

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOULOUD MAMMARI DE TIZI-OUZOU



FACULTE DU GENIE ELECTRIQUE ET D'INFORMATIQUE
DEPARTEMENT D'INFORMATIQUE

Mémoire de Fin d'Etudes de MASTER PROFESSIONNEL

Domaine : **Mathématiques et Informatique**

Filière : **Informatique**

Spécialité : **Ingénierie des systèmes
d'informations**

Présenté par

MANSOURA Hacene

MEGHNINE Kamel

Thème

**Conception d'une collection d'open data basée
sur une ontologie disciplinaire**

Mémoire soutenu publiquement le 30/06/2016 devant le jury composé de :

Président : M^{er} HABET Mohammed-Saïd

Encadreur : M^{me} BOUARAB Farida

Examineur : M^{er} CHEBOUBA Lokmane

Examineur : M^{er} DIB Ahmed

Remerciement

Nous tenons à remercier dans un premier temps notre encadreur : Mme. Dahmani Bourab pour son soutien et ses conseils durant toute cette année.

Nous tenons également à remercier tous les enseignants qui ont contribué à notre formation.

Nous tenons à remercier les membres du Jury pour avoir accepté d'examiner ce modeste travail.

Nous aimerions aussi remercier nos familles, nos amis, ainsi tous ceux qui ont contribué de prêt ou de loin à la réalisation de ce travail.

Sommaire

Résumé

Organisation du mémoire

Liste des Figures

Liste des Tableaux

Introduction générale

Chapitre 01 : Etude bibliographique2

Introduction 2

1. Origine du web sémantique 3

2. Linked Open data 4

2.1. Linked data..... 4

2.2. Open data 5

3. Les principes du Linked Open data..... 7

3.1. Rôle des URIs dans Linked Open data 8

3.2. Rendre les URIs déréférencables 10

3.3. Les URIs 303 11

4. Topologie du web des données..... 11

4.1. Données multi-domaines 12

5. les bases principales..... 12

5.1. L'architecture web 12

5.1.1 Les Ressource 12

5.1.1.1. Ressource Informationnelle 13

5.1.1.2. Ressource non-Informationnelle 13

5.1.2 Identification de la ressource 13

5.1.3 Représentation de la ressource 13

5.1.4	Négociation du contenu	14
5.1.5	URI alias	15
5.2.	Le modele RDF	16
5.2.1	Les litteraux RDF	16
5.2.2	Les Liens RDF	16
6.	Bien choisir un URI	20
7.	Choix de vocabulaire pour représenter une information	21
7.1.	La réutilisation des termes existants	21
7.2.	Définir un terme	23
8.	La description RDF	24
8.1.	Description autorisé	24
8.2.	Description non-autorisé	25
8.3.	RDFa	26
9.	Définir un lien RDF vers un dataset	26
9.1.	Définir les liens RDF manuellement.....	27
9.2.	Générer les liens RDF automatiquement	27
9.2.1	Algorithmes simples	28
9.2.2	Algorithmes complexes.....	29
10.	Publier les données en Linked data.....	29
10.1.	Utilisation des fichiers RDF statiques.....	30
10.2.	Utilisation des base de données relationnelles... Erreur ! Signet non défini.	
10.3.	Utilisation d'autre type d'informations..... Erreur ! Signet non défini.	
10.4.	Création des wappers autours des APIs	34
10.5.	Choisir une méthode adéquate pour la publicationde Linked Open data ...	35

Chapitre 02 : L'utilisation des LOD en milieu socio-économique et en éducation.....37

Introduction	37
1. Utilisation des LOD en milieu socio-économique	37
1.1. L’Open data au service de la démocratie	37
1.2. L’Open data au service d’optimisation des villes	38
2. Utilisation des LOD en milieu éducation.....	39
2.1. L’Open data au service du défusion de savoirs.....	39
2.2. Interet de Linked Open data sur la recherche d’informations.....	39
2.2.1. Une meilleures interopérabilité des données de la bibliothèque.....	40
2.2.2. Une meilleure possibilité de recherche	40
2.2.3. Une meilleur sérendipité	42
2.2.4. Une meilleure visibilité de données sur le web.....	43
Chapitre 03: Conception d’un LOD disciplinaire basé sur l’ontologie de domaine ONTO-TDM	45
Introduction	45
1. Ressource pédagogique	45
1.1. Quelques définitions de ressources pédagogiques	46
2. Objet pédagogique.....	46
2.1. Quelques définitions	47
3. Les ressources pédagogiques face à l’apprentissage en ligne	48
3.1. L’apprentissage en ligne	49
4. les normes d’apprentissage en ligne.....	49
4.1. LOM.....	50
4.2. SCORM.....	51
4.3. IMS-LD.....	52
5. L’apprentissage en ligne et les ontologies.....	53
5.1. Représentation des ressources pédagogiques en se basant sur Ont-TDM ..	54

**Chapitre 04:Implimentation d'un prototype d'un site LOD pour les ressources
disciplinaire.....56**

Introduction 56

1. Planification stratégique 56

1.1. Objectif du projet 56

1.2. Connaitre son environnement 56

1.2.1. Notre plateforme face aux données du département informatique 56

1.2.2. Les enseignants du dépatement informatique de l'UMMTO..... 57

2. Conception..... 57

2.1. Présentation des fonctions de base de la plateforme..... 57

2.2. Architecture de l'information 58

2.2.1. Organisation du contenu 58

2.2.2. Plan de la plate forme 62

2.2.3. Conception de la base de données 62

2.2.4. Conception du générateur de dataset disciplinaire 65

3. Le Zoning de la plateforme..... 67

3.1. Zoning de la première page maître 67

3.2. Zoning de la deuxième page maître 67

4. Représentation graphique de la démarche de modalisation du site 69

4.1. Spécification des besoins & cas d'utilisation..... 69

4.1.1. Identification des acteurs 69

4.1.2. Identification des cas d'utilisations 69

4.1.3. Identification des scénarios..... 70

4.1.4. Diagramme de cas d'utilisation 72

Chapitre 04:Réalisation.....74

1. Environnement de développement..... 74

1.1.	Microsoft Visual Studio	74
1.2.	Internet Information Services.....	75
2.	Les langages de programmations	75
2.1.	Le langage XHTML (Extensible HyperText Markup Language)	75
2.2.	RDFa	76
2.3.	Le langage de script C#.....	76
2.4.	JavaScript.....	76
2.5.	Requêtes SQL	76
2.6.	CSS	77
3.	Présentation des interfaces	77
3.1.	Page d'accueil	77
3.2.	Page de consultation des datasets	78
3.3.	Page de consultation des ressources	78
3.4.	Page d'authentification	79
3.5.	Page d'inscription	79
3.6.	Page d'accueil dans l'espace enseignant.....	80
3.7.	Page de création d'un dataset.....	80
3.8.	Page de publication d'une ressource	81

Conclusion générale

Références bibliographiques

Références webographies

Liste des Figures

Figure 01 : Les cinq étoiles de LOD	5
Figure 02 : les URI sont utilisées pour identifier des gens et les relations qui les joignent.....	9
Figure 03 : Statistiques des jeux de données du CKAN	12
Figure 04 : Recherche géographique sur le site data.bnf.fr.....	41
Figure 05 : Recherche de relation entre deux personnes	42
Figure 06 : Une seule interface pour les ressources variées du Centre Pompidou, ici sur Max Ernst.....	43
Figure 07 : Recherche par type d'entité	44
Figure 08 : La pyramide des objets pédagogiques	48
Figure 09 : Décomposition des notions du domaine d'enseignement	55
Figure 10 : Enumération du contenu	58
Figure 11 : Catégorisation du contenu	59
Figure 12 : Structure méta data du data set	59
Figure 13 : Structure externe du contenu	60
Figure 14 : Structure interne du contenu	61
Figure 15 : Plan de la plateforme	62
Figure 16 : Entité Association.....	64
Figure 17 : Représentation des métadonnées d'un dataset	65
Figure 18 : Fonctionnement de générateur de dataset.....	66
Figure 19 : Le zoning de la première page maître	67
Figure 20 : Le zoning de la deuxième page maître	68
Figure 21 : Diagramme de cas d'utilisation « visiteur »	72
Figure 22 : Diagramme de cas d'utilisation « enseignant »	73
Figure 23 : Visual studio 2012	74

Figure 24 : Serveur web IIS	75
Figure 25 : Page d'accueil	77
Figure 26 : Page consultation des datasets	78
Figure 27 : Page de consultation des ressources	78
Figure 28 : Page d'authentification	79
Figure 29 : Page d'inscription	79
Figure 30 : Page d'accueil dans l'espace enseignant	80
Figure 31 : Page de création d'un dataset	80
Figure 32 : Page de Publication d'une ressource	81

Liste des Tableaux

Tableau 01 : Présentation du dictionnaire de données	63
Tableau 02 : Présentation du dictionnaire de métadonnées	63
Tableau 03 : Spécification des tâches	69
Tableau 04 : Les scénarios de visiteur	70
Tableau 05 : Les scénarios de l'enseignant	71

Résumé :

Ces dernières années, une nouvelle vision et des nouveaux défis du web sont apparus. Parmi ces nouvelles visions on cite le web des données et le web des données ouvertes (OPEN DATA et LINKED OPEN DATA). Pour appliquer ces deux technologies, Les utilisateurs du web doivent suivre et respecter les différentes exigences et les nouveaux concepts que nous avons bien détaillés dans notre mémoire.

Dans notre travail, nous avons appliqué ces deux dernières technologies pour lier et ouvrir en public les données disciplinaires du département informatique de l'université de Mouloud MAMMERI de Tizi ousou. Pour ce faire, nous avons utilisé une ontologie de domaine disciplinaire pour créer des métas données et le RDF pour la création des triplets des ressources.

Mots clé : Open data, Linked data, Onto-TDM, RDfa, ressources disciplinaires.

Introduction générale :

Web sémantique et web des données représentent une nouvelle vision du web reposant sur un ensemble de techniques. Proposé par Tim-Berners Lee et le W3C (l'organisme qui spécifie l'architecture du web), le web sémantique n'a pas pour objectif de remplacer le web actuel, mais d'en étendre les possibilités. Il préconise l'indexation des pages du web et l'extraction de la sémantique à l'aide de différentes techniques dont la base est le RDF (Resource Description Framework). Toute ressource du web peut être décrite à l'aide de séries de triplets composés d'un sujet, d'un prédicat et d'un objet. Chaque élément du triplet est identifié à l'aide d'un URL du type URI permettant une identification unique et pérenne. Des ontologies (ou vocabulaires) permettent de catégoriser les sujets, prédicats et objets.

Le web sémantique vise à mettre en place une indexation des pages du web, mais aussi potentiellement des concepts absents du web mais dont une description sur le web est possible (personne, lieu, date, élément abstrait, objet physique). Le même objet peut être vu au format HTML adapté à la lecture, ou bien en RDF, le format du web sémantique dérivé de XML qui contient les listes de triplets relatifs à l'objet. Si les logiciels du web 2.0 (gestionnaires de contenu, wikis, plateformes d'archivage) permettent déjà l'affichage des métadonnées en HTML, RDF ou RDFa (une combinaison de HTML et RDF), les nouveaux outils intègrent mieux les standards du domaine que sont RDF, RDFS, OWL, SKOS, SPARQL, RDFa, etc.

Le web sémantique intéresse particulièrement les acteurs du web et les organismes de grande taille confrontés à la multitude et à la diversité des données. Les moteurs de recherche comme Google, Yahoo et Bing sont à l'origine de schema.org, une spécification permettant d'introduire des informations sémantiques dans le code HTML. Par ailleurs, Google tire parti de "Knowledge Graph", une base de connaissance multilingue résultant de la compilation de plusieurs bases du web sémantique, pour afficher des données factuelles lors d'une recherche par exemple. Les gouvernements, les villes, les universités, les bibliothèques, les archives, les encyclopédies, les entreprises, la recherche, mais aussi certaines unités de recherche de taille plus modeste s'impliquent. L'un des intérêts majeur du web sémantique est de rendre possible la liaison entre objets et données d'origine différente et la réutilisation dans de nouveaux services. Le web sémantique est alors appelé "web des données" ou "web des données liées". Si les données sont protégées avec une licence ouverte (Creative Commons, Etalab, Domaine public, etc), on parle de "données ouvertes" ou "open data".

Notre projet de fin d'étude porte sur l'application de ces nouvelles visions de web « open data » afin de lier les bases de données disciplinaires de département informatique de l'université de Mouloud MAMMRI de Tizi-Ouzou et de les rendre accessibles au public. Ce choix de thème est motivé en premier lieu par l'importance de ces technologies de web dans une économie développée basée sur la possibilité de l'accès aux bases de données publiques par tous les acteurs économiques en temps réel et en deuxième lieu par le rôle de ces nouvelles technologies pour simplifier la vie d'un citoyen. Pour le choix de l'application de ces nouvelles versions de web sur les bases de données de département informatique de l'université de Mouloud MAMMRI de Tizi-Ouzou est justifié par notre appartenance à ce département et à la facilité d'accès aux bases de données, contrairement aux autres entreprises privées ou publiques que nous avons consultées afin de nous communiquer leurs bases de données. Donc afin de proposer une solution et les différentes interfaces de notre plateforme d'étude, notre projet est articulé sur cinq chapitres :

Le premier chapitre : Dans ce chapitre nous avons parlé sur l'historique et l'apparition de ce nouveau concept du web, nous avons défini ses notions de bases, nous avons parlé de ses principes, ses bases et les technologies utilisées pour bien respecter et suivre ses principes. Nous avons ainsi spécifié les ontologies utilisées dans le linked open data.

Le deuxième chapitre : Dans le deuxième chapitre nous avons parlé de Linked open data et son impacte sur les différents services sociaux, économiques, sur le secteur éducatif et sur la démocratie.

Le troisième chapitre : ce chapitre est consacré sur les ressources disciplinaires et leurs organisations sur le web, l'ontologie du domaine disciplinaire et les différentes normes d'apprentissage utilisées sur le web.

Le quatrième chapitre : Dans ce chapitre nous avons présenté les différentes techniques et spécifications que nous avons utilisées pour la conception d'un prototype de solution qui se base sur les principes du Linked open data définis précédemment pour la gestion et l'organisation des ressources du domaine disciplinaire du département informatique de l'université mouloud Mammeri de Tizi-Ouzou.

Le cinquième chapitre : Dans ce chapitre nous avons représenté les outils et les technologies utilisées pour l'implémentation de la solution et les différentes interfaces de notre plateforme.

Introduction :

Linked data permet aux utilisateurs de publier des données structurées sur le web aussi facile que publier un document. Cette nouvelle technologie consiste à publier et lier les données structurées sur le web. Le but derrière le linked data c'est pouvoir lier les données entre elles et faciliter l'accès aux ressources recherchées.

