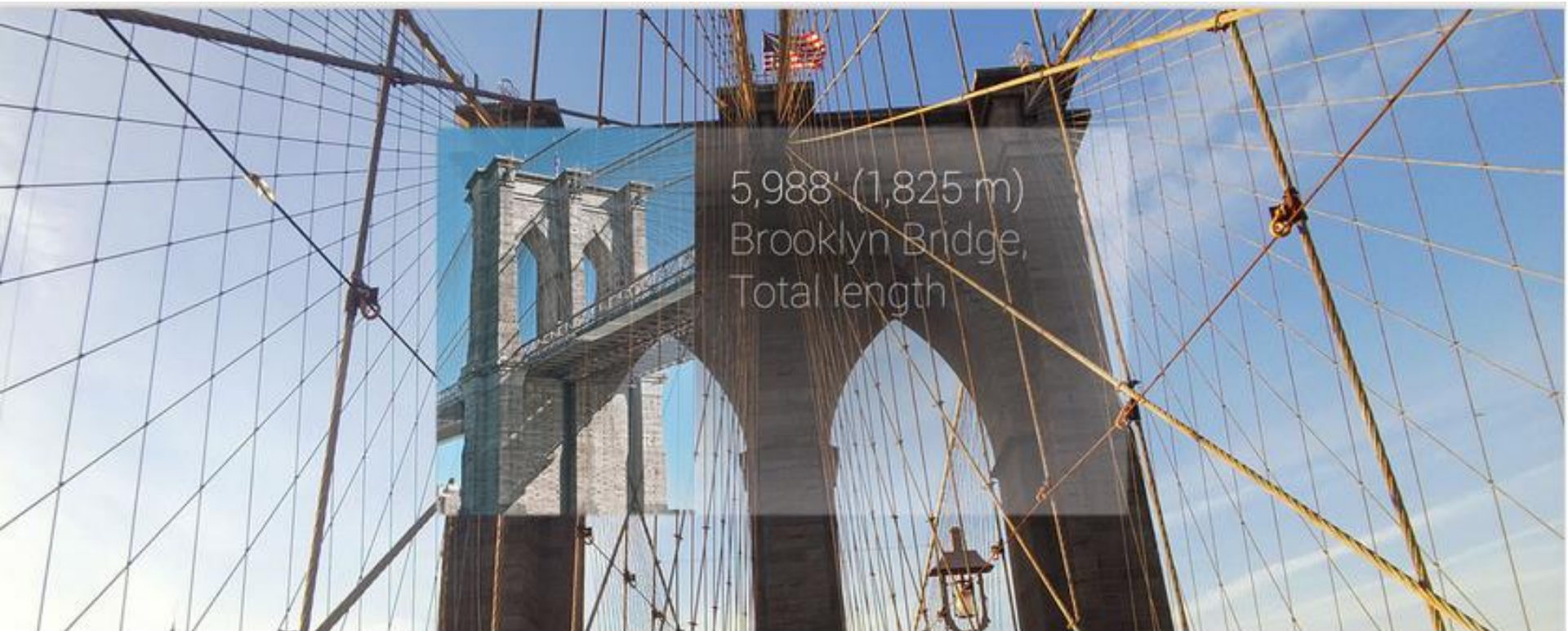
Even share what you see. Live.



Directions right in front of you.



Speak to send a message.



Ask whatever's on your mind.



Translate your voice.



Answers without having to ask.

But du projet : Avoir accès à internet partout dans le monde

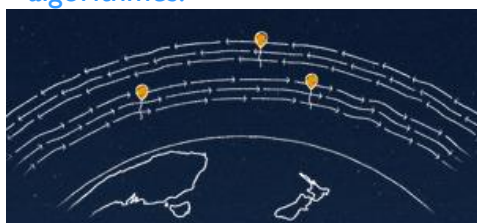
Comment : en créant un réseau de ballons (Loons) à 20km du sol

Le pilote a été lancé en Nouvelle-Zélande mi juin : 30 Loons gonflés à l'hélium ont été envoyés dans la stratosphère

Stratosphère et ses vents

Les Loons sont envoyés à 20km du sol : au dessus des montagnes, nuages et avions, dans la stratosphère

Le but est d'utiliser les vents de la stratosphère pour déplacer les Loons là où il y a besoin d'aller, ce qui est déterminé par des algorithmes.



