## Lier le Web des données

Le Web actuel est configuré comme une série de données ou ‘documents’. Alors que les documents puisent dans des sources riches en données, celles-ci sont masquées par une page destinée à être vue par des humains. Lors de ce module, nous verrons ce qui se passerait si l’on supprimait toutes les pages et les documents du Web.

Imaginez que vous n’avez que les données brutes, toutes ouvertes, toutes utilisables et toutes liées par un réseau ou ‘Web’ de données.

Ce module traite aussi du Web des données ouvertes et liées (‘ The Web of Linked Open Data’) et explique comment les 5 étoiles du ‘Linked Open Data’ (les données ouvertes et liées) donnent des consignes pour arriver à faire de cette vision une réalité.

##### Dans ce module, nous allons explorer les points suivants :

* Qu’est-ce que le Web des données ?
* Comment utiliser les identifiants du Web
* Ce à quoi ressemble un Web de données ouvertes et liées.

## Comprendre le Web des données

Le Web des données ouvertes et liées change notre façon de parcourir et naviguer l’open data.

### Qu’est-ce que le Web des données ?

Le Web actuel est configuré comme une série de données ou ‘documents’. Ces documents sont crées par des humains avec une volonté de les rendre esthétiques pour que d’autres humains puissent les lire et en retirer de l’information.

Il y a une différence entre les données ‘sur le Web’ et le ‘Web des données’.

Les données ‘sur le Web’ sont des fichiers téléchargeables d’une page Web. Le ‘Web des données’ est une série de ressources de données étroitement liées immédiatement lisibles et compréhensibles par l’humain ou la machine.

### Lier le Web des données

Imaginez que les documents esthétiques soient tous supprimés du Web. Comment vous y retrouveriez-vous ?

Pour naviguer, il vous faut des panneaux vous indiquant votre emplacement actuel par rapport à vos alentours.

Le ‘Web des données ouvertes et liées’ fournit ces panneaux indicateurs en exigeant que toute donnée soit connectée à d’autres données via des liens. Les liens permettent aux humains ainsi qu’aux machines de naviguer entre les points de données.

### Un Web de données ouvertes et liées, pour quoi faire ?

Le Web des données est déjà une réalité dans certaines parties du Web.

De grandes corporations telles que Microsoft ont déjà lancé des recherches sur des outils qui nous permettraient de naviguer le Web de manière complétement différente, en utilisant des données au lieu de pages Web pour se déplacer.

[Take me to Microsoft Pivot talk](http://www.ted.com/talks/gary_flake_is_pivot_a_turning_point_for_web_exploration?language=en)

## Construire un Web des données

Afin de construire un Web des données ouvertes et liées, deux éléments sont nécessaires :

* Des données que l’on peut référencer sur le Web
* Des gens qui fournissent des liens vers ces données à partir de leurs propres données.

Ce sont les deux conditions nécessaires pour satisfaire les exigences des 4ème et 5ème étoiles du ‘Linked Open Data’.

## 4ème étoile : identifiants ouverts

La 4ème étoile traite de la capacité à référencer les choses sur le Web.

### Les identifiants ouverts

Un identifiant est une façon simple et rapide de référencer quelque chose de plus complexe.

Les codes ISBN pour les livres ou les codes postaux pour les lieux en sont des exemples courants.

Cependant, il y a beaucoup d'autres identifiants utilisés par les secteurs public et privé.

La difficulté de nombreux identifiants est qu'il n’est pas toujours évident de savoir à quoi le nombre se réfère. La plupart des gens comptent donc sur les moteurs de recherche pour leur en dire plus quant à ces identifiants. Les identifiants basés sur le Web ne posent pas ce problème car on peut tout simplement cliquer sur l'identifiant pour en savoir plus.