Les bases principales de linked data :

- Utilisation du modèle rdf pour publier les données structurées sur le web.
- Utilisation des liens rdf pour lier les données de différentes sources.

L'utilisation des deux principes nous mène à créer des données communes sur le web, un espace où un internaute ou une organisation peut publier et utiliser les données existantes. On appelle aussi cette nouvelle technologie le web des données ou le web sémantique. On accède au web des données en utilisant des navigateurs et des moteurs de recherche linked data comme on utilise les moteurs de recherche du web documentaire qui utilisent le HTML. Les moteurs de recherche linked data permettent aux utilisateurs de naviguer sur les différentes sources de données en suivant les liens rdf, cela permet à un internaute de faire sa recherche sur un dataset puis aller jusqu'à accéder aux données trouvées sur d'autres datasets grâce aux liens existants entre datasets. Par exemple un utilisateur de linked data fait une recherche à propos d'une personne sur un dataset X, et l'utilisateur s'intéresse sur la ville où la personne habite, en suivant les liens on peut accéder à des informations enregistrées sur un dataset Y à propos de la ville.

Les liens RDF sont des relations entre les données, et ces dernières peuvent être de différents types, par exemple un lien connectant deux personnes peut être une relation de connaissance, et entre un livre et une personne peut être que la personne est l'auteur du livre. Pour publier et structurer les données sur le web sous forme de l'open linked data il faut suivre quelques étapes :

- Respecter et appliquer les principes de linked data.
- Choisir une bonne méthode d'identification de données par URI
- Choisir un vocabulaire connu et un espace de noms adéquat pour bien représenter les informations

- Spécifier quelles informations sera incluses dans la description rdf pour qu'elle soit publiée sur le web.
- Choisir une bonne approche pour générer les liens rdf entre les données de différentes sources.
- Exposer les différentes manières de publication des différentes type d'informations en utilisant les outils de linked data et choisir la plus adéquate.
- Faire des teste unitaire pour les linked data source qu'on veut publier.

1. Origine du web sémantique

Le web sémantique tire ses origines de travaux des années 1990 portant sur la gestion des connaissances. Mais il ne commence à se concrétiser qu'en 1999 avec les recommandations du W3C portant sur les "langages" RDF et RDFS, et avec la publication en 2007 par DBpedia des premières "données ouvertes liées". Les bases de données relationnelles transformées en données stockées dans des "Triple Store" sont publiées en nombre croissant. La liaison des données permet la navigation et le partage des informations entre sites différents. On assiste ainsi à la mise en place progressive d'un réseau dont l'importance est mondiale. Une vue de Septembre 2011 permet de visualiser les principales données ouvertes liées (Linked Open Data ou LOD) et les relations entre réservoirs de données, dessinant le nuage des données ouvertes liées (Lod Cloud). Des données encyclopédiques multidisciplinaires (DBpedia, Freebase ou YAGO) constituent le cœur du nuage. DBpedia provient de l'extraction régulière des données de Wikipedia. D'autres données concernent la géographie, les objets des bibliothèques, des universités et de la recherche, des archives, les personnes des réseaux sociaux, les médias (musiques et films), la météorologie, les cartes (OpenStreetMap), les données gouvernementales (résultats électoraux, recensements, statistiques diverses), etc. Certains sites web et applications pour mobiles tirent parti des données ouvertes liées pour offrir des services à haute valeur ajoutée, croisant les sources et les données entre elles. On pense que des services bâtis sur ces données librement réutilisables conduiront à l'émergence d'un nouveau secteur économique s'appuyant principalement sur les tablettes et les smartphones.

2. Linked Open Data

Le terme Linked Open Data résulte de la fusion de deux concepts distincts : Linked Data et Open Data. Nous proposons ci-dessous une définition de ces deux concepts.

2.1 Linked data

Linked Data est un terme créé en 2006 par Tim Berners-Lee dans le contexte du web sémantique. L'objectif du web sémantique est de briser les silos de données sur le web et de les rendre interopérables (**Byrne, 2010**). Le web *classique* non sémantique se base sur un réseau hypertexte dans lequel des documents, les pages web, sont liés entre eux. Le web sémantique quant à lui veut créer un nouveau réseau, mais au sein duquel les données sont reliées entre elles, en plus des pages web. Pour obtenir des données liées ou des Linked Data, quatre principes doivent être respectés (**Berners-Lee, 2009**) :

1. Utiliser des adresses URI unique pour identifier les choses.
2. Utiliser des adresses URI HTTP qui existent sur le Web. Une Erreur HTTP 404 indique que l'URI utilisé est peu fiable et ne doit pas être réutilisé pour décrire d'autres données ;
3. Fournir à travers l'URI des renseignements lisibles par les humains et par les machines. En utilisant le mécanisme de redirection HTTP (code 302) et la variable User Agent contenue dans les entêtes des requêtes HTTP, un serveur peut afficher une page en XML ou RDF pour une machine ou une page HTML pour le navigateur d'une personne ;
4. Ajouter des URI externes aux données pour améliorer la découverte d'autres informations sur le Web.



Figure 01 : Les cinq étoiles de LOD

En d'autres termes, il faut donc attribuer des identifiants aux ressources sous forme d'adresses URI HTTP, les décrire en utilisant des standards du web sémantique et créer des liens vers des données externes pour créer un réseau. Si plusieurs acteurs du web utilisent les mêmes standards et les mêmes modèles, il devient plus facile d'échanger, de réutiliser et de lier les données. Mais ce scénario n'est possible qu'à condition que les données soient disponibles sous licence ouverte. C'est pour cela que les Linked Data à elles seules ne suffisent pas dans l'optique du web sémantique : elles doivent également être des Open Data, ou données ouvertes en français.

2.2 Open Data

Open Data est défini ainsi par l'Open Knowledge Foundation (2012) :

« Une donnée ouverte est une donnée qui peut être librement utilisée, réutilisée et redistribuée par quiconque - sujette seulement, au plus, à une exigence d'attribution et de partage à l'identique. »

Par déduction, cette définition stipule donc qu'une donnée ouverte doit pouvoir être réutilisée à des fins commerciales également.

De manière générale, pour faciliter la réutilisation des données, trois aspects doivent être pris en compte :

1. Aspect technique : un format le plus ouvert possible
2. Aspect juridique : une licence ouverte
3. Aspect économique : des redevances nulles ou limitées

L'Open Data s'inscrit dans une mouvance plus générale d'ouverture des connaissances, incluant entre autres l'Open Source (pour les logiciels), l'Open Research Data ou encore l'Open Access (pour les ressources d'information). Le concept inclut par ailleurs l'Open Government Data, qui concerne les données produites par des institutions de droit public.

L'interopérabilité technique et la capacité sémantique des Linked Data, couplées à la libre utilisation juridique et à la gratuité de l'Open Data, forment la puissance du concept de Linked Open Data. L'aspect ouvert des données permet ainsi au web sémantique de se populariser et facilite la création de données liées en donnant accès à des jeux de données externes de manière libre.

L'idée consiste à appliquer l'architecture du World Wide Web [**Jacobs,2004**] pour partager des données structurées à une échelle globale. Afin de comprendre ces principes, il est au préalable nécessaire de comprendre l'architecture du document Web classique.

Le Web est construit sur un ensemble de standards simples :

1. des URI (Uniform Resource Identifiers, identifiants de ressource uniformes) comme mécanisme d'identification unique et global [**Berners-Lee, 1998**].
2. HTTP (HyperText Transfer Protocol, protocole de transfert hypertexte), le mécanisme d'accès universel [**Fielding,1999**];
3. HTML (HyperText Markup Language, langage de balisage hypertexte), le format de contenu largement utilisé [**Raggett,1999**].

De plus, il s'appuie sur le principe de liens existant entre des documents pouvant résider sur des serveurs différents. Les hyperliens permettent aux internautes de naviguer entre les serveurs et aux moteurs de recherche d'explorer le Web afin de fournir des fonctionnalités

sophistiquées a partir du contenu explore. C'est pourquoi ces hyperliens sont centraux dans la connexion du contenu de différents serveurs afin de créer un *espace d'information* unique et global. En combinant simplicité, décentralisation et ouverture, le Web semble atteindre l'architecture idéale, ce que sa croissance de ces vingt dernières années ne cesse de souligner.

Les données liées se fondent sur l'architecture du Web et l'appliquent au partage de données à l'échelle globale.

3. Les principes du Linked Open Data

Le premier principe des données liées recommande d'utiliser des références URI pour identifier non seulement des documents et du contenu digital, mais aussi des objets réels et des concepts abstraits. On peut y inclure des éléments tangibles (des gens, des endroits, des voitures) ou plus abstraits (les relations comme *connaître quelqu'un*, l'ensemble des voitures vertes dans le monde, voire la couleur verte elle-même). Cela peut être vu comme une extension des principes du Web pour comprendre tout objet ou concept.

Le protocole HTTP est le mécanisme universel d'accès au Web. Dans le Web classique, des URI HTTP combinent identification unique et mécanisme de récupération simple. Ainsi, le deuxième principe recommande l'utilisation de ces dernières pour identifier des objets et des concepts abstraits, afin que ces URI soient *déréférencées* (autrement dit, que l'on puisse récupérer le contenu pointé) par le protocole HTTP et traduites en une description de l'objet identifié ou du concept.

Pour que davantage d'applications différentes accèdent au contenu web, il est important d'utiliser un format de contenu standardisé. Le choix de HTML comme format de documents dominant a été un facteur prépondérant dans la croissance du Web.

Le troisième principe conseille donc un modèle unique pour publier des données structurées. Il s'agit de RDF, un modèle de données simple fondé sur une structure en graphe conçue spécifiquement pour le contexte du Web [Klyne, 2004].

Le quatrième principe propose d'utiliser les hyperliens afin de connecter toutes sortes d'éléments et pas seulement des documents web : ainsi, un hyperlien peut être défini entre une personne et un lieu ou entre un lieu et une entreprise. Si dans le Web classique, les hyperliens

n'ont généralement pas de type défini, ceux qui connectent des éléments dans un contexte de données liées ont des types qui décrivent les relations entre les éléments. Par exemple, un hyperlien de type *ami de* pourrait être défini entre deux personnes, un autre de type *habite près de* entre une personne et un endroit. Pour les distinguer, ces hyperliens sont appelés *liens RDF*.

De nombreux serveurs doivent répondre aux requêtes pour déréférencer des URI HTTP dans de nombreux espaces de noms et, dans un contexte de données liées, pour renvoyer des descriptions RDF correspondant aux ressources identifiées par ces URI. Ainsi, dans un tel contexte, si un lien RDF connecte des URI dans des espaces de noms différents, il connecte au final des ressources dans des jeux de données distincts.

De la même manière que les hyperliens dans le Web classique connectent des documents dans un espace global d'information, les données liées utilisent les hyperliens pour connecter des données disparates dans cet espace global, ce qui permet aux applications de pouvoir naviguer dans cet espace de données. Par exemple, une application, qui a connecté une URI et reçu des données RDF décrivant une personne, peut suivre des liens de ces données vers d'autres données sur d'autres serveurs web.

Au final, le Web de données s'adosse à des standards et à un modèle de données communs. Il est donc possible d'implémenter des applications génériques qui se fondent sur l'espace de données complet : par exemple, les navigateurs de données liées, qui permettent à l'utilisateur de voir des données d'une source puis d'en suivre les liens RDF vers d'autres sources, ou des moteurs de recherche de données liées, qui explorent le Web de données et ont la capacité de répondre à des requêtes sophistiquées à partir de l'espace des données.

En résumé, les principes des données liées posent les fondations de l'extension du Web dans un espace de données global, basé sur les mêmes principes d'architecture que ceux du Web de documents. Les sections suivantes expliquent en détail les techniques de réalisation de ces principes.

3.1 Rôle des URIs dans Linked Open Data

Pour publier des données sur le Web, il faut d'abord identifier les éléments du domaine d'intérêt. Il s'agit des éléments dont les propriétés et les relations seront décrites dans les données ; il peut s'agir de documents web, d'entités réelles et de concepts abstraits.

Puisque les données liées s'appuient directement sur l'architecture du Web [**Jacobs,2004**], le terme *ressource* est réutilisé pour nommer ces éléments dignes d'intérêt, qui sont, à leur tour, identifiés par des URI HTTP. La **Figure 02** montre l'utilisation d'URI HTTP pour identifier des entités réelles et leurs relations. Sur cette photo de l'équipe de tournage de *Big Lynx* au travail, on voit le cameraman principal, Matt Briggs, avec ses deux assistants, Linda Meyer et Scott Miller, identifiés par des URI HTTP de l'espace de noms *Big Lynx*. La relation (ils se connaissent) est représentée par des lignes, avec une URI de type `http://xmlns.com/foaf/0.1/knows`.



Figure 02 : les URI sont utilisées pour identifier des gens et les relations qui les joignent.

Ces données liées n'utilisent donc que des URI HTTP (on évite les autres schémas d'URI, comme URN [**Moats,1997**] et DOI [**Paskin,2006**]) et cela pour deux raisons :

1. Les URI HTTP fournissent une manière simple de créer des noms globalement uniques, de façon décentralisée, puisque n'importe qui possédant un nom de domaine peut créer ou déléguer la création de références URI.
2. Elles servent de nom mais aussi de moyen d'accès à l'information décrivant l'entité identifiée.

S'il vous semble étrange de penser à des URI HTTP en tant que noms pour des choses plutôt qu'à des adresses pour des documents web, alors nous vous recommandons fortement la lecture des références [Walsh,2006] et [Thompson,2006], qui méritent des visites régulières.

3.2 Rendre les URI déréférencables :

Toute URI HTTP devrait être déréférencable, autrement dit les clients HTTP devraient pouvoir visiter l'URI conformément au protocole HTTP et récupérer une description de la ressource identifiée. Cela s'applique aux URI nécessaires pour identifier des documents HTML classiques autant qu'à celles utilisées dans un contexte de données liées pour identifier des objets du monde réel et des concepts abstraits.

Les descriptions des ressources sont incarnées par des documents web. Celles qui sont prévues pour être lues par des êtres humains sont généralement représentées en HTML et celles qui sont prévues pour être consommées par des machines le sont par des données RDF.

Lorsque les URI identifient des objets réels, il est primordial de ne pas confondre les objets eux-mêmes avec les documents web qui les décrivent. Pour cela et afin d'éviter toute ambiguïté, on utilise couramment des URI différentes pour identifier l'objet réel et le document qui le décrit. Cette pratique permet d'effectuer des déclarations séparées sur un objet et sur le document qui le décrit. Par exemple, la date de création d'une personne peut être différente de celle du document qui la décrit. Distinguer les deux par des URI différentes est essentiel pour la cohérence du Web des données.

Le Web est conçu comme un espace de données utilisable tant par les humains que par des machines. Les deux parties devraient être capables de récupérer des représentations des ressources dans une forme qui répond à leurs besoins (HTML pour les humains et RDF pour les machines). Cela peut être réalisé avec le mécanisme HTTP appelé la *négociation du contenu* [Fielding, 1999].

