# Culture numérique

- Enseignement commun
  Université de Caen - 2009-2010
- **Le Web des données**

**http://herve.perso.info.unicaen.fr**


- Hervé Le Crosnier
- Université de Caen
- herve@info.unicaen.fr

# Documents et données

- « Free raw data » Tim Berners Lee
- Contraintes juridiques
- Enjeux économiques et de citoyenneté
- Les objets communiquants
- Données et vie privée

# Des données au document

- Exemple du « passeport »
  - Données conservées par l'administration
  - Document rendu disponible, visible par tous (vous et le douanier,...)
- Du « billet d'avion électronique » à la carte d'embarquement
- Document = forme transmissible des données
  - + interprétation des données

# Structures de données

- Terme informatique pour désigner la façon dont les données sont organisées et peuvent être utilisées par les programmes
- Par extension :
  - Formats de données
  - Interopérabilité
  - Ouverture
- XML : eXtended Markup Language
- RDF : Resource Description Framework

# Capteurs

- La photographie d'un artiste est un document
- Les photographies automatiques prises par les satellites ?
- L'enregistrement d'une interprétation : document
- L'enregistrement des bruits d'une rivière ?
  - Saveoursounds

# Volontaires

- De nombreux internautes souhaitent participer à la construction de lots de données
- Sans revendiquer leur caractère auctorial
  - Wikipedia → ré-organisé en données par DBpedia
  - Openstreetmap (200 000 contributeurs fin 2009

# Free Raw Data

- Un appel de Tim Berners Lee à TED (janvier 2009)
- Web sémantique → web des données
- Les données sont organisées dans un but (par exemple le recensement)
- Les libérer permet de créer de nouveaux usages qui n'entrent pas dans l'orbite des créateurs de données

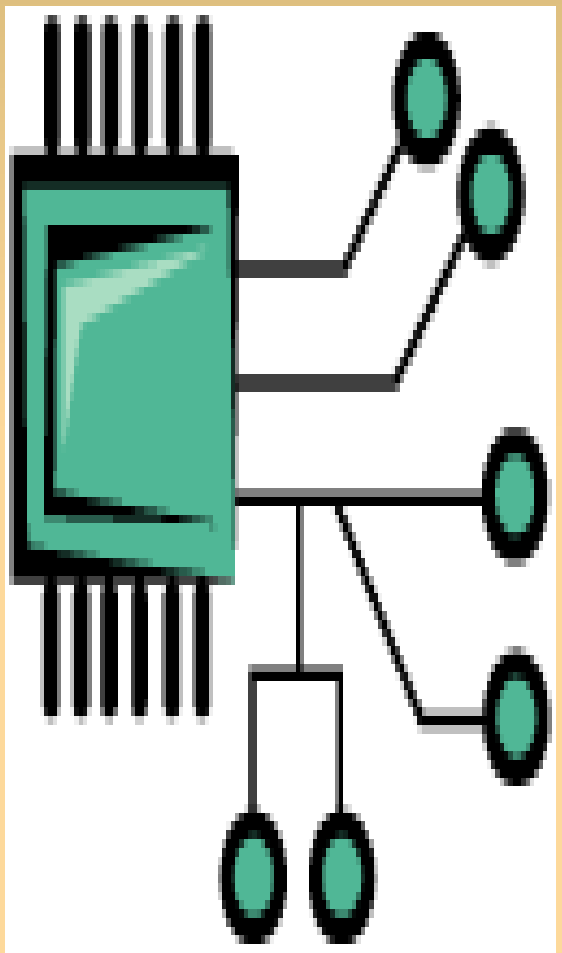
# Données scientifiques

- Publication = document
- Données d'expérience = moyen de faire évoluer les recherches (controverses scientifiques)
- Participe du communalisme scientifique
- Lutte contre la fraude scientifique
- Ou le conflit d'intérêt
  - « We want raw data, now » (BMJ, dec. 2009 à propos Tamiflu)