Une antenne spécifique communique avec un Loon, puis ce Loon communique avec les autres Loons ce qui crée un réseau.

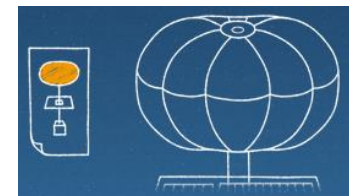
La portée de signal d'un Loon est de 40km de diamètre à une vitesse comparable à la 3G

Le réseau



Composition du Loon

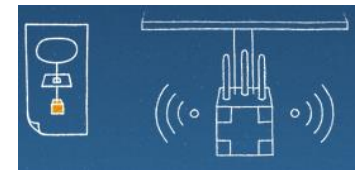
Un ballon en toile spécifique pour être résistant aux conditions, avec un parachute en cas de problème



Des panneaux solaires pour être autonome en énergie



Equipements qui permettent aux Loons de communiquer entre eux, avec les antennes au sol et la batterie pour stocker l'électricité des panneaux solaires



<http://www.google.com/loon/how/#tab=navigation&tab=equipment>

Historique

- La voiture sans conducteur de Google, parfois appelée Google Car, est un projet de **voiture automate** développé par Google.
- Il est dirigé par Sebastian Thrun, également directeur du Stanford Artificial Intelligence Laboratory et co-inventeur de Google Street View.
- En octobre **2010**, Google annonce avoir **conçu un système de pilotage automatique pour automobile**, déjà installé sur huit véhicules qui ont parcouru plus de 200 000 km en Californie, sans avoir provoqué d'accident
- Elle est **autorisée** dans 3 Etats aux USA : Le **Nevada, la Californie et la Floride**
- Google prévoit sa **commercialisation dans 5 ans**
- Sergey Brin voit plus loin. Il rappelle que les voitures sont utilisées d'une façon bien peu optimisée. « Elles servent à rouler seulement 3 % du temps », ce qui a des répercussions en termes d'espaces urbains dédiés à des places de stationnement. Les Google Cars pourraient donc changer profondément les villes et leurs banlieues, en facilitant le partage de véhicule, en réduisant le nombre de places de parking nécessaires, etc.

Technologies

- La **téledétection par laser** (aussi appelée "Lidar") :
 - c'est concrètement un capteur multidirectionnel rotatif posé sur le toit du véhicule, d'une portée de plus de 60 mètres, qui va pouvoir générer une **carte précise en 3 dimensions de l'environnement de la voiture**. Il s'agit plus précisément du modèle HDL-64 E de la marque Velodyne. C'est le cœur du système, et aussi son élément le plus cher (70 000 dollars)
- Un **capteur de mouvement** :
 - il sera posé au-dessus de la roue arrière gauche. Il est chargé de mesurer et enregistrer tous les mouvements, même les plus petits, effectués par la voiture. Il aide la voiture à se géolocaliser correctement.
- Une **caméra** :
 - elle est installée près du rétroviseur intérieur. Détectant les feux et toute la signalisation, elle permet aussi de reconnaître les obstacles mobiles, comme les pétons ou les cyclistes.
- Un **radar** :
 - quatre capteurs placés sur les pare-chocs, trois devant et un à l'arrière, permettent de déterminer la distance entre le véhicule et les objets. Leur portée est suffisante pour aider la voiture à repérer les autres véhicules qui roulent vite sur les autoroutes.