L'idée sous-jacente suppose que les clients HTTP envoient des en-têtes HTTP avec chaque requête pour indiquer les types de documents qu'ils préfèrent. Les serveurs inspectent ces en-têtes et répondent de façon appropriée : si l'en-tête indique que le client préfère le HTML, le serveur lui envoie un document HTML et, s'il préfère le RDF, le serveur lui envoie un document RDF.

Il existe deux stratégies pour créer des URI qui identifient des objets réels déréférencables. Toutes deux s'assurent que les objets et les documents qui les décrivent ne sont pas confondus et que les humains autant que les machines peuvent récupérer des représentations appropriées. Ces stratégies sont appelées *URI 303* et *URI avec ancre*. La note du groupe d'intérêt *Cool URIs for the Semantic Web* du W3C [Sauermaun, 2008] les décrits et les justifie en détail. Les sections suivantes les résumant et les illustrent avec un exemple de session HTTP.

3.3 URI 303

Des objets réels, comme des maisons ou des gens, ne peuvent pas être transmis par le protocole HTTP. Il n'est donc pas possible de déréférencer directement les URI qui les identifient. Par conséquent, dans cette stratégie, au lieu d'envoyer l'objet lui-même par le réseau, le serveur répond au client par le code de réponse http 303 See Other et par l'URI d'un document web qui décrit l'objet réel. C'est ce que l'on appelle une "redirection 303". Dans un second temps, le client déréférence cette nouvelle URI et récupère un document web décrivant cet objet réel.

4. Topologie du Web de données

Cette section propose un aperçu de la topologie du Web de données telle qu'elle était en septembre 2011. Les jeux sont classés selon les domaines suivants : géographie, gouvernement, médias, bibliothèques, sciences de la vie, commerce, contenu généré par les utilisateurs, jeu de données à la croisée de plusieurs domaines. La **Figure 03** présente un aperçu du nombre de triplets à cette date, ainsi que le nombre de liens RDF par domaine. Il montre les liens sortant d'un jeu d'un domaine vers un autre jeu. Ces chiffres proviennent du document State of the LOD Cloud1 présentant l'état du nuage de LOD, document qui compile à intervalles réguliers des statistiques au sujet des jeux de données du catalogue CKAN.

<i>Domaines</i>	<i>Jeux de données</i>	<i>Triplets</i>	<i>Pourcentage</i>	<i>Liens RDF</i>	<i>Pourcentage</i>
Multidomaines	41	4 184 635 715	13,23	63 183 065	12,54
Géographie	31	6 145 532 484	19,43	35 812 328	7,11
Gouvernements	49	13 315 009 400	42,09	19 343 519	3,84
Médias	25	1 841 852 061	5,82	50 440 705	10,01
Bibliothèques	87	2 950 720 693	9,33	139 925 218	27,76
Sciences de la vie	41	3 036 336 004	9,60	191 844 090	38,06
Contenu utilisateur	20	134 127 413	0,42	3 449 143	0,68
Total	295	31 634 213 770		503 998 829	

Figure 03 : Statistiques des jeux de données du CKAN

4.1 Données multi-domaines

Parmi les premiers jeux de données, certains n'étaient pas propres à un domaine en particulier mais en couvraient plusieurs. Ces ponts entre domaines sont essentiels pour connecter des jeux spécifiques, créer un seul espace interconnecté de données et, ainsi, éviter la fragmentation du Web de données en îlots thématiques.

5. Les bases principales :

On définit quelques notions importantes de linked data sur l'architecture web et sur son utilisation du langage rdf pour la représentation des ressources.

5.1 L'architecture web :

L'architecture web telle que connue se base sur les protocoles et les langages à balise pour router et transmettre des données et bien les représentées.

5.1.1 Ressource :

Pour publier des données sur le web, on commence par identifier l'article qui nous intéresse et cela en décrivant ses propriétés et son thème. En architecture web tout article est

une ressource. Le «W3C Technologie Architecture Groupe (TAG) » définit deux types de ressources : ressource informationnelles et ressources non-informationnelles. Cette différence est très importantes en linked data. Toutes les ressources qu'on trouve dans le web documentaire sous le format document, image ou vidéo sont des ressources informationnelles, par contre il existe beaucoup de ressources non-informationnelles comme les lieux géographiques, les humains et les produits physiques. Tous ce que existe à l'extérieure du web sont considérés comme ressource non-informationnelles. [Chris, 2016]

5.1.1.1 Ressource informationnelle :

La ressource informationnelle est directement identifier par une URI déréférencé donc une nouvelle représentation de la ressource est très facile, et le serveur envoie la ressource au client en utilisant le http réponse code 200 ok. [Chris,2016]

5.1.1.2 Ressource non-informationnelle :

Une ressource non-informationnelle ne peut pas être identifiée par une URI déréférencé directement, le serveur répond par une redirection 303 vers une ressource informationnelle. [Chris,2016]

5.1.2 Identification de la ressource :

Une ressource est identifiée sur le web par un URI. En linked data on utilise seulement les http URI et on évite les autres types d'URI comme les URNs et DOIS. Les http URIs ont deux grands avantages pour l'identification des ressources : elles permettent la création simple et unique des identifications de ressources en premier lieu, et en deuxième lieu un http URI peut jouer un rôle de nom de la ressource comme elle peut définir le vrai sens et la définition de la ressource. [Chris,2016]

5.1.3 Représentation de la ressource :

Une ressource informationnelle peut avoir une représentation. Une représentation d'une ressource est un fichier de plusieurs formats HTML, RDF/XML ou JPEG. Une facture est une ressource informationnelle, elle peut être représentée sous différents formats (document PDF, document RDF ou sous format html...). Une ressource informationnelle peut avoir plusieurs représentations. Le W3C TAG a différencié la représentation en URI des ressources informationnelles et non-informationnelles. [Chris,2016]

5.1.4 Négociation du contenu

Les navigateurs peuvent afficher les fichiers RDF sous forme de ligne de code RDF comme ils donnent la possibilité de les télécharger sous forme de fichier sans les afficher. La représentation de la ressource sous forme html et RDF permet à l'utilisateur et à la machine de mieux comprendre le contenu de la ressource.

Les serveurs répondent aux clients sous différents formats selon la demande, les serveurs inspectent les en-têtes des requêtes http dont le client spécifier le type de la représentation. Si l'en-tête indique que le client préfère HTML le serveur génère une représentation HTML et si le client préfère RDF il génère rdf.

La négociation du contenu pour les ressources non-informationnelles est généralement mise en œuvre de la manière suivante : quand une URI identifie une ressource non-informationnelle le serveur envoie une redirection 303 vers une ressource informationnelle appropriée. Donc la représentation d'une ressource non-informationnelle est faite de 3 URI, par exemple on prend le cas de l'Algérie comme ressource :

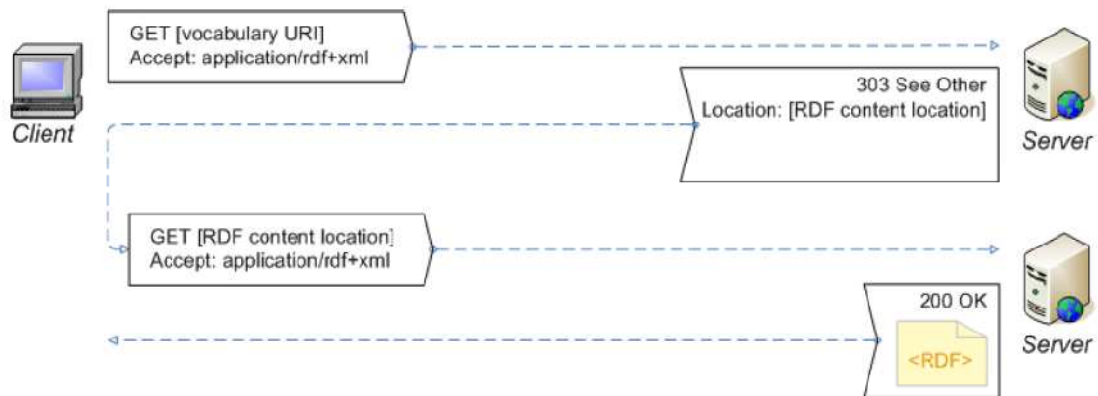
- <http://www4.wiwiss.fu-berlin.de/factbook/resource/Russia> (URI identifying the non-information resource Algeria).
- <http://www4.wiwiss.fu-berlin.de/factbook/data/Russia> (information resource with an RDF/XML representation describing Algeria).
- <http://www4.wiwiss.fu-berlin.de/factbook/page/Russia> (information resource with an HTML representation describing Algeria).

L'image représente comment le contenu d'une ressource non-informationnelle identifié par une URI déréférencé.

1. Le client effectue une requête http get sur une URI identifiant une ressource non-informationnelle, dans notre cas si le client est un moteur de recherche linked data et préfère la représentation rdf/xml de la ressource, il envoie un *Accept :application/RDF+XML* à l'en-tête de la requête. Et dans le cas d'un navigateur HTML il envoie *Accept : test/html* à l'en-tête de la requête.
2. Le serveur reconnaît l'URI pour identifier une ressource non-informationnelle. comme le serveur ne peut pas retourner une représentation à cette ressource, il répond en utilisant *http 303 see other* et envoie au client l'URI d'une ressource informationnelle

qui décrit la ressource non-informationnelle. Dans le cas de RDF il envoie *RDF content location*.

3. Le client maintenant demande au serveur pour obtenir une représentation à cette ressource informationnelle en demandant a nouveau *application/rdf+xml*.
4. Le serveur envoie au client un document RDF/XML contenant une description de l'origine de la ressource *vocabulary URI*. [Chris,2016]



5.1.5 URI alias

Dans un environnement ouvert comme le Web, il arrive souvent que les différents fournisseurs d'informations parlent de la même ressource non-informationnelle, par exemple d'un emplacement géographique ou d'une personne célèbre. Comme ils ne savent pas les uns des autres, ils introduisent différents URIs pour identifier le même objet du monde réel. Par exemple: DBpedia une source de données fournissant des informations qui ont été extraites de Wikipedia utilise l'URI <http://dbpedia.org/resource/Berlin> pour identifier Berlin. Geonames est une autre source de données fournissant des informations sur des millions d'emplacements géographiques utilise le <http://sws.geonames.org/2950159/> URI pour identifier Berlin. Comme les deux URIs se réfèrent à la même ressource non-informationnelle, ils sont appelés URIs alias. Les URIs alias sont communs sur le Web des données, car il n'est raisonnablement d'attendre que tous les fournisseurs d'informations seront d'accord sur les mêmes URIs pour identifier une des ressources non-information. Afin de suivre les différentes informations qui parlent de la même ressource non-informationnelle, les fournisseurs d'information mettent des liens owl: sameAs vers les URIs alias qu'ils connaissent. [Chris,2016]

5.2. Le modèle RDF

Lors de la publication de données liées sur le Web, on représente les informations sur les ressources à l'aide du cadre RDF. Il fournit un modèle de données qui est extrêmement simple d'une part, mais strictement adaptée à l'architecture Web d'autre part.

En RDF, une description d'une ressource est représentée par un nombre de triplets. Les trois parties de chaque triplet sont appelés : sujet, prédicat, et objet. Un triple miroir de la structure de base d'une phrase simple, comme celui-ci:

Chris has the **email address** **chris@bizer.de** .

(subject) *(predicate)* *(object)*

Le sujet d'un triplet est l'URI identifiant la ressource décrite. L'objet peut être soit une valeur littérale simple, comme une chaîne, un nombre, ou la date; ou un URI d'une autre ressource qui est en quelque sorte lié au sujet. Le prédicat, au milieu, indique ce qui existe type de relation entre le sujet et l'objet. Le prédicat est un URI aussi. Ces URIs prédicat proviennent de vocabulaires, des collections d'URIs qui peuvent être utilisés pour représenter des informations sur un certain domaine.

On distingue deux types de triplets RDF, les Triplets littéraux et les liens RDF :

5.2.1 Les littéraux RDF

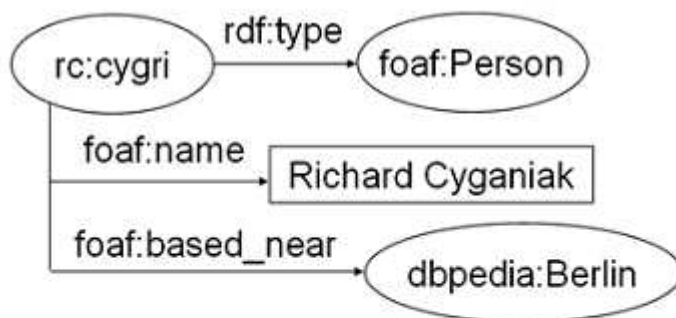
Un littéral RDF peut être une chaîne de caractères, un entier ou une date. Les triplets littéraux sont utilisés pour décrire les propriétés des ressources. Par exemple ce genre de triplet peut être utilisé pour décrire le nom ou la date de naissance d'une personne.

5.2.2 Les liens RDF

Les liens RDF sont des relations entre les ressources. Ils se composent de trois références URI. Les URIs dans le sujet et la position de l'objet de la liaison identifier les ressources liées entre elles. L'URI dans la position de prédicat définit le type de lien. Par exemple, un lien RDF peut affirmer qu'une personne est employée par une organisation. Un autre lien RDF peut affirmer que les personnes connaissent certaines autres personnes.

Les liens RDF sont la base du Web de données. L'URI déréférencé qui apparaît comme la destination d'un lien donne une description de la ressource liée. Cette description contient habituellement des liens RDF supplémentaires qui pointent vers d'autres URIs qui à leur tour peuvent également être déréférencé, et ainsi de suite. Voici comment les descriptions de ressources individuelles sont tissées dans le Web de données. Ceci est également la façon dont le Web de données peut être consulté en utilisant un navigateur de données liées ou exploré par le robot d'un moteur de recherche.