# Environnement

- 1995 :rapport du Comité sur les données géophysiques et environnementales du NRC (Etats-Unis)
  - Pour comprendre des phénomènes globaux, il faut avoir des données globales
  - Il est plus intéressant pour chaque pays de partager ses données que de collecter partout
  - Pour produire des propositions politiques sur l'environnement, il faut la participation de tous (dès l'origine des données)
- Partage des données climatiques

# Données publiques



- Données produites sur fonds publics
- Afin d'améliorer le bien-être collectif
- Etats-Unis : data.gov
  - Dec 2009 :*Open government directive* de la Maison Blanche
  - 600 banques de données publiques
- Royaume Uni : data.gov.uk
  - Ouverture janvier 2010
- France ?
  - APIE (Agence pour le Patrimoine Immatériel de l'Etat)

# Inventer les usages des données

- Crowdsourcing de l'analyse
  - The Guardian et les données sur les dépenses des députés britanniques
- Mashup
  - Carte de santé, des accidents, des consommations électriques,...
- Nouveaux services
  - Associatifs ou commerciaux

# Demandeurs

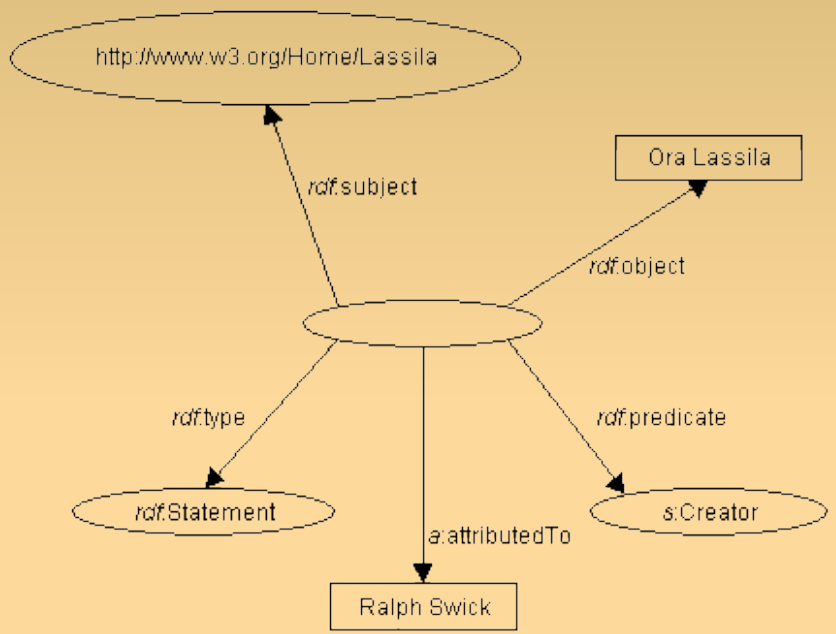
- 5 grands types de promoteurs de l'ouverture des données publiques
  - Aspect « public relation » de l'ouverture
  - Défense de la « transparence » pour un meilleur gouvernement
  - Ceux qui veulent utiliser les données pour améliorer leurs conditions de vie
  - Ceux qui veulent bénéficier de l'argent dépensé en leur nom
  - Ceux qui veulent toutes les données
- Concerne les particuliers et les associations
- … tout comme les entreprises