<http://www.youtube.com/watch?v=c dg Qpa l p U UE>

Linked  **\$65**

175 millions de profils professionnels

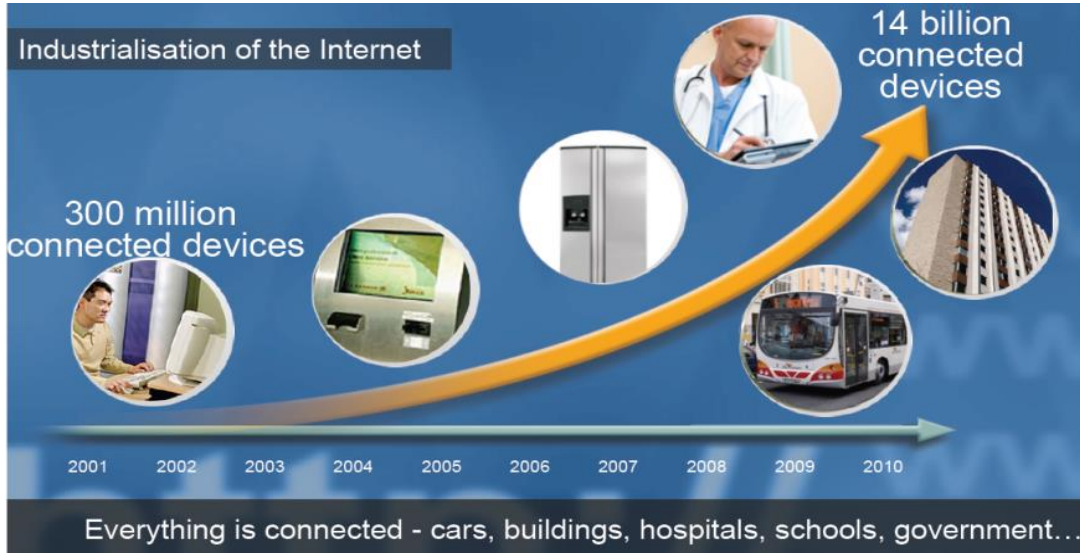
Valorisation boursière : \$11,3 Milliards

\$170 / €130

Aujourd'hui

Quelle valeur
demain ?