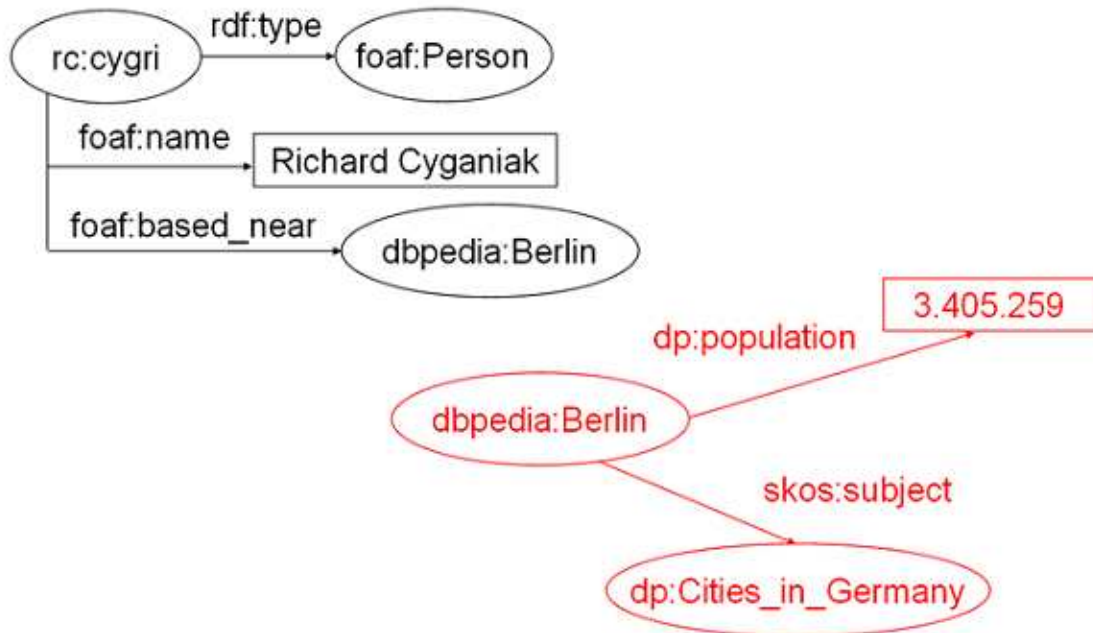
Prenons un navigateur RDF comme Disco ou Tabulator comme un exemple. L'internaute utilise le moteur de recherche pour afficher des informations sur Richard de son profil FOAF. Richard lui-même identifié avec l'URI `http://richard.cyganiak.de/foaf.rdf#cygri`. Lorsque l'internaute tape cet URI sur l'espace de recherche, le moteur de recherche déréférence cette URI sur le web et demandant le type de contenu `application : /rdf+xml` puis afficher les informations récupérées. Dans son profil, Richard a déclaré qu'il habite près de Berlin, en utilisant l'URI DBpedia `http://dbpedia.org/resource/Berlin` comme URI alias pour la ressource non-information de Berlin. Comme l'internaute s'est intéressé à Berlin, il indique au navigateur de déréférencer cet URI en cliquant dessus. Le navigateur déréférence maintenant cet URI demandant `application/rdf+xml`.



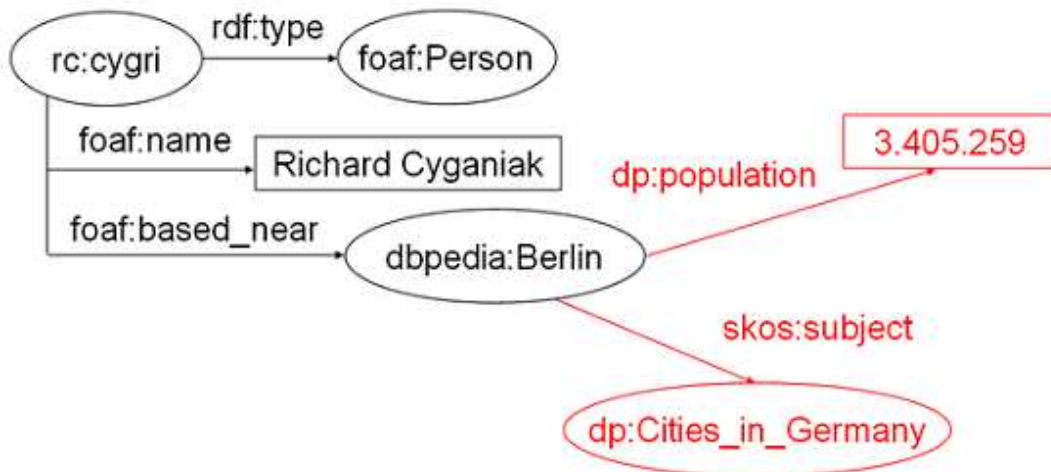
GET /resource/Berlin HTTP/1.0
Accept: application/rdf+xml

Après avoir été redirigée avec un code de réponse HTTP 303, le navigateur récupère un graphe RDF décrivant Berlin plus en détail. Une partie de ce graphique est illustré ci-dessous.

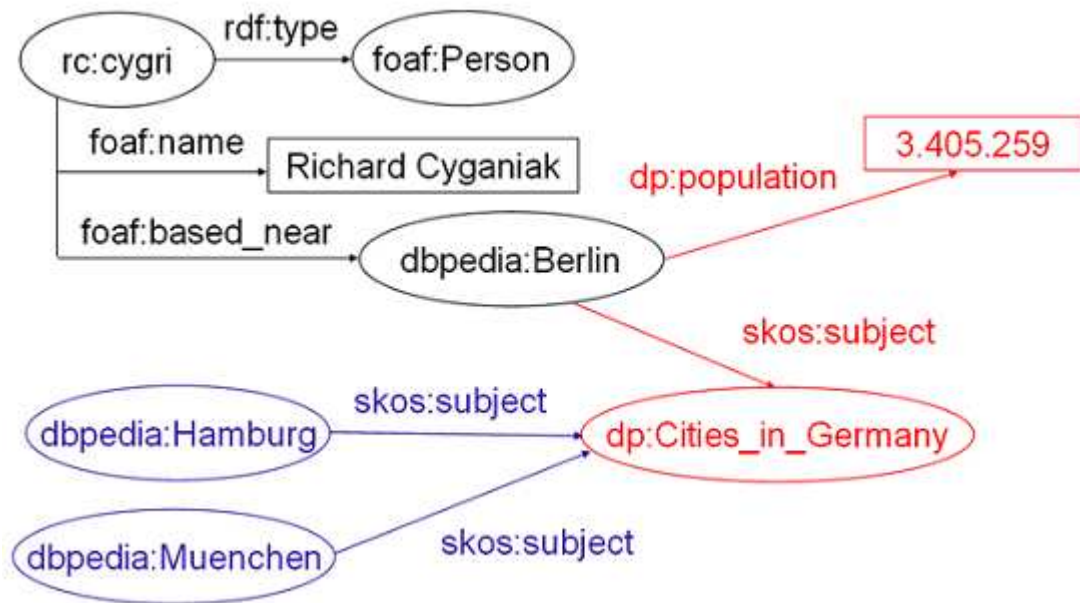
Le graphique contient un littéral triplet indiquant que Berlin a 3.405.259 habitants et un autre lien RDF à une ressource représentant une liste des villes allemandes.



Comme les deux graphes RDF partagent l'URI <http://dbpedia.org/resource/Berlin>, ils fusionnent naturellement ensemble, comme indiqué ci-dessous.



L'internaute peut également être intéressé par d'autres villes allemandes. Par conséquent, il laisse le navigateur déréférencer l'URI identifiant la liste. Le graphe RDF récupéré contient plus de liens RDF vers les villes allemandes, par exemple, Hambourg et Munich, comme illustré ci-dessous.



Vu du point de vue de Web, les liens les plus précieux RDF sont ceux qui relient une ressource à des données externes publiées par d'autres sources de données, car ils relient différents thèmes de données dan un site Web. Techniquement, un tel lien externe RDF est un triplet RDF qui possède un sujet URI à partir d'une source de données et un prédicat URI à partir d'une autre source de données. L'exemple ci-dessous contient différents liens RDF externes tirés de différentes sources de données sur le Web.

Examples of External RDF Links

Two RDF links taken from DBpedia

```
<http://dbpedia.org/resource/Berlin>
  owl:sameAs <http://sws.geonames.org/2950159/> .
<http://dbpedia.org/resource/Tim_Berners-Lee>
  owl:sameAs <http://www4.wiwiss.fu-berlin.de/dblp/resource/person/100007> .
```

RDF links taken from Tim Berners-Lee's FOAF profile

```
<http://www.w3.org/People/Berners-Lee/card#i>
  owl:sameAs <http://dbpedia.org/resource/Tim_Berners-Lee> ;
  foaf:knows <http://www.w3.org/People/Connolly/#me> .
```

RDF links taken from Richard Cyganiaks's FOAF profile

```
<http://richard.cyganiak.de/foaf.rdf#cygri>
  foaf:knows <http://www.w3.org/People/Berners-Lee/card#i> ;
  foaf:topic_interest <http://dbpedia.org/resource/Semantic_Web> .
```

Les principaux avantages de l'utilisation du modèle de données RDF dans un contexte Linked Data sont les suivantes:

- Les clients peuvent consulter tous les URIs dans un graphe RDF sur le Web pour récupérer des informations supplémentaires.
- Informations provenant de différentes sources sont fusionnées naturellement.
- Le modèle de données permet de définir des liens RDF entre les données provenant de différentes sources.
- Le modèle de données permet de représenter l'information qui est exprimée à l'aide de schémas différents dans un modèle unique.
- Combiné avec des langages de schéma comme RDFS ou OWL, le modèle de données permet d'utiliser autant ou aussi peu de structure dont on a besoin, ce qui signifie qu'on peut représenter des données bien structurées ainsi que des données semi-structurées.

6. Bien choisir un URI

Les ressources sont nommées avec des références URI. Lors de la publication de données liées, on doit consacrer un certain effort pour choisir de bons URIs pour les ressources. [Chris,2016]

D'une part, un URI doit être valide pour que d'autres fournisseurs d'information puissent l'utiliser en toute confiance pour créer un lien vers cette ressource dans leurs propres sources données. D'autre part, on doit mettre l'infrastructure technique en place pour les rendre déréférencable.

Pour bien choisir un URI valide on doit garder en esprit cette liste :

- Utiliser HTTP URIs pour toutes les données. Le http: // est le seul schéma d'URI qui est largement pris en charge dans les outils et l'infrastructure d'aujourd'hui. Tous les autres systèmes nécessitent un effort supplémentaire pour les services Web, traitant avec les registres d'identification, et ainsi de suite. Les arguments en faveur de l'utilisation de HTTP sont discutés en plusieurs endroits.

- Définir les URIs dans un espace de noms HTTP sous le contrôle du créateur, où on peut réellement les définir dérivable. Ne pas les définir dans un autre espace de noms.
- Garder les URIs stable et persistante. Changer les URIs va briser les liens déjà établis, il est donc conseillé de consacrer une réflexion supplémentaire pour eux à un stade précoce.
- Une ressource non-informationnelle est définie par trois URI :
 1. Un identificateur de la ressource
 2. Un identificateur pour une ressource informationnelle relative, pour les navigateurs html
 3. Un identificateur pour une ressource informationnelle relative, pour les navigateurs rdf

On aura souvent besoin d'utiliser une sorte de clé primaire à l'intérieur des URIs, pour assurer que chacun est unique. Si on peut, on utilise une clé qui est significatif dans un domaine de travail. Par exemple, lorsqu'ils traitent avec des livres, ce qui rend le numéro ISBN partie de l'URI est mieux que d'utiliser la clé primaire d'une table de base de données interne. Cela rend également l'exploitation minière d'équivalence pour dériver RDF liens plus facile.

7. Choix de vocabulaire pour représenter l'information

Afin que les applications clientes traitent les données facilement, il est préférable de réutiliser les termes de vocabulaires bien connus chaque fois que possible. On peut définir de nouveaux termes si on ne les trouve pas dans les vocabulaires existants. [Chris,2016]

7.1. La réutilisation des termes existants :

Un ensemble de vocabulaires bien connus a évolué dans la communauté du Web sémantique. On vérifie toujours si les données peuvent être représenté avec les vocabulaires existant avant de les définir. Les vocabulaires les plus utilisés sont :

Friend-of-a-Friend (FOAF), vocabulaire pour décrire les gens.

Dublin Core (DC), définit les caractéristiques générales de métadonnées. Voir aussi leurs nouveaux domaines et projet

Semantically-Interlinked Online Communities (SIOC), vocabulaire pour représenter les communautés en ligne.

Description of a Project (DOAP), vocabulaire pour décrire les projets.

Simple Knowledge Organization System (SKOS), vocabulaire pour représenter les taxonomies et les connaissances vaguement structuré

Music Ontology fournit des termes pour décrire les artistes, les albums et les pistes.

Review Vocabulary, vocabulaire pour représenter les revues.

Creative Commons (CC), vocabulary for describing license terms.

Une liste plus complète des vocabulaires bien connus est maintenue par le projet communautaire Linking Open Data SWEOW3C dans le ESW Wiki. Une liste des 100 espaces de noms RDF les plus courants (Août 2006) est fournie par UMBC Ebiquity Group.

Il est plus pratique de mélanger des termes de vocabulaires différents. Il est notamment recommandé d'utiliser de rdfs :label et foaf :depiction propriétés à chaque fois que ces termes sont bien supportés par les application clients.

Si on a besoin des URIs référencés pour les lieux géographiques, les domaines de recherche, des thèmes généraux, des artistes, des livres ou des CD, on utilise les URIs prédéfini des sources de données du projet communautaire SWEOW3C Linking Open Data, par exemple Geonames, DBpedia, Musicbrainz, dbtune ou RDF Book Mashup. Les deux principaux avantages de l'utilisation des URIs de ces sources de données sont les suivantes:

- Les URIs sont dérivable, ce qui signifie que la description du concept peut être récupérée à partir du Web
- Les URIs sont déjà liés à URIs d'autres sources de données.

7.2 Définir un terme

Lorsqu'on ne trouve pas de bons vocabulaires existants qui couvrent toutes les classes et les propriétés dont on a besoin. On doit définir de nouveaux termes. Les classes et les propriétés RDF sont aussi des ressources identifiées par des URIs et publier sur le web, donc tout ce qu'on a dit sur la publication des données liées s'applique sur eux-mêmes. [Chris, 2016]

On peut définir des vocabulaires en utilisant le vocabulaire RDF Description Language 1.0: RDF Schema ou le Web Ontology Language (OWL).

Pour mieux définir de nouveaux termes il est préférable de respecter les aspects suivant :

- Ne pas définir de nouveaux vocabulaires à partir de zéro
- Fournir pour les humains et les machines.
- Faire URIs terme dérivable.
- Exploitez les termes d'autres personnes.
- Bien expliquer les informations importantes.
- Ne créez pas trop de contraintes.

L'exemple suivant contient une définition d'une classe et une propriété en suivant les règles ci-dessus. L'exemple suivant utilise la syntaxe Turtle.

Definition of the class "Lover"

```
<http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LoveVocabulary#Lover>
  rdf:type rdfs:Class ;
  rdfs:label "Lover"@en ;
  rdfs:label "Liebender"@de ;
  rdfs:comment "A person who loves somebody."@en ;
  rdfs:comment "Eine Person die Jemanden liebt."@de ;
  rdfs:subClassOf foaf:Person .
```

Definition of the property "loves"

```
<http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LoveVocabulary#loves>
  rdf:type rdf:Property ;
  rdfs:label "loves"@en ;
  rdfs:label "liebt"@de ;
  rdfs:comment "Relation between a lover and a loved person."@en ;
  rdfs:comment "Beziehung zwischen einem Liebenden und einer geliebten Person."@de ;
  rdfs:subPropertyOf foaf:knows ;
  rdfs:domain <http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LoveVocabulary#Lover> ;
```

rdfs:range foaf:Person .