# Coût de mise en oeuvre

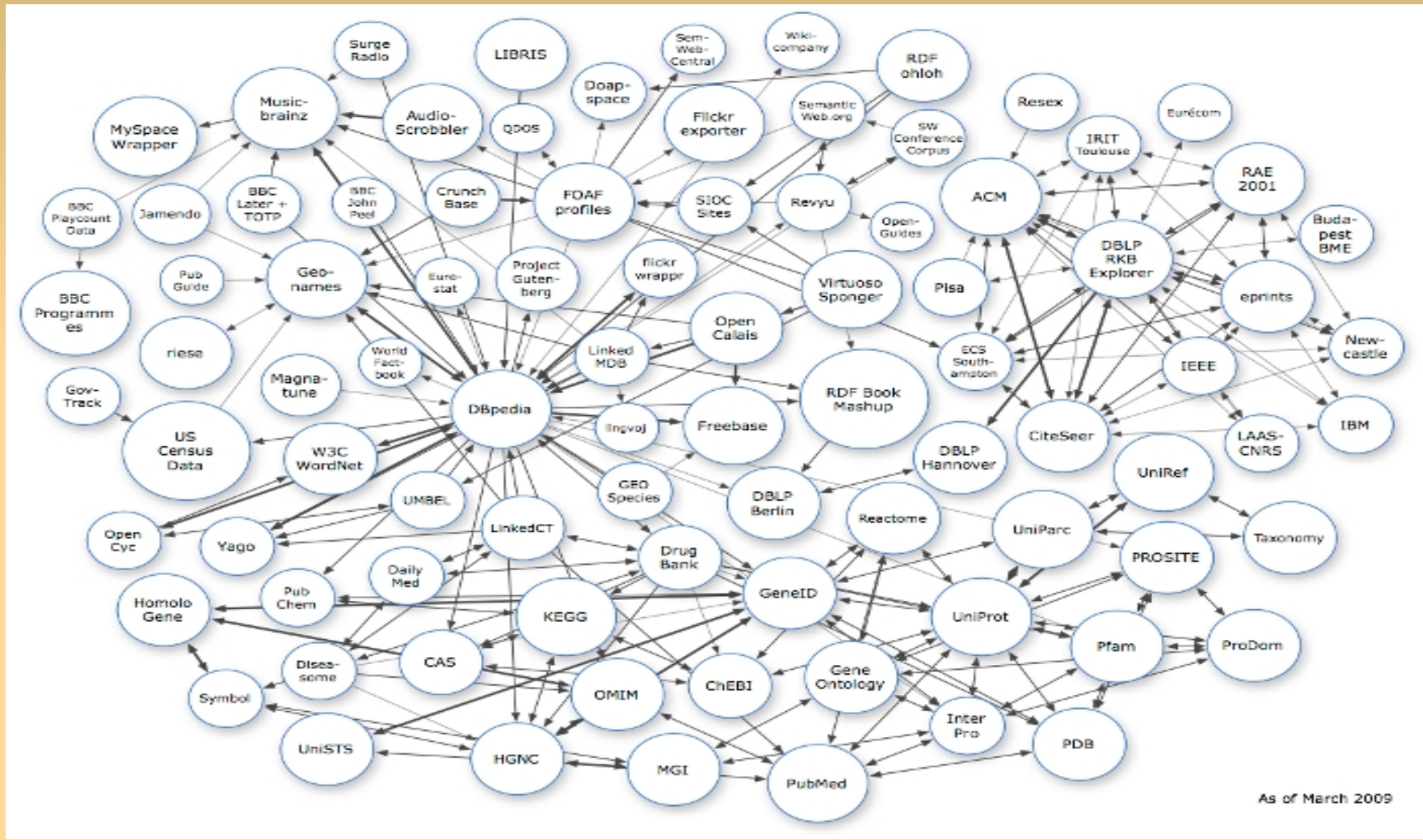
- Ouvrir les données publiques coûte cher
- Transformer les données dans des formats ouverts
- Et réutilisables
- Mise en place d'API pour l'accès
- Ou de possibilité de télécharger les datasets
- → Ouvrir les données les plus utiles
  - Ce qui n'est pas dans la culture des administrations « paternalistes »

# Linked Data

- Principes pour l'interopérabilité
  - Utiliser des URI
  - De type HTTP, avec un serveur délivrant les données
  - Délivrer les données en format RDF
  - Inclure des liens vers d'autres sources de données