Prenons l’exemple du numéro ISBN d'un livre spécifique qui serait 9781568848624. Pour trouver le livre en utilisant cet ISBN, vous devez taper ce code dans un moteur de recherche pour savoir à quoi l’identifiant fait référence avant de trouver la ressource. En revanche, un identifiant pour ce même livre basé sur le Web vous dira immédiatement qu’il s’agit d’un livre et quel en est le titre, ce qui vous permet de le suivre directement à partir de n’importe quel navigateur Web.

<http://bookshop.blackwell.co.uk/jsp/id/Netscape_and_the_World_Wide_Web_For_Dummies/9781568848624>

### Utiliser la 4ème étoile

La 4ème étoile requiert la création et l’utilisation d’identifiants sur le Web pour toutes choses, qu’elles soient physiques ou numériques.

Actuellement, ces identifiants sont souvent des ensembles de caractères, comme les codes postaux ( ex.: EC2A 4JE) ou simplement des chiffres (08030289). Ces identifiants sur le Web permettent l’utilisation du service de résolution de noms HTTP pour rechercher ces choses ou même en obtenir une copie.

<http://id.southampton.ac.uk/building/59>

<http://data.ordnancesurvey.co.uk/id/postcodeunit/EC2A4JE>

### L’open data de la 4ème étoile

La 4ème étoile exige aussi que si un utilisateur clique sur un lien qui ne peut pas lui apporter la chose exacte (un livre, un objet, une personne ou un bâtiment par exemple), il reçoive de l’information contextuelle au sujet de cette chose.

Ce qui arrive souvent est que le client (humain ou machine) est redirigé vers une représentation appropriée de l’objet. Il peut s’agir d’une page Web, ou encore de données brutes sur cette page Web.

### À vous !

Certains identifiants basés sur le Web sont [cool.](http://www.w3.org/TR/cooluris/)

Une URI ‘cool’ indique au lecteur ce à quoi l’identifiant fait référence grâce au texte de l’URL. Essayez de deviner à quel identifiant le lien ci-dessous fait référence :

<http://data.ordnancesurvey.co.uk/id/postcodeunit/SO171BJ>

Notez qu'une fois chargée, l’adresse de la page est différente de l'exemple ci-dessus.

L’exemple du lien ci-dessus envoie une demande au serveur pour un objet physique. Pour satisfaire cette demande, le serveur a renvoyé une page Web sur l’objet physique (comme il ne pouvait pas vous envoyer l’objet lui-même).

Votre navigateur s’est soumis à un processus appelé ‘négociation de contenu avec le serveur’ : ce que vous obtenez est ce qui se rapproche le plus de ce que le serveur devait vous donner - dans ce cas il s’agit de données au sujet de l’objet sous une forme spécifique au Web. Si un autre format correspond mieux à vos besoins, vous pouvez utiliser des extensions pour poursuivre la négociation de contenu avec le serveur.

Pourquoi ne pas tenter de négocier du contenu avec le lien vers Ordnance Survey ci-dessus en ajoutant des extensions comme '.json' et '.xml' à la fin de l'URL?

## 5ème étoile : ‘Linked Data’ (données liées)

Ici, il s’agit de lier des points individuels de données entre eux.

### Le lien

La 5ème étoile introduit le concept de lier des points de données entre eux (le concept du Web de données).

Ces liens permettent aux humains et aux machines de naviguer entre les sources de données sur le Web.

Le fait de lier des points de données directement les uns aux autres permet de diriger la personne ou la machine vers la bonne information au sujet de la donnée, alors que sinon chacun est responsable de la maintenance de ses propres information sur ces données. Le Web des données fonctionne comme un site officiel, en donnant à tous l’accès à une ressource faisant autorité sur le sujet. Le Web des données permet d’éviter de répéter des données vu qu’on partage une seule ressource parmi un nombre infini de séries de données grâce à des URIs.