8. La description RDF

Supposant qu'on a représenté toutes les données sous forme de triplets RDF.

- **La description:** La représentation doit inclure tous les triplets de dataset qui ont l'URI de la ressource en tant que sujet. Ceci est la description immédiate de la ressource.
- **Backlinks:** La représentation devrait également inclure tous les triplets de dataset qui ont l'URI de la ressource en tant que objet. Ceci est redondant, car ces triplets peuvent déjà être récupérés à partir de leur objet URIs, mais il permet aux navigateurs et les robots de traverser des liens dans les deux sens.
- **descriptions connexes:** on peut inclure toute information supplémentaire sur les ressources connexes qui peuvent être d'intérêt dans les scénarios d'utilisation typiques. Par exemple, on peut envoyer des informations sur l'auteur ainsi que des informations au sujet d'un livre, parce que beaucoup de clients intéressés par le livre peuvent également être intéressés par l'auteur.
- **Métadonnées:** La représentation doit contenir toutes les métadonnées qu'on souhaite joindre aux données à publiées, comme un URI identifiant l'auteur et l'authentification. Ceux-ci doivent être enregistrés comme des descriptions RDF de la ressource d'information qui décrit une ressource non-information; autrement dit, l'objet du triplet RDF devrait être l'URI de la ressource d'informations.
- **Syntaxe:** Il existe différentes façons de sérialiser une description RDF. Le dataset devrait au moins fournir des descriptions RDF comme RDF / XML qui est la seule syntaxe officielle pour RDF. Comme RDF / XML est pas très lisible par l'homme, le dataset pourrait en outre fournir une description Turtle quand on lui demande le type MIME application / x-turtle.

8.1. Description autorisé

L'exemple suivant illustre des parties de la représentation Turtle d'une ressource d'information http://dbpedia.org/data/Alec_Empire. La ressource décrit le musicien allemand Alec Empire. Utilisation de Web architecture terminologie, c'est une description retourné après une redirection 303 par le propriétaire de l'URI http://dbpedia.org/resource/Alec_Empire.

Metadata and Licensing Information

```
<http://dbpedia.org/data/Alec_Empire>
  rdfs:label "RDF description of Alec Empire" ;
  rdf:type foaf:Document ;
  dc:publisher <http://dbpedia.org/resource/DBpedia> ;
  dc:date "2007-07-13"^^xsd:date ;
  dc:rights
    <http://en.wikipedia.org/wiki/WP:GFDL> .
```

The description

```
<http://dbpedia.org/resource/Alec_Empire>
  foaf:name "Empire, Alec" ;
  rdf:type foaf:Person ;
  rdf:type <http://dbpedia.org/class/yago/musician> ;
  rdfs:comment
    "Alec Empire (born May 2, 1972) is a German musician who is ..."@en
  rdfs:comment
    "Alec Empire (eigentlich Alexander Wilke) ist ein deutscher Musiker. ..."@de ;
  dbpedia:genre <http://dbpedia.org/resource/Techno> ;
  dbpedia:associatedActs <http://dbpedia.org/resource/Atari_Teenage_Riot> ;
  foaf:page <http://en.wikipedia.org/wiki/Alec_Empire> ;

  foaf:page <http://dbpedia.org/page/Alec_Empire> ;
  rdfs:isDefinedBy <http://dbpedia.org/data/Alec_Empire> ;
  owl:sameAs <http://zitgist.com/music/artist/d71ba53b-23b0-4870-a429-cce6f345763b> .
```

Backlinks

```
<http://dbpedia.org/resource/60_Second_Wipeout>
  dbpedia:producer <http://dbpedia.org/resource/Alec_Empire> .
<http://dbpedia.org/resource/Limited_Editions_1990-1994>
  dbpedia:artist <http://dbpedia.org/resource/Alec_Empire> .
```

8.2 Description non-autorisé:

L'exemple suivant montre la représentation Turtle de la ressource d'information <http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/ChrisAboutRichard> qui est publié par Chris pour fournir des informations sur Richard. On note qu'en Turtle, la syntaxe `<>` peut être utilisé pour faire référence à l'URI du document en cours. Richard possède l'URI <http://richard.cyganiak.de/foaf.rdf#cygri>, donc il est le seul qui peut fournir une description autorisé pour cette URI. Ainsi, en utilisant la terminologie Architecture Web, Chris peut fournir des informations non-autorisées sur Richard. [Chris,2016]

Metadata and Licensing Information

```
<>
  rdf:type foaf:Document ;
  dc:author <http://www.bizer.de#chris> ;
  dc:date "2007-07-13"^^xsd:date ;
```

```
cc:license <http://web.resource.org/cc/PublicDomain> .
```

The description

```
<http://richard.cyganiak.de/foaf.rdf#cygri>  
foaf:name "Richard Cyganiak" ;  
foaf:topic_interest <http://dbpedia.org/resource/Category:Databases> ;  
foaf:topic_interest <http://dbpedia.org/resource/MacBook_Pro> ;  
rdfs:isDefinedBy <http://richard.cyganiak.de/foaf.rdf> ;  
rdf:seeAlso <> .
```

Backlinks

```
<http://www.bizer.de#chris>  
foaf:knows <http://richard.cyganiak.de/foaf.rdf#cygri> .  
<http://www4.wiiss.fu-berlin.de/is-group/resource/projects/Project3>  
doap:developer <http://richard.cyganiak.de/foaf.rdf#cygri> .
```

8.3 RDFa

RDFa (pour « Resource Description Framework dans des Attributs ») est une recommandation du W3C définissant une syntaxe permettant d'ajouter des données structurées dans une page HTML ou n'importe quel document XML. Ainsi formellement décrites, les données peuvent alors faire l'objet de traitements automatisés complexes, via des outils adaptés. Le code RDFa est invisible pour l'internaute et n'affecte pas ce qui est affiché.

9. Définir un lien RDF vers un autre dataset

Les liens RDF permettent aux moteur de recherche linked data de naviguer entre différents sources de données (datasets) et découvrir des nouveaux.

Le domaine d'application permet de déterminer les propriétés RDF et sont utilisées comme prédicats. Prenant le cas d'utilisation des propriétés de liaison entre des personnes comme foaf: knows, foaf: based_near et foaf: topic_interest. Des exemples de la combinaison de ces propriétés avec des valeurs de propriété de DBpedia, la bibliographie de DBLP et le RDF Book Mashup se trouvent dans les profils de Tim Berners-Lee et FOAF Ivan Herman.

La propriété owl:sameAs est utilisé pour définir un autre dataset qui contient des information sur une ressource non-informationnelle. Les liens owl:sameAs indiquent que deux URIs identifient une même ressource, ils sont aussi utilisé pour relier deux URI alias.

Les liens rdf peuvent êtres défini manuellement, ce qui est généralement le cas pour les profils FOAF, ou ils peuvent être générer par des algorithmes automatique, cette approche est généralement utilisé pour relier les grand datasets.

9.1 définir les liens RDF manuellement

Avant de définir un lien RDF manuellement, on doit avoir quelques informations sur le datasets cible. Apres avoir choisi un dataset cible parmi les datasets existant sur le web, on peut faire une recherche manuelle pour trouver son URI et le relier avec notre dataset. Et si ce dataset n'a pas une interface de recherche comme un SPARQL endpoint ou une interface web, on utilise un moteur de recherche linked data avec qui on explore le dataset et on trouve son URI.

On peut également utiliser des services tels que Uriqr ou Sindice pour rechercher URIs existants et de choisir le plus populaire si on trouve plusieurs candidats. Uriqr permet de trouver les URIs pour les personnes qu'on connaisse, tout simplement en cherchant leurs noms. Mais on aura besoin d'appliquer un peu de l'intelligence humaine pour choisir l'URI le plus approprié à utiliser parmi les résultats. Sindice indexe le Web sémantique et peut définir la source de certains URIs. Par conséquent, ce service peut aider à choisir l'URI le plus populaire pour un concept. Sans oublier que les datasets peuvent utiliser le http 303 redirects pour rediriger les clients d'un URI qui identifie une ressource non-informationnelle vers un autre qui identifie une ressource informationnelle décrivant la ressource non-informationnelle. Dans ce cas il faut faire attention a ne pas définir le lien vers le URI du document qui décrit la ressource non-informationnelle au lieu de le définir vers l'URI de cette dernière.

9.2 générer les liens RDF automatiquement

La définition manuelle des liens RDF entre datasets est utilisée lors qu'on travaille sur des datasets de petite taille. Par contre si on veut interconnecter 70 000 place de Geonames vers DBpedia on est obligé d'utiliser un algorithme de liaison des enregistrements pour

générer des liens entre les deux datasets. Ce genre d'algorithmes ont connu beaucoup de problèmes avec les bases de données. [Chris,2016]

Il y a encore un manque de bons outils faciles à utiliser pour générer automatiquement des liens RDF. Par conséquent, il est plus pratique de mettre en œuvre des algorithmes de liaison d'enregistrements spécifiques aux datasets pour générer des liens RDF. il existe deux type d'algorithmes : simples et complexes.

9.2.1 Algorithmes simples

Dans le domaine de la publication, il y a les numéros ISBN, dans le domaine financier il y a les identificateurs ISIN. Si ces identificateurs sont utilisés comme une partie d'URI HTTP identifiant une ressource particulière, il est possible d'utiliser des algorithmes pour générer des liens RDF entre ces ressources. [Chris,2016]

RDF Book Mashup est un exemple de dataset utilisant les numéros ISBN comme une partie de ses URIs, qui, par exemple, utilise l'URI <http://www4.wiwiss.fu-berlin.de/bookmashup/books/0747581088> pour identifier le livre ' Harry Potter et le prince de sang mêlé'. Avoir le numéro ISBN dans ces URIs facilite à DBpedia de générer des liens owl: sameAs entre les livres enregistrés dans DBpedia et Book Mashup. DBpedia utilise l'algorithme suivant:

- Itérer sur tous les livres dans DBpedia qui ont un numéro ISBN.
- Créer un lien owl: sameAs entre l'URI d'un livre dans DBpedia et l'URI correspondant au même livre dans Book Mashup, en utilisant ce modèle d'URIs pour Book Mashup: <http://www4.wiwiss.fu-berlin.de/bookmashup/books/{ISBN nombre}>.

L'exécution de cet algorithme pour tous les livres dans DBpedia a abouti à 9000 liens RDF qui ont été fusionnées avec le dataset DBpedia. Par exemple, le lien résultant pour le livre Harry Potter est:

http://dbpedia.org/resource/Harry_Potter_and_the_Half-Blood_Prince

owl:sameAs <http://www4.wiwiss.fu-berlin.de/bookmashup/books/0747581088>

9.2.2 Algorithmes complexes:

Dans les cas où il n'existe pas des identifiants communs entre les datasets, il est nécessaire d'utiliser des algorithmes de liaison plus complexes. On présente deux algorithmes ci-dessous : [Chris,2016]

Interconnexion DBpedia et Geonames. Information sur les lieux géographiques apparaissent dans la base de données Geonames ainsi que dans DBpedia. Afin d'identifier les lieux qui apparaissent dans les deux ensembles de données (datasets), l'équipe Geonames utilise une heuristique qui est basée sur les titres d'article et leurs informations sémantiques, comme la latitude et la longitude, mais aussi le pays, la division administrative, le type d'entité, la population et les catégories. L'exécution de cette heuristique entre les deux datasets a abouti à 70500 correspondances qui ont été fusionnées en tant liens Geonames owl: sameAs avec les deux datasets.

Interconnexion Jamendo et MusicBrainz. : Cet algorithme permet de créer des liens entre les informations décrivant les ressources (informations personnelles, identifiant des disques).

10. Publier les données en linked data

Ce chapitre fournit les étapes pratiques pour la publication de différents types d'informations sous forme de Linked data sur le Web. Pour publier une ressource en tant que donnée liée sur le Web on doit satisfaire les exigences suivantes:

- _ La ressource doit être identifiée avec une URI HTTP déréférencable.
- _ Si un URI déréférencé demandant pour un type MIME *application / rdf + xml*, le dataset doit retourner une description RDF / XML de la ressource identifiée.
- _ URIs identifiant les ressources non-informationnelles doivent être définies dans l'une de ces façons :

Toute source de données doit renvoyer une réponse HTTP contenant un HTTP 303 redirigé vers une ressource informationnelle décrivant la ressource non-information. Ou l'URI

de la ressource non-informationnelle doit être formé en prenant l'URI de la ressource informationnelle qui la décrit et en ajoutant un identifiant de fragment.

_ En plus que RDF permet la création des liens entre ressources de même dataset, il permet également de créer des liens entre des ressources appartenant aux différents datasets, de sorte que les clients peuvent naviguer sur le Web de données en suivant les liens RDF.

10.1 Utilisation des fichiers RDF statiques :

La façon la plus simple d'utiliser les linked data c'est de créer un fichier statique RDF et de l'enregistrer sur un serveur web. Cette technique est utilisée :

- Quand les fichiers RDF sont créés manuellement généralement utilisés pour publier les fichiers personnels FOAF ou les vocabulaires RDF.
- Quand les fichiers RDF sont générés par une partie d'un logiciel

Mais y a quelques exigences pour appliquer cette technique.

Configuration du serveur pour un type MIME correcte :

Les anciens serveurs web sont parfois pas encore configurés pour renvoyer le type MIME correct en servant des fichiers RDF / XML. Les navigateurs de données liées peuvent ne pas reconnaître les données RDF servis de cette façon parce que les serveurs exigent le texte brut et non plus RDF / XML. Donc pour résoudre ce problème il faut configurer le serveur web de sorte qu'il reconnait les fichiers xml/rdf.

Taille du fichier

Les documents Web, il est considéré comme une mauvaise forme de publier d'énormes pages HTML, car ils se chargent très lentement dans les navigateurs et consomment de la bande passante nécessaire. Le même est vrai lors de la publication de données liées. Les fichiers RDF ne devraient pas dépasser quelques centaines de kilo-octets. Si les fichiers sont plus grands et décrivent des ressources multiples, on doit les diviser en plusieurs fichiers RDF, ou utilisez Pubby pour les servir en morceaux.

Lors de l'utilisation de plusieurs fichiers RDF, on doit s'assurer qu'ils sont liés entre eux par des triplets RDF impliquant des ressources décrites dans des fichiers différents.

Choisir les URIs pour les ressources non informationnelles

La méthode de fichier statique ne supporte pas les 303 redirections nécessaires pour les URIs des ressources non-informationnelles. Il existe une autre méthode pour identifier les ressources non-informationnelles, qui fonctionne très bien avec les fichiers RDF statiques conformes aux normes. Cette méthode repose sur URIs de hachage.