# Relier les ressources

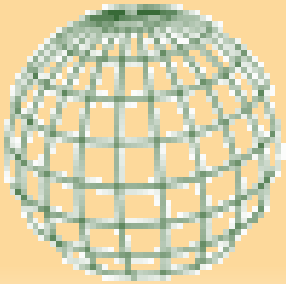


# Quelques exemples

- OpenCalais
  - Reuters :interpétation des textes pour retrouver des lieux ou des événements

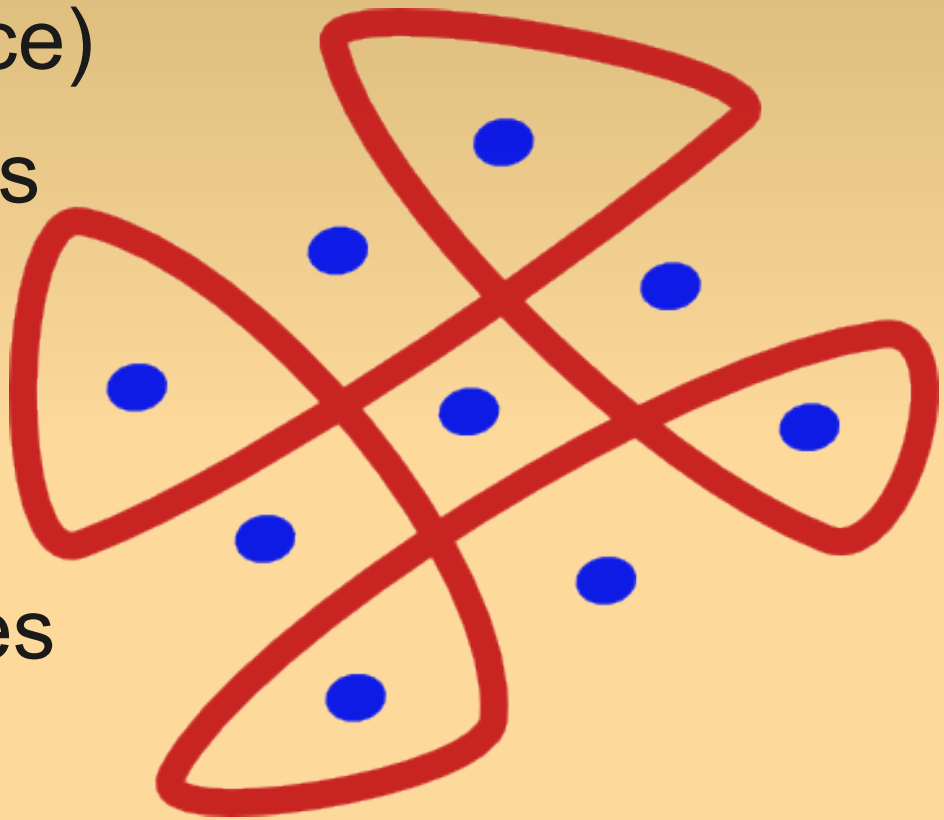
- FreeDB
  - Garantir l'ouverture des données produites par les usagers (output de CDDB → Gracenote)

- Geonames
  - Données statistiques et linguistiques sur les cartes



# Licences pour les données

- Droit de propriété des bases de données (1998 en France)
- Garantir la maintenance des biens communs
  - Dans le temps
  - Face aux récupérations
- Trouver des règles adaptées aux divers usages
  - Réutilisation dans des calculs
  - Présentation des données hors contexte,...



# Exemples

- CC0
  - Étendre les Creative commons
  - Voir aussi le projet ScienceCommons
- Open Knowledge Foundation
  - Public Domain Dedication and Licence (PDDL)
  - Open Database Licence (OdbL)

# Des données aux personnes

- Les traces, volontaires ou non permettent de remonter aux personnes
- Quelle protection des individus ?
  - Niveau d'agrégation des données
  - Suppression de tout identifiant
  - Information sur les conséquences du partage des données