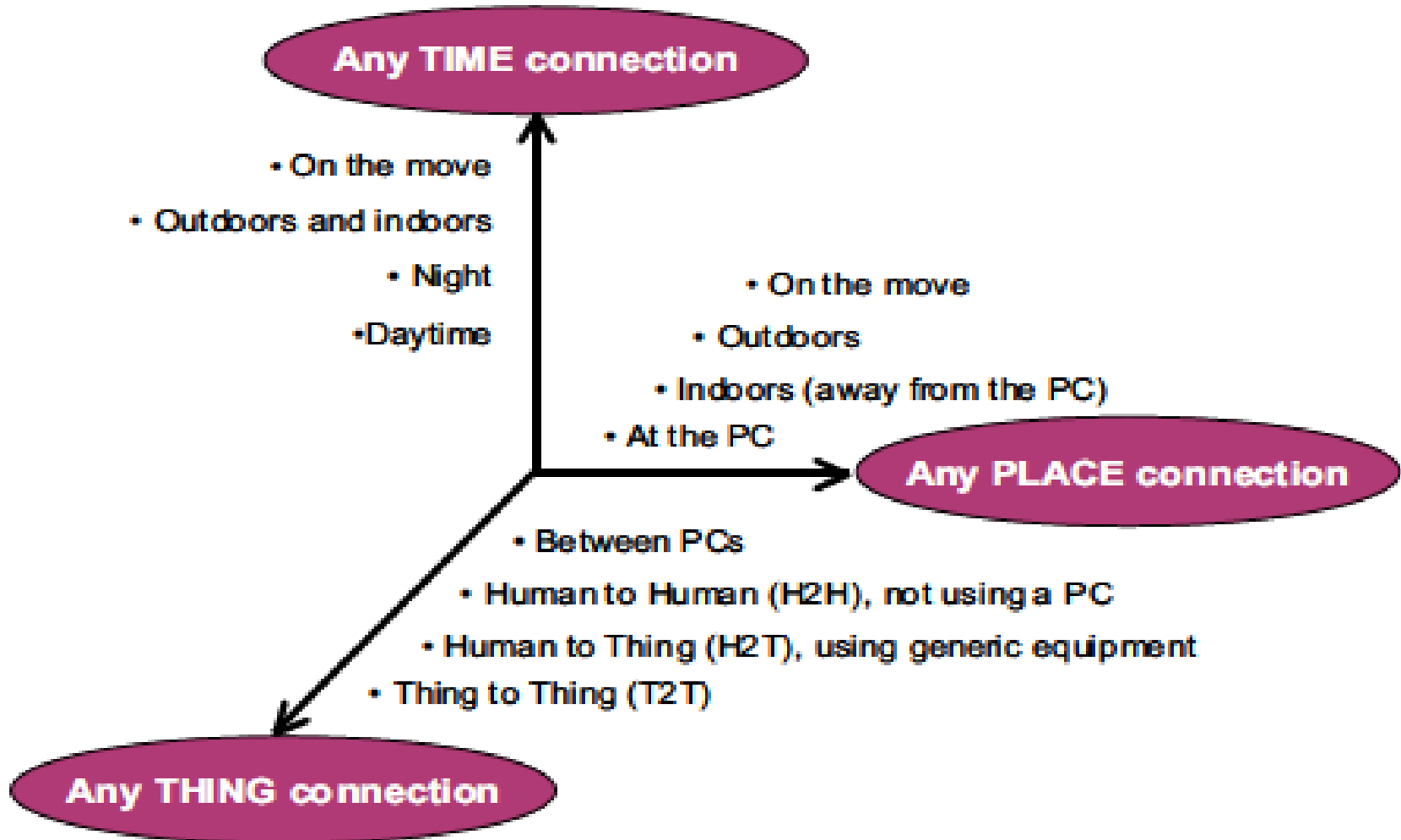




Aujourd'hui, 9 milliards d'équipements connectés à l'internet

En 2020, cela devrait représenter entre 24 et 100 milliards d'équipements





- On peut donc estimer que la valeur va être multipliée par un facteur >100
- $> 17\ 000\$ / 13\ 000\ €$ par personne
- C'est évidemment un trésor qui va éveiller les convoitises
- Qui va gagner cette course ?





Phase 2 : Star Wars





- Facebook a récemment publié dans la timeline les conversations personnelles des utilisateurs notamment français
- Facebook est propriétaire de tous les contenus que les utilisateurs publient