Lorsqu'on utilise un fichier RDF statique, disons, `http://example.com/people.rdf`, alors on doit nommer les ressources non-informationnelles décrites dans le fichier en ajoutant un identifiant de fragment à l'URI du fichier. L'identifiant doit être unique au fichier. De cette façon, on se retrouve avec des URIs comme celui-ci pour les ressources non-informationnelles:

`http://example.com/people.rdf#alice`

`http://example.com/people.rdf#bob`

`http://example.com/people.rdf#charlie`

Cela fonctionne parce que les clients HTTP déréférencent les URIs découpés en enlevant la partie après le découpage et déréférencent l'URIs résultant. Les navigateurs Linked data cherchent alors dans la réponse (le fichier RDF dans ce cas), et trouvent les triplets qui contient plus d'informations sur la ressource non-informationnelle.

L'inconvénient de cette approche de nommage est qu'elle ne respecte pas les aspects de création des URIs. L'extension `.rdf` est une référence à ce format de représentation. Et si on décide de renommer le fichier RDF plus tard, ou de diviser les données en plusieurs fichiers, tous les identificateurs vont changer et liens existants seront brisés.

Donc cette méthode est utiliser que si la structure globale et la taille de l'ensemble de données n'auront pas beaucoup de mise à jours à l'avenir, mais ça reste une solution rapide pour les données transitoires où la stabilité du lien est pas si important.

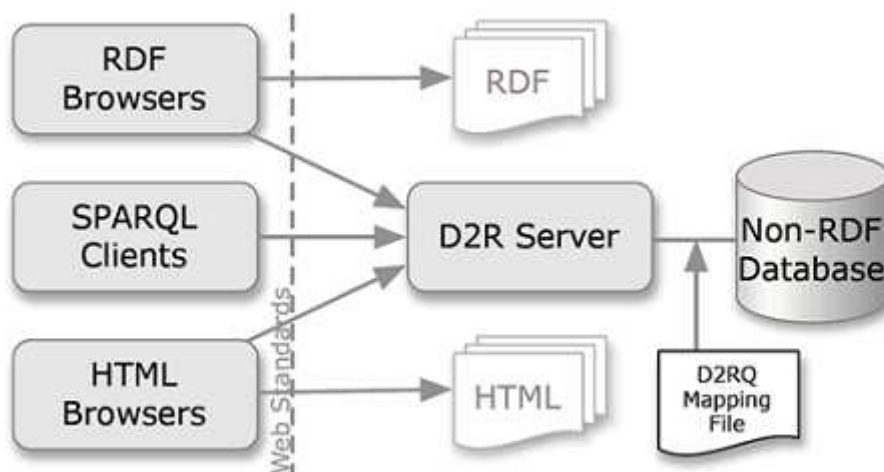
Extension de la recette pour 303 redirections et négociation de contenu

Cette approche peut être étendue pour utiliser 303 redirections et même pour supporter la négociation de contenu, Mais ce processus dépend du serveur Web et sa configuration. Le W3C a publié plusieurs articles qui montrent comment faire pour le serveur Web Apache. Le document est officiellement destiné aux éditeurs de vocabulaires RDF, mais ces articles parlent aussi d'autres types de données RDF servi à partir de fichiers statiques.

10.2 Utilisation des bases des données relationnelles

Dans le cas où les données sont stockées dans une base de données relationnelle, il est généralement plus pratique de les laisser telles qu'elles sont et de publier une vue Linked Data de cette base de données.

Le serveur D2R est un outil pour créer des vues de données liées à partir des bases de données relationnelles. Le serveur D2R repose sur une méthode déclarative entre les schémas de la base de données et les termes RDF. Sur la base de cette méthode, D2R serveur crée une vue de données liées sur la base de données cible et fournit pour elle un point de terminaison (endpoint) SPARQL .



Il existe plusieurs serveurs D2R en ligne, par exemple

- Berlin DBLP Bibliographie Server,
- Hannover DBLP Bibliographie Server,

- Systems Web @ serveur FU Berlin Group
- EuroStat Pays et Régions Server.

Il existe aussi d'autres outils pour publier les bases de données relationnelles et les publier en linked data :

- Utilisation de l'OpenLink Virtuoso pour publier votre base de données relationnelle comme Linked Data.
- Triplify, un petit plugin pour les applications Web, qui révèle les structures sémantiques crypté dans des bases de données relationnelles et rendre le contenu de base de données disponible en RDF, JSON ou Linked Data.

10.3 Utilisation d'autre type d'informations

Si les informations sont représentée dans des formats tels que CSV, Microsoft Excel, ou BibTEX et on veut les publier sous forme de données liées sur le Web, on suit les étapes suivantes :

- Convertir les données en RDF en utilisant l'outil de RDFizing.
- Stocker les données converties dans un répertoire RDF
- Le répertoire choisi doit être compatible avec l'interface de linked data pour que ses données seront accessibles sur le web. Comme beaucoup de répertoires n'ont pas encore implémenté des interfaces linked data, on peut choisir un répertoire qui fournit un endpoint SPARQL et Pubby comme interface de données liées.

L'approche décrite ci-dessus est supposé par le projet DBpedia. Le projet utilise des scripts PHP pour extraire des données structurées à partir des pages de Wikipédia. Ces données sont ensuite converties en RDF et stocké dans un répertoire Virtuoso OpenLink qui fournit un point d'extrémité SPARQL. Afin d'obtenir une vue Linked Data, Pubby est mis en face de point d'extrémité SPARQL.

Si le dataset est suffisamment petit pour tenir complètement dans la mémoire principale du serveur web, alors on peut le faire sans le répertoire RDF, et au lieu d'utiliser *Pubby conf:loadRDF* pour charger les données RDF à partir d'un fichier RDF, le chargement se fait directement dans Pubby. Cela apparait plus simple, mais contrairement à un véritable

répertoire RDF, Pubby va tout garder dans la mémoire principale et ne propose pas de point d'extrémité SPARQL.

10.4 Création des wrappers autour des API

Un grand nombre d'applications Web ont commencé à rendre leurs données disponibles sur le Web via les API Web comme eBay, Amazon, Yahoo, Google et Google Base. Différentes API fournissent des interfaces d'interrogation et renvoient des résultats en différents formats tels que XML, JSON ou ATOM. Il existe trois inconvénients pour l'utilisation des APIs web sont :

- Le contenu d'une API Web ne peut pas être analysé par les moteurs de recherche
- Les APIs Web ne sont pas accessibles par des navigateurs de données génériques
- Les Mashup sont utilisés pour quelques sources de données et ne peut être utilisé pour les nouvelles sources de données.

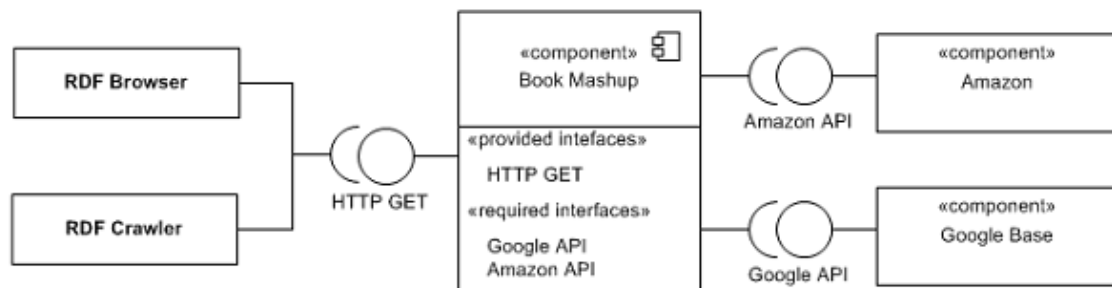
Pour dépasser ces inconvénients il faut mettre en œuvre des wrappers des données liées autour de l'API. En général, les wrappers de données liées procèdent comme suit:

1. Elles attribuent les URIs HTTP aux ressources non-informations sur lesquelles l'API fournit des données.
2. Lorsque l'un de ces URIs est déréférencé demandant *Application / rdf + xml*, le wrapper réécrit la requête du client en une requête de l'API.
3. Les résultats des requêtes de l'API sont transformés en RDF et renvoyés au client.

Quelques exemples de wrappers de données liées :

Le RDF Book Mashup

Le RDF Book Mashup rend l'information sur les livres, leurs auteurs, des critiques et des librairies en ligne disponibles sous RDF sur le Web. Le RDF Book Mashup assigne un URI HTTP à chaque livre qui a un numéro ISBN. Chaque fois que l'un de ces URIs est déréférencé, le Livre Mashup demande des données au sujet du livre, son auteur ainsi que les critiques et les offres de ventes de l'API Amazon et l'API Google Base. Ces données sont ensuite transformées en RDF et renvoyé au client.



Le RDF Book Mashup est implémenté comme un petit script PHP qui peut être utilisé comme modèle pour la mise en œuvre des wrappers similaires autour d'autres API Web

SIOC exportateurs pour WordPress, Drupal, phpBB

Le projet SIOC a développé wrappers de données liées pour plusieurs moteurs de blogs populaires, les systèmes de gestion de contenu et forums de discussion. Le projet prévoit également une API PHP Export qui permet aux développeurs de créer de nouveaux outils SIOC d'exportation sans avoir à entrer dans les détails techniques sur la façon dont l'information est représentée en RDF.

Virtuoso Sponger

Virtuoso Sponger est un Framework pour le développement des wrappers de données liées (appelées cartouches) autour de différents types de sources de données. Les sources de données peuvent ranger les données structurées à partir de pages html dans les APIs.

10.5 Choisir la méthode adéquate pour la publication des linked open data

:

Selon la taille :

Dans le cas où on veut publier une centaine de triplet RDF, de préférence utiliser les fichiers RDF statiques. Par contre si les données sont assez grandes il vaut mieux utiliser Pubby.

Type de données :

Dans la cas ou les données sont stocké dans une base de données relationnelle on doit utiliser D2R serveur, et si elles sont disponibles via une API on doit utiliser un wrappers autour de cette API. Et si les informations sont représentées dans un autre format Excel, csv ... on doit utiliser d'abord un convertisseur RDF.

Temps de mise à jour :

Dans le cas ou les données changent fréquemment il est préférable d'utiliser les outils qui génèrent des vues RDF sur les données tel que D2R serveur et les wrappers.

Afin de le rendre facile non seulement pour les machines mais aussi pour l'homme à découvrir les données publiées, il est nécessaire d'ajouter cet ensemble de données à la liste des datasets sur l'ESW Wiki, et inclure quelques exemples URIs de ressources intéressantes, pour que les gens auront des points de départ pour la navigation.

Introduction :

La thématique émergente des Linked Open Data (LOD) représente peut-être l'aube d'une révolution en milieu socio-économique et en éducation. Dans ce chapitre, nous avons souhaité brosser un aperçu des enjeux auxquels seront probablement bientôt confrontés les professionnels de l'information.

1. Utilisation des LOD en milieu socio-économique :

1.1. L'open data au service de la démocratie :

Si l'open data séduit les gouvernements, c'est qu'il propose une vision nouvelle de la démocratie. Au lendemain de sa prise de fonction en janvier 2009, Barack Obama a été l'un des premiers chefs d'État à soutenir l'open data au sein de son gouvernement. Dans sa continuité, et grâce au travail acharné de mouvements citoyens, d'autres collectivités ont ensuite participé au mouvement. En France, des associations comme LiberTIC ou Regards Citoyens ont beaucoup fait avancer l'État sur le sujet. Auparavant dans l'opposition, ces groupes aident désormais le gouvernement, qui se veut plus ouvert.

L'open gouvernement fait beaucoup de promesses. Il symbolise le souhait d'une démocratie plus participative, où les citoyens s'impliqueraient davantage et comprendraient des politiques souvent difficiles à saisir. Des sites comme « Nos Députés » ou *Where does my money go* montrent comment mieux inclure les citoyens dans le procédé démocratique. Ils rendent la gestion des budgets, de la participation des élus, et d'autres événements politiques (élections, nouvelles lois, etc) plus transparents. Une nouvelle expertise citoyenne pourrait ainsi se développer autour des données.

Outre le besoin de transparence, l'open data a aussi un autre intérêt pour les gouvernements : responsabiliser les citoyens. Lors d'un entretien, Emmanuel Raviart, directeur technique d'etalab (la plateforme française de l'open data affiliée au gouvernement) expliquait que les États ont été surpassés par le secteur privé pour la gestion de certains domaines. Le support citoyen apporté grâce à l'open data permet alors de maintenir des services dans le secteur public. L'exemple de la cartographie est frappant : Google Maps est plus performant que l'État, et il

est difficile d'imaginer une parade face à ce géant. Pourtant, Open Street Map propose une alternative viable : ce sont les citoyens eux-mêmes qui cartographient les lieux qu'ils connaissent, en partenariat avec le service de cartographie d'État et les cadastres mis à disposition par les collectivités. On cumule ainsi des données justes, contextualités, sur des fonds de cartes ouverts et exploitables par qui en a besoin.

1.2. L'open data au service d'optimisation des villes

En plus d'être un nouvel outil pour la démocratie, l'open data propose une meilleure compréhension et optimisation des flux dont notre société se compose. Une nouvelle façon de gérer la ville se profile grâce à l'analyse des déplacements des habitants, de la gestion des transports, de la pollution, ainsi que tout paramètre accessible sous forme de données. Bien entendu, une partie conséquente des données n'est pas ouverte pour l'instant. Cependant, les collectivités, les services publics et certaines entreprises réalisent peu à peu l'intérêt de fournir des jeux de données à la société civile. C'est le phénomène des « smart cities ». Grâce à des combinaisons de capteurs, au volontariat des habitants ou aux traces qu'ils laissent en parcourant la ville, il devient possible de mieux comprendre les comportements, de les prévoir et d'aménager la ville en fonction.

Pour en revenir à Open Street Map, l'intérêt de la participation citoyenne via l'open data a été démontré lors du séisme à Haïti, en 2010. En moins d'un mois, quelque six cents personnes ont participé à la cartographie du pays sinistré. Ce travail a été d'une grande aide pour les secours : en connaissant les routes praticables, les villages isolés, les endroits sécurisés, les associations humanitaires ont pu être plus efficaces [Beckett,2004].

Dans un autre genre, les Transports Publics Genevois ont ouvert plus de mille jeux de données depuis septembre 2013. Toute personne qui en a l'envie et les capacités techniques peut en faire une application utilitaire, une datavisualisation, une œuvre d'art ou quoi que ce soit d'autre. Par exemple, l'application « UncrowdTPG », qui indique quels bus sont surchargés, a été créée grâce aux données ouvertes.