### Utiliser la 5ème étoile

Exemple - Un jeu de données sur tous les bâtiments d’une entreprise utilisera souvent un code postal pour chaque bâtiment de son bloc de données:

Le nom du bâtiment : EEE Building

Le numéro du bâtiment : 32

Le code postal du bâtiment : SO17 1BJ

Ajoutez du Web des données :

Le nom du bâtiment : EEE Building

Le numéro du bâtiment : 32

Le code postal du bâtiment :<http://data.ordnancesurvey.co.uk/id/postcodeunit/SO171BJ>

### Essayez d’utiliser la 5ème étoile

Vous aurez peut-être constaté certains liens entre les données explorées jusqu’à présent dans ce module.

Essayez de charger le lien utilisé plus haut :

<http://data.ordnancesurvey.co.uk/id/postcodeunit/SO171BJ>

Maintenant, chargez ce même lien dans un navigateur de données :

[Lien vers Graphile linked data browser](http://graphite.ecs.soton.ac.uk/browser/)

Le navigateur de données fait automatiquement une négociation de contenu des données (pas de contenu de page Web) et affiche les données brutes sous un format lisible par l’humain.

Tout texte en rouge dans le navigateur est un lien dans les données, qui devrait mener vers plus de données.

Utilisez l'URL suivante pour voir ce que vous pouvez découvrir au sein du Web des données ouvertes et liées.

<http://www.bbc.co.uk/programmes/b006q2x0>

## Êtes-vous prêt(e) à surfer le Web des données ?

Il y a quatre éléments utiles à prendre en compte lorsqu’on surfe le Web des données ouvertes et liées. Saurez-vous vous en souvenir ?

## Qu’est-ce que le Web des données ?

Le Web des données, c’est...

* une série de ressources de données interconnectées dans le Web.
* un Web de portails de données où on peut télécharger des données.
* un nouveau réseau social.

##### Vouz avez raison!

Le Web des données est une série de ressources de données interconnectées immédiatement lisibles et compréhensibles par les humains ou les machines.

##### Êtes vous sûr(e) ?

Le Web des données est une série de ressources de données interconnectées immédiatement lisibles et compréhensibles par les humains ou les machines.

## Qu’identifie le lien suivant?

<http://id.southampton.ac.uk/room/32-3077>

* Une page Web
* un bâtiment
* une pièce

##### Vouz avez raison!

Rappelez-vous qu’un identifiant basé sur le Web indique la chose physique, et non pas la page Web où elle se trouve.

##### Êtes vous sûr(e) ?

Rappelez-vous qu’un identifiant basé sur le Web indique la chose physique, et non pas la page Web où elle se trouve.

## Lesquels des identifiants suivants font référence au l’artiste Taylor Swift?

* <http://www.bbc.co.uk/music/artists/20244d07-534f-4eff-b4d4-930878889970>
* <https://musicbrainz.org/artist/20244d07-534f-4eff-b4d4-930878889970>
* <http://dbpedia.org/resource/Taylor_Swift>

##### Vouz avez raison!

En réalité, tous trois sont des identifiants. Les deux derniers exemples utilisent le même sous identifiant, mais dépendent d’autorités différentes. Le troisième est l’identifiant wikipedia (de dbpedia) de l'artiste. Ils sont tous liés les uns aux autres et partagent leurs données. Saurez-vous trouver les liens ?

##### Êtes vous sûr(e) ?

En réalité, tous trois sont des identifiants. Les deux derniers exemples utilisent le même sous identifiant, mais dépendent d’autorités différentes. Le troisième est l’identifiant wikipedia (de dbpedia) de l'artiste. Ils sont tous liés les uns aux autres et partagent leurs données. Saurez-vous trouver les liens ?

## Surfer le Web des données ouvertes et liée

Le Web des données est une série de ressources de données étroitement liées qui peuvent être lues et comprises instantanément par les humains ou les machines.

En 1989 Sir Tim Berners-Lee a inventé le Web des documents et nous avons mis plus de 20 ans à comprendre comment l’utiliser. Il fait à présent partie intégrante de notre société.

En 2001, Tim a envisagé le Web des données ouvertes et liées et nous commençons petit à petit à réaliser de quelle façon cela va à nouveau révolutionner le Web.