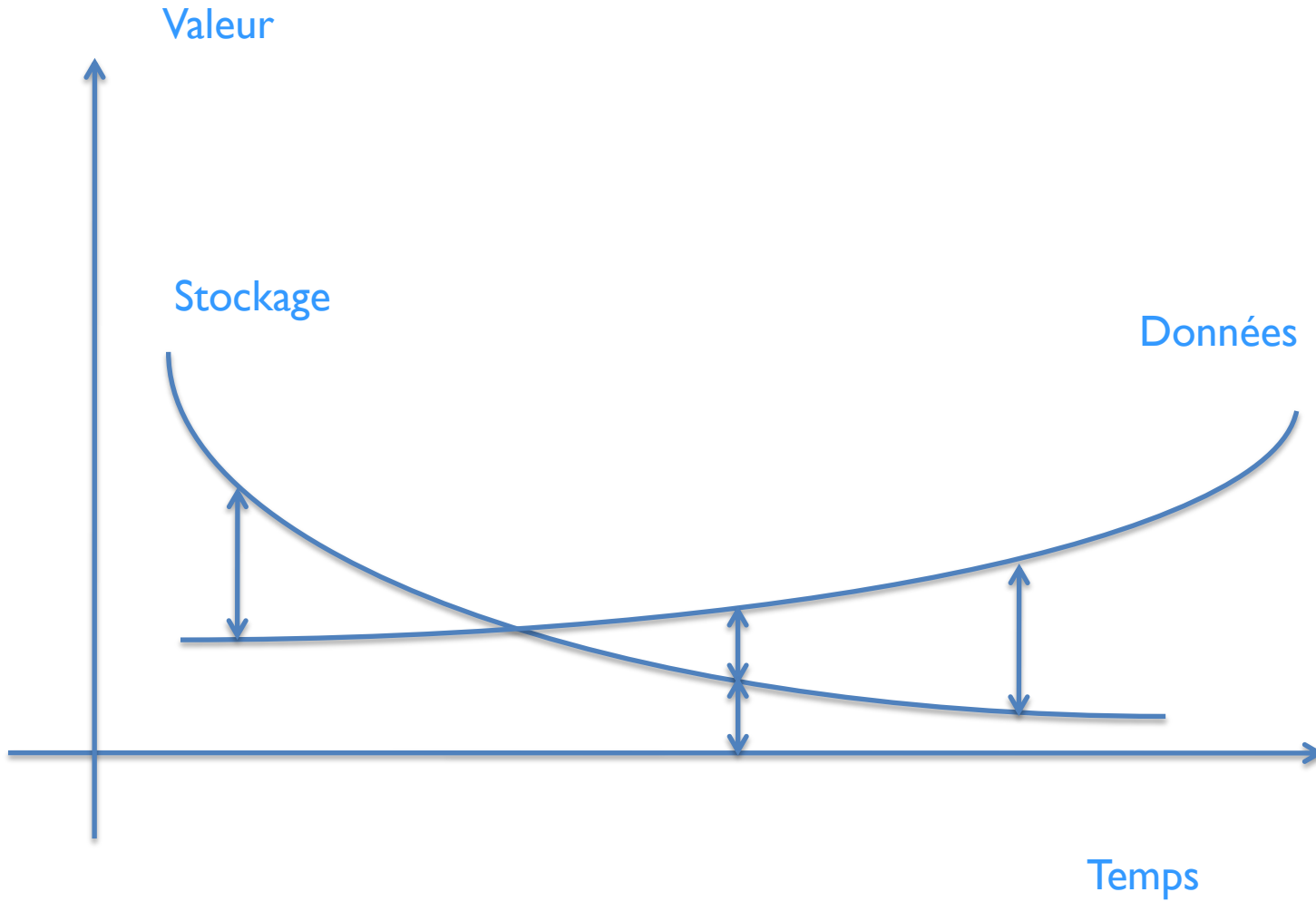


Drew Houston



- A commencé à jouer avec les ordinateurs à l'âge de 5 ans.
- En 2007, Il a 24 ans quand il crée Dropbox
- Ycombinator finança Dropbox en juin 2007, Sequoia en septembre 2007, suivi l'année suivante par \$6M d'Accel et Sequoia
- 9 employés en 2008 (pour 200 000 utilisateurs) et 14 personnes en 2010 pour 2 millions d'utilisateurs
- En 2011, Dropbox a atteint les \$240M de chiffre d'affaires, provenant de seulement 4% de sa base de 50 millions d'utilisateurs. 70 employés et bénéficiaire
- Être profitable n'a pas empêché Dropbox de lever \$250M avec Index, Greylock, Benchmark et les investisseurs déjà présents. Avec une valorisation de \$4B
- Tout fichier déposé sur Dropbox aurait pu appartenir à Dropbox