« Sense your City » de Data Canvas est aussi un exemple de projet qui illustre le développement des smart cities. L'organisation a fait émerger des communautés dans sept villes à travers le globe, leur a fourni des kits de capteurs à construire soi-même, qui recueillent en temps réel humidité, pollution, bruits, luminosité, et d'autres informations. Ces

informations sont automatiquement publiées sur une base de données en ligne, sous forme d'open data. Des outils de visualisation de ces données sont aussi proposés sur le site, permettant ainsi de voir directement ce qui est capté. Une fois la récolte commencée, tout est possible : croiser les bases de données, les villes, les paramètres, pour mieux appréhender la vie de la cité.

2. Utilisation des LOD en milieu éducation

2.1. L'open data au service de diffusion du savoir

D'une façon générale, l'open data a cette capacité de connecter les savoirs. Le souhait de mettre en commun les connaissances est présent dans de nombreux milieux. Une prise de conscience se fait peu à peu : ouvrir des données, c'est multiplier les chances d'en tirer des externalités positives auxquelles le premier détenteur n'aurait pas pensé. Les scientifiques sont parmi les premiers à avoir ouvert et connecté leurs bases de données pour s'appuyer sur des expériences déjà réalisées, des résultats fiables et vérifiables. Au niveau des institutions muséales, bibliothèques, archives et galeries, les mêmes possibilités d'ouverture voient le jour malgré les obstacles juridiques (droit d'auteur par exemple) et techniques. René Schneider, professeur à la Haute École de Gestion de Genève et spécialiste des linked open data explique que les bibliothécaires travaillent depuis des siècles sur les données et leur catalogage, mais de façon fermée. Grâce à l'open data, il estime qu'ils ont désormais le pouvoir et le devoir de diffuser le savoir accumulé. L'Open Knowledge Foundation est l'un des acteurs open data le plus investi dans l'ouverture de la connaissance.

2.2. L'intérêt de linked open data sur la recherche d'information

Les LOD représentent un thème évoqué de plus en plus fréquemment en recherche d'information, mais dont la substance est parfois difficile à saisir. L'aspect très technique qui caractérise le web sémantique, ainsi que le jargon qui l'accompagne, contribuent largement à en faire une thématique mal comprise parmi les professionnels de l'information.

Afin de concrétiser ce sujet, nous avons réalisé une étude proposant une analyse de diverses applications déjà existantes et basées sur les LOD en bibliothèque. D'après les

résultats de ce travail, quatre grandes catégories de plus-values ont été identifiées. Elles sont présentées ci-dessous, illustrées par des exemples réels.

2.2.1 Une meilleure interopérabilité des données de la bibliothèque

Les données des bibliothèques sont actuellement enregistrées sous forme de notices en format MARC. Il s'agit d'un standard métier établi dans le monde entier depuis plus de 50 ans. Néanmoins, du fait de son utilisation en bibliothèque uniquement, ce format pose certains problèmes d'interopérabilité. Les autres acteurs de l'information utilisent actuellement des standards basés sur XML ou sur d'autres formats plus modernes.

MARC s'appuie fortement sur les chaînes de caractères : trop peu de champs sont contrôlés ou utilisent des identifiants, ce qui limite le traitement automatique des données par des machines. De plus, ce format est difficilement extensible et adaptable aux évolutions. En effet, MARC est un standard ancien et répandu dans le monde entier. « [S]on adaptation à de nouveaux besoins a conduit à le complexifier et le risque d'introduire une divergence avec les données existantes freine les évolutions majeures ».

Les LOD se basent par contre sur RDF, un standard du web flexible et extensible. Ainsi, des données bibliographiques de structures très diverses (MARC, Dublin Core, EAD, etc.) peuvent être converties en un format unique basé sur RDF. Dans ce cas de figure, il est possible de proposer une recherche fédérée bien plus performante que lors de l'interrogation de plusieurs bases de données séparées. C'est ce qu'a réalisé avec succès la Bibliothèque nationale de France, à travers son service data.bnf.fr (<http://data.bnf.fr/>), qui permet de rechercher – au moyen d'une seule requête – dans le catalogue des imprimés, la bibliothèque digitale et d'autres données non issues des catalogues traditionnels.

L'utilisation d'un standard répandu sur le web et d'une licence ouverte permet également aux données des bibliothèques d'être réutilisées par des acteurs externes au domaine. Au final, cette meilleure interopérabilité peut déboucher concrètement sur des gains de temps et d'argent pour les bibliothèques.

2.2.2 Une meilleure possibilité de recherche

Les LOD se caractérisent par des liens avec d'autres jeux de données. Grâce à ces informations externes supplémentaires, des requêtes plus précises sont rendues possibles. L'enrichissement par des données géographiques telles que latitudes et longitudes permettent ainsi la mise en place de fonctionnalités de recherche au moyen d'une carte.



Figure 04 : Recherche géographique sur le site data.bnf.fr

Sur le même principe, des liens avec des thésaurus similaires dans d'autres langues peuvent constituer la base de fonctionnalités de recherches multilingues. Le portail *The European Library* (<http://www.theeuropeanlibrary.org/tel4/>) a implémenté une fonction de ce genre proposant des termes de recherche en trois langues à l'utilisateur. Le portail finlandais Kulttuurisampo propose quant à lui de nombreuses fonctionnalités expérimentales, parmi lesquelles la recherche de relations entre deux personnes comme les peintres Rubens et Miró.

Person 1 Rubens, Peter Paul (flaamilainen taiteilija, 1577-1640)

Person 2 Miró, Joán (espanjalainen taiteilija, 1893-1983)

Rubens, Peter Paul (Flemish painter and draftsman, 1577-1640)
patron was
Louis XIV, King of France (French monarch, collector, and patron, 1638-1715)
knows
Louis XV, King of France (French ruler, patron, and collector, 1710-1774)
grandparent of
Charles X, King of France (French ruler, patron, and collector, 1757-1836)
patron of
Lhuillier, Charles (French painter and sculptor, born 1824, died 1898 or 1899)
teacher of
Dufy, Raoul (French painter, designer, and ceramicist, 1877-1953)
collaborated with
Llorens Artigas, Josep (Spanish ceramicist, 1892-1980)
collaborated with
Miró, Joán (Spanish painter, printmaker, and ceramicist, 1893-1983)

Figure 05 : Recherche de relation entre deux personnes

2.2.3 Une meilleure sérendipité

Au-delà de l'amélioration des fonctionnalités de recherche, les liens vers des jeux de données externes augmentent la sérendipité de l'utilisateur. Plusieurs applications en proposent, par exemple la Deutsche digitale Bibliothek (<https://www.deutsche-digitale-bibliothek.de/>). Ces liens permettent de rediriger l'utilisateur vers des ressources complémentaires, comme des notices d'autorités d'autres bibliothèques ou des articles Wikipédia. VIAF (<http://viaf.org/>), un agrégateur de ressources de bibliothèques sur les personnes et les organisations, permet de faciliter la création de ce genre de relations.

Une institution culturelle ne possède pas que des données bibliographiques. En intégrant tous types de données au sein d'une application unique basée sur RDF, on offre la possibilité à l'utilisateur de trouver des informations qu'il ne cherchait pas forcément. Ainsi, le Centre Pompidou à Paris a converti des données de bibliothèque et de musée en un seul format, et les a reliées grâce à RDF. De cette manière, une seule recherche donne accès à des livres, des œuvres d'art numérisées, des dossiers pédagogiques ainsi que d'autres types de documents.

MAX ERNST

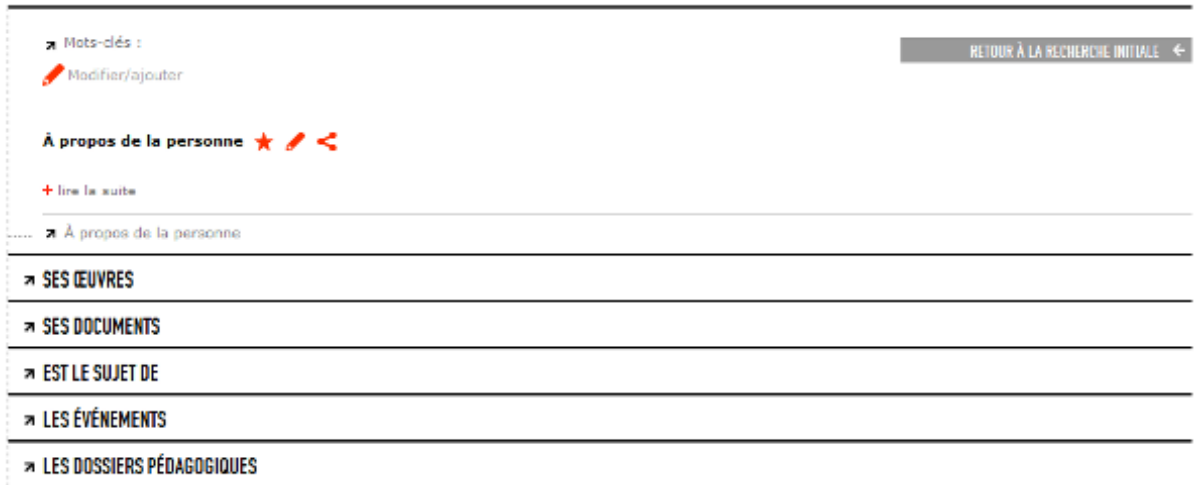


Figure 06 : Une seule interface pour les ressources variées du Centre Pompidou, ici sur Max Ernst

La variété des données liées ouvre des possibilités intéressantes de représentation de l'information. On s'oriente alors vers une recherche par navigation plutôt que par requêtes comme cela se fait dans les catalogues traditionnels. Dans ce sens, l'application data.bnf.fr propose la recherche géographique, mais également des représentations chronologiques des données.

Les méthodes de recherche évoluent également vers de nouvelles fonctionnalités d'auto-complétion, notamment grâce aux classes sémantiques attribuées aux données. L'application bêta d'Europeana (<http://eculture.cs.vu.nl/europeana/>) offre ainsi des propositions de types d'entités au cours de la saisie de la requête

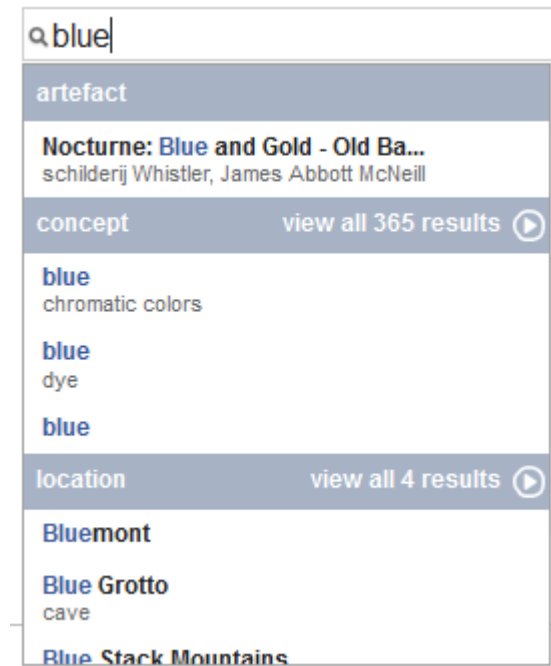


Figure 07 : Recherche par type d'entité

2.2.4 Une meilleure visibilité de données sur le web

Se basant sur des standards du web, les LOD proposent une interopérabilité appréciée par les moteurs de recherche, qui ont plus de facilité à indexer les contenus (Ostrowski & Pohl, 2012 : p. 272). A l'inverse, les données MARC paraissent en silos sur le web, car elles ne sont accessibles qu'au travers de portails, et ne communiquent pas avec le reste du web. L'utilisation d'identifiants uniques et pérennes pour les LOD permet de créer ces relations manquantes entre les ressources. Grâce à cela, des liens tiers pointant vers le catalogue améliorent la visibilité des données des bibliothèques. Par exemple, près de 80 % des visiteurs du service data.bnf.fr proviennent des moteurs de recherche, alors même que le catalogue traditionnel n'y est absolument pas visible (Simon, 2014)

Introduction :

Les ressources pédagogiques correspondent à tout ce qui concerne l'enseignement et les apprentissages dans leur globalité. Elles se réfèrent en particulier au contenu pédagogique.

Les matériaux sont donc créés, sélectionnés et édités spécifiquement pour les apprenants et les enseignants. Elles peuvent contenir à la fois du texte et à la fois des multimédias. Elles ne sont pas nécessairement limitées à du simple matériel de lecture. Leur taille et leur nature sont variées et peuvent se présenter sous différentes formes. Elles peuvent jouer un rôle différent selon les besoins des acteurs et sont autonomes. En définitive, les ressources pédagogiques constituent l'ensemble des informations et des matériaux qui soutiennent l'enseignement et permettent les apprentissages.

1. Ressource pédagogique

Certains chercheurs s'accordent à dire qu'une ressource pédagogique est fortement liée aux objets pédagogiques. Cette liaison semble à la classification de Pernin qui a spécialisé trois principales catégories d'objets pédagogiques :

- Les Unités d'Apprentissage (UA), qui permettent de structurer la formation et de l'organiser dans l'espace et dans le temps ;
- Les Activités Pédagogiques (AP), qui définissent les modalités précises d'acquisition, de validation, de communication d'une ou de plusieurs connaissances ;
- Les ressources pédagogiques, physiques ou numériques, nécessaires à la réalisation des activités ;

D'après cette classification nous remarquons que les ressources pédagogiques sont une classe ou spécificité d'objets pédagogiques.

1.1. Quelques définitions de ressources pédagogiques :

Gomez définit une ressource pédagogique comme étant toute entité (numérique ou non), utilisée dans un contexte d'enseignement, de formation ou d'apprentissage et qui est durable (doit pouvoir résister aux changements technologiques), adaptable (doit permettre une modulation sur mesure c'est-à-dire tenant en compte les spécificités physiques des apprenants), gérable (doit pouvoir être administrée facilement), fiable (doit répondre aux exigences pour lesquelles on l'a créée), abordable (doit pouvoir être accessible), évaluable (doit pouvoir être estimée facilement), interopérable (permettre l'utilisation de contenus et composants développés par une organisation sur une plate-forme donnée par d'autres organisations sur d'autres plates-formes), retrouvable (doit pouvoir être accessible facilement), réutilisable (doit pouvoir être utilisé dans des contextes et des buts différents), indexable (doit être décrite par des métadonnées ou des annotations sémantiques) **(PETER-SAHA, 2010)**.