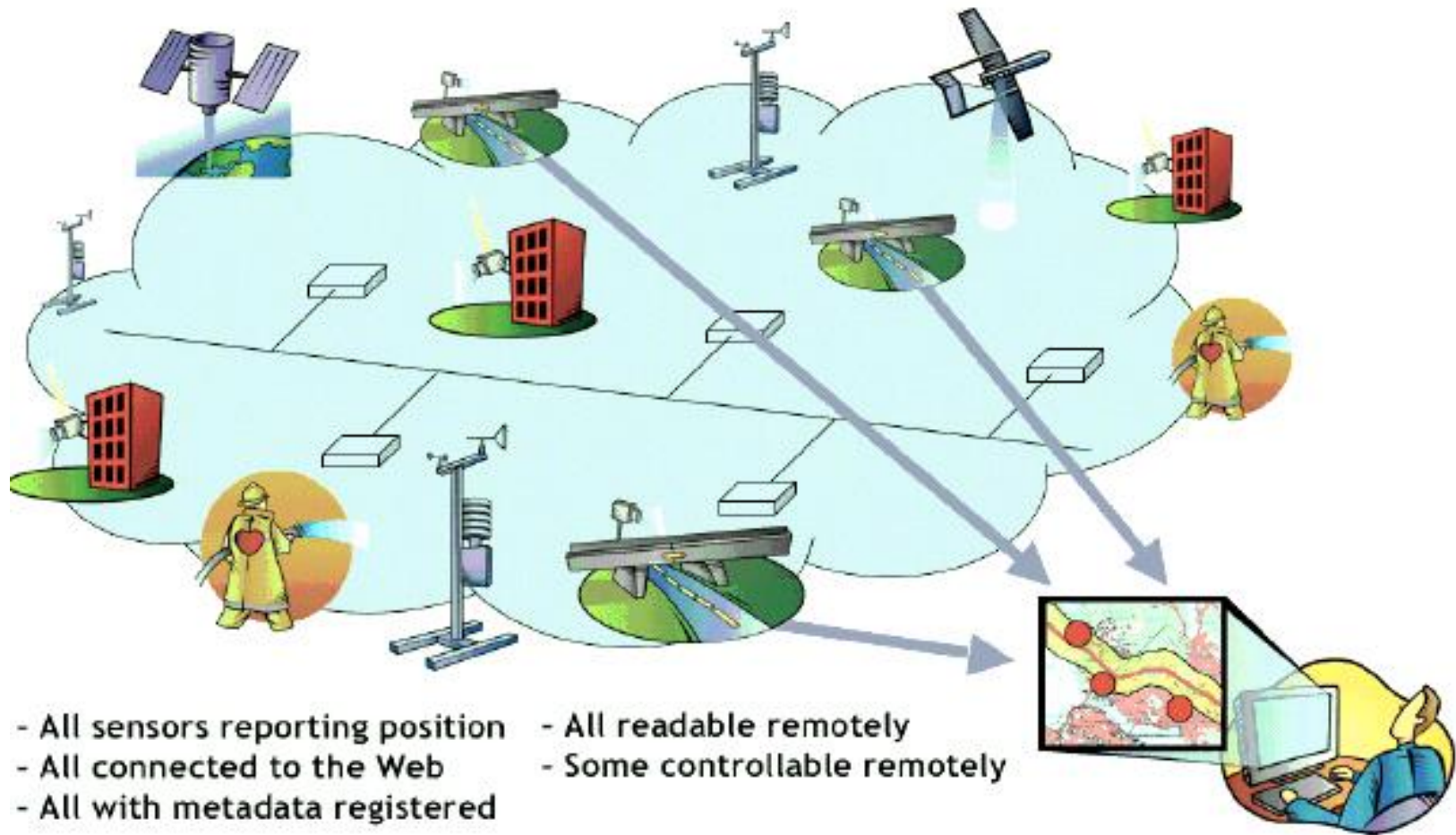


Phase 3 :
The Good,
The Bad,
The Ugly



- Telefónica has provided a welcome boost to the industry by announcing the creation of a new global business unit aimed at deriving revenue from Big Data
- The organisation has launched with one product, Smart Steps, which will use what it promises is fully anonymised and aggregated mobile network data to help companies “to measure, compare and understand” what factors influence the number of people visiting a location at any time.
- Specifically, Telefónica said the insights could help retailers tailor local offerings for existing stores, and determine the best locations and most appropriate formats for new stores
- In addition, Telefónica said market research firm Gfk, with whom it has signed a strategic partnership, would also help with product development – initially in Germany, the UK and Brazil.
- “Unlocking the potential of customer data should provide a brand new source of revenue.”



<http://www.opengeospatial.org/projects/groups/sensorweb>

- Développement des solutions communicantes in-vivo.
- Ces solutions nomades et autonomes implantées directement dans l'organisme permettent d'établir une communication continue entre l'homme et l'animal



- Une start-up propose d'offrir un service de carte bancaire gratuite et monétise les informations des transactions



- **Nous allons tous devenir des potentielles stars d'un Reality Show Mondialisé !!!**