Une définition très large, même si le contexte pédagogique apparaît de manière plus forte de Gérard Puimatto **(PUIMATTO, 2004)** « La notion de ressource pédagogique (ou éducative) est fortement liée à celle de document, et aux approches documentaires. Dans le domaine du multimédia, les "ressources multimédias pour l'éducation" constituent l'ensemble des informations, documents, logiciels, programmes, banques de données, etc. qui permettent de véhiculer, de transmettre ou d'appréhender des concepts et contenus d'enseignements »

Robert Bibeau propose encore une définition des ressources numériques pour l'éducation : « Les ressources numériques pour l'éducation et la formation (ressources pédagogiques) ce sont tous les logiciels d'édition et de communication (portails, moteurs de recherche, répertoires, logiciels outils, applicatifs de formation) ainsi que les données, les informations et les œuvres numérisées (données statistiques ou informationnelles, références générales, œuvres littéraires, artistiques ou autres) utiles à l'enseignant ou à l'apprenant dans le cadre d'une activité d'enseignement ou d'apprentissage.»

2. L'objet pédagogique

Aujourd'hui, le terme *d'objet pédagogique* (ou learning object en anglais) est devenu central sur le terrain des environnements informatiques pour l'apprentissage humain et fait

l'objet de nombreux travaux au sein des instances internationales de normalisation. Le concept d'objet pédagogique parvient à fédérer les efforts poursuivis par différents groupes d'intérêt parce qu'il rassemble un certain nombre d'atouts reconnus à différents niveaux économique, pédagogique ou technique. Au niveau économique, au cours des dernières décennies, le développement de produits ou services reposant sur l'assemblage et la réutilisation de composants s'est largement répandu. Au niveau pédagogique, le concept d'objet s'accorde bien avec les notions de "formation tout au long de la vie" et de "parcours individualisé à la carte". Construire une offre de formation revient alors à assembler un ensemble de composants adaptés aux besoins spécifiques de l'apprenant. Au niveau technique, les apports de l'approche par objets dans le domaine du génie logiciel sont indiscutables. L'application de l'approche par objets dans le domaine des composants pédagogiques a contribué à la volonté d'en décrire précisément les caractéristiques et les services afin d'en assurer le partage et la réutilisation [Koper, 01].

2.1. Quelques définitions:

Selon le groupe de travail IEEE (Institute of Electrical and Electronics Ingeniers) connus sous le nom LTSC (Learning Technologies Standards Committee), auteur du LOM (Learning Object Metadata) : un objet pédagogique peut être défini comme « toute entité numérique ou non, qui peut être utilisée, réutilisée ou référencée lors d'une formation dispensée à partir d'un support technologique» (traduction proposée par l'Association des Informaticiens de Langue Française¹⁰).

Flory (**FLORY, 2004**) a ajouté à la définition d'IEEE que l'objet pédagogique est utilisé dans différents contextes. L'enseignant va constituer son cours en assemblant les différents éléments, comme des fragments de texte, des annotations, des images, des sons, des simulations... qui vont constituer des chapitres, puis des cours ou modules et enfin des formations complètes, comme l'illustre le schéma suivant:



Figure 08 : La pyramide des objets pédagogiques.

Philippe Parmentier dans sa définition dite que « Chaque objet pédagogique doit être durable, adaptable, gérable, fiable, abordable, évaluable, interopérable, retrouvable, réutilisable, indexable ». (FLORY, 2004)

3. Les ressources pédagogiques face à l'apprentissage en ligne

L'apprentissage en ligne est une méthode d'apprentissage reposant sur la mise à disposition de contenus pédagogiques à travers des scénarii pédagogiques dans un environnement numérique. Outre l'élaboration des contenus, la création de telles ressources pose de nombreuses difficultés à ses concepteurs. L'utilisation de ces ressources dans des scénarii pédagogiques implique non seulement que la ressource soit découpée en objets pédagogiques cohérents mais également les activités associées aux objectifs pédagogiques soient définies. Les ressources existantes sont difficilement réutilisables car elles n'ont pas été conçues dans cet objectif ; elles correspondent souvent à de simples présentations en ligne de documents qui n'ont pas été créées spécifiquement pour leur exploitation dans des environnements d'apprentissage. (Psyché et al., 2005) constate l'insuffisance ou l'absence de l'application d'une approche pédagogique, que ce soit au niveau de la présentation des ressources pédagogiques ou du séquençement des activités d'apprentissage dans les outils actuels. D'autre part, pour être ré-utilisable, une ressource doit être indexée de façon à en connaître à la fois le contenu, les objectifs pédagogiques visés, les conditions de réutilisabilité, etc...

3.1. L'apprentissage en ligne :

L'apprentissage en ligne est une activité pédagogique qui vise à acquérir ou à approfondir des connaissances tout en repoussant les contraintes de temps et d'espace entre l'apprenant et l'enseignant, par l'utilisation des nouvelles technologies de l'information et de la communication (E-TUD) (**Boutmedjet, 2004**).

Un système d'apprentissage en ligne doit permettre:

Pour les enseignants :

- l'accès aux ressources pédagogiques pertinentes grâce à une bonne indexation des ressources (**Gasevic, 2005**), (**Psyche, 2005**), (**Lenne, 2005**), (**Abel, 2003**) ;

- la réutilisabilité des objets et des scénarii pédagogiques (**Knight, 2005**),
- la conception et la mise à jour des ressources pédagogiques par les enseignants (**Lenne, 2005**), (**Abel, 2003**) ;

Pour les apprenants :

- une interaction et navigation au sein des ressources pédagogiques suivant une pédagogie d'apprentissage adéquate mise en place (**Psyche, 2005**) ;
- le suivi individualisé des apprenants (**IMSLD, 2003**).

4. les normes d'apprentissage en ligne

Les normes actuelles d'e-learning offrent une solution partielle pour la ré-utilisabilité. En particulier, SCORM (**SCORM, 2004**) et LOM (**LOM, 2002**) aident à l'homogénéisation des représentations des ressources pédagogiques et facilitent l'interopérabilité. LOM propose

différentes méta-données nécessaires pour la description des ressources pédagogiques mais cette description n'est pas suffisante, il y manque la représentation sémantique des contenus.

SCORM de son côté, permet la structuration des contenus des ressources pédagogiques et leurs relations avec l'environnement d'utilisation. Enfin, IMS-LD (IMSLD, 2003) considère la pédagogie d'apprentissage et de son déroulement en intégrant la notion de scénario pédagogique. Cependant, les représentations obtenues via ces normes ne sont pas suffisantes pour permettre et assurer la ré-utilisabilité de ressources ou de parties de ressources car elles ne prennent pas en compte ni le contenu sémantique des méta-données qu'elles introduisent ni les relations qui existent entre chaque objet pédagogique.

4.1 LOM

Le standard LOM spécifie la syntaxe et la sémantique des méta-données décrivant des ressources pédagogiques numériques ou non et définit les attributs nécessaires à une description complète des ressources pédagogiques à partir des 78 attributs divisés en neuf catégories :

1. Généralité : caractéristiques indépendantes du contexte comme l'identifiant, le titre, la langue de la ressource...
2. Cycle de vie : caractéristiques relatives au cycle de vie, comme la version ou l'état (brouillon, final, révisé, non disponible) ;
3. Méta-métadonnées : caractéristiques de la description elle-même comme : la contribution (personnes ayant participé à l'élaboration), catalogue, ...
4. Technique : caractéristiques techniques comme le format (du logiciel nécessaire pour accéder à la ressource), taille...
5. Pédagogie : caractéristiques pédagogiques : type d'interactivité, type de ressource; niveau d'interactivité, densité sémantique, rôle présumé de l'utilisateur final, contexte, tranche d'âge, difficulté, temps d'apprentissage moyen, description, langage ;
6. Droits : coûts, copyrights, description...
7. Relation : caractéristiques exprimant les liens avec d'autres ressources ;

8. Annotation : commentaires sur l'utilisation pédagogique de la ressource

9. Classification : caractéristiques de la ressource décrites par des entrées dans des systèmes de classification : but, classification de références, chemin...

LOM est très complet et permet d'identifier tout type de contenu, et plus spécifiquement les contenus éducatifs, avec un format commun décrivant le type de contenu, son auteur, la meilleure manière de l'utiliser, etc. Les profils d'applications ont été introduits afin de répondre aux besoins spécifiques et concrets des utilisateurs. Un profil d'application est une instance d'un modèle comme LOM dans un contexte particulier. De fait, cela signifie interpréter, raffiner, étendre ou parfois même simplifier les syntaxes et les sémantiques des méta-données LOM.

4.2. SCORM

SCORM (Sharable Content Object Reference Model) (**SCORM, 2004**) de Advanced Distributed Learning (ADL) est une référence pour le partage de contenus et d'objets.

SCORM permet l'assemblage des contenus web et un environnement d'apprentissage pour les objets pédagogiques. Il a pour vocation la mise en place de la bonne structuration du contenu du cours et de ses interactions avec son environnement.

SCORM traite les éléments suivants :

- Packaging : il a pour objectif la transmission d'un contenu d'une plate-forme vers une autre, l'importation ou l'exportation de contenus d'objets pédagogiques pour les mettre à disposition d'autres. Il s'intéresse également à la structuration des objets pédagogiques ;
- Méta-données : elles sont issues de LOM et ont pour objectif de partager les informations standards qui décrivent la nature et l'objectif du contenu.
- Communication ou environnement d'exécution : détermine la communication avec un environnement web. La notion d'environnement est également présente dans IMS LD ;
- Séquencement et navigation : définit une méthode de représentation de la navigation entre objets d'apprentissages. Spécifiquement, il décrit les branchements et le flux d'activités d'apprentissages en termes d'arbre d'activité ;

- Agrégation de contenu : elle distingue trois niveaux de ressources :

N. Hernandez et al.

- -La *ressource numérique élémentaire* (asset) constitue la brique élémentaire : Il peut s'agir d'un document simple (image JPEG ou GIF, son WAV ou MP3, page Web) mais également de tout ensemble d'informations pouvant être délivré vers un client Web (document Flash, code JavaScript, applet Java, etc.),
- Un *objet de contenu partageable* (SCO) est un ensemble cohérent de ressources numériques élémentaires. Il peut être contrôlé depuis une plate-forme d'apprentissage en ligne (LMS). Respectant le protocole d'exécution SCORM, il représente le plus bas niveau de granularité pouvant faire l'objet d'un suivi,
- Un *agrégat de contenu* (Content Aggregation) est un ensemble de ressources pédagogiques structurées de façon cohérente au sein d'une entité de plus haut niveau, telle qu'un cours, un chapitre, un module, etc.

La structuration du contenu des modules d'enseignement suivant la norme SCORM permet de les réutiliser dans d'autres modules pour différentes formations ou systèmes. De plus, elle améliore le dialogue entre les objets pédagogiques et le système d'une part, et entre les acteurs et le système d'autre part. Dans notre modèle, nous utilisons SCORM pour représenter les structures des modules d'enseignement d'une formation et ainsi garantir leur interopérabilité.

4.3 IMS-LD

Un Learning Design ou scénario pédagogique est défini comme une description d'une méthode permettant à un apprenant d'atteindre certains objectifs d'apprentissage par la réalisation de certaines activités pédagogiques ordonnées, dans un environnement d'apprentissage.

En complément de SCORM, qui s'intéresse surtout aux pédagogies de transfert, IMS-LD (**IMS-LD, 2003**) ou Instructional Management System Learning Design est une norme qui vise à apporter des éléments de pédagogie dans un système d'apprentissage en ligne. Il s'agit d'un langage de modélisation des processus d'apprentissage basé sur les travaux de Koper (**Koper, 2001**). Il a été conçu pour la définition de scénarii d'apprentissages et d'interaction

Chapitre 3 : Conception d'un LOD disciplinaire basé sur l'ontologie de domaine ONTO-TDM

pour les créateurs de contenu ou de cours. Il aide les concepteurs à modéliser : qui fait quoi, quand et avec quelles ressources et quels services pour réaliser des objectifs d'apprentissages. Il définit la structure d'une unité d'apprentissages comme « pièce » : un ensemble « d'actes » composés de « partitions » associant des « activités » à des « rôles » (enseignant, apprenant,...).

Pour faciliter la production de la spécification ainsi que son implémentation, LD a été divisé en trois parties :

- Niveau A : contient l'ensemble des structures de base incluant : Activités, Environnements, Pièces, Actes, Rôles, Services ;
- Niveau B : ajoute des Propriétés et Conditions au niveau A. Ceci permet la personnalisation, le séquençement et l'interaction plus élaborés en se basant sur le profil de chaque apprenant ;
- Niveau C : ajoute les notifications au niveau B. Une notification est déclenchée par la réalisation d'un Résultat, et rend une activité exécutable et disponible pour un Rôle.

Chaque niveau est représenté dans des fichiers XML séparés. Le niveau B et C intègrent et étendent le niveau précédent.

Dans notre modèle, nous nous appuyons sur IMS-LD pour définir le déroulement des interactions Homme-Machine pendant la phase d'exécution et d'utilisation des objets pédagogiques. Les normes définies dans le contexte de l'apprentissage en ligne permettent de s'assurer d'une certaine interopérabilité et utilisabilité au travers de scénarii.

5. L'apprentissage en ligne et les ontologies

L'utilisation des méta-données telle qu'elle est préconisée par les normes ne suffit pas à résoudre les problèmes de ré-utilisabilité et accessibilité. En effet, ces problèmes ne seront résolus que si le système et les acteurs partagent le même sens accordé aux valeurs des métadonnées.

D'autre part, les liens et relations comme la composition, l'ordre d'apprentissage, et les dépendances de pré-requis entre objets pédagogiques doivent être mentionnés pour permettre non seulement de réaliser des traitements ou tâches automatiques sur ces objets mais aussi faciliter leur réutilisation. L'utilisation des ontologies dans le modèle sous-jacent à un système

d'apprentissage en ligne sont une solution intéressante. En effet, une ontologie regroupe les concepts qui représentent l'ensemble des connaissances d'un domaine en une

spécification explicite et formelle (**Studer et al., 1998**). Elle montre les relations ainsi que les règles d'associations qui existent entre ces concepts. Elle permet ainsi, d'une part à l'ordinateur la production de nouvelles connaissances par le biais d'inférences, et d'autre part à l'homme et à l'ordinateur d'accorder des sens communs aux termes utilisés dans un domaine d'activité afin de lever toute ambiguïté pendant les traitements. Ainsi, une ontologie peut fournir le vocabulaire et sa sémantique, pour exploiter le contenu des méta-données associées aux ressources annotées. Différents travaux de la littérature ont étudié l'utilisation des ontologies.

5.1 Représentation des ressources pédagogiques en se basant sur Ont-TDM

On utilise dans notre cas pour la représentation des ressources pédagogiques de suivre les concepts de l'ontologie d'un domaine d'enseignement Ont-TDM qui sont :

- Les notions : Ce sont les composantes du domaine d'enseignement, elles peuvent jouer le rôle de titre ou racine de la ressource.
- Les sous-notions : une notion peut être décomposée en plusieurs sous-notions pour des raisons pédagogiques ou structurelles et une sous-notion peut jouer le rôle d'une notion.
- Les items de connaissances : un item de connaissance est la sous-notion indécomposable qui se trouve au plus bas niveau de la représentation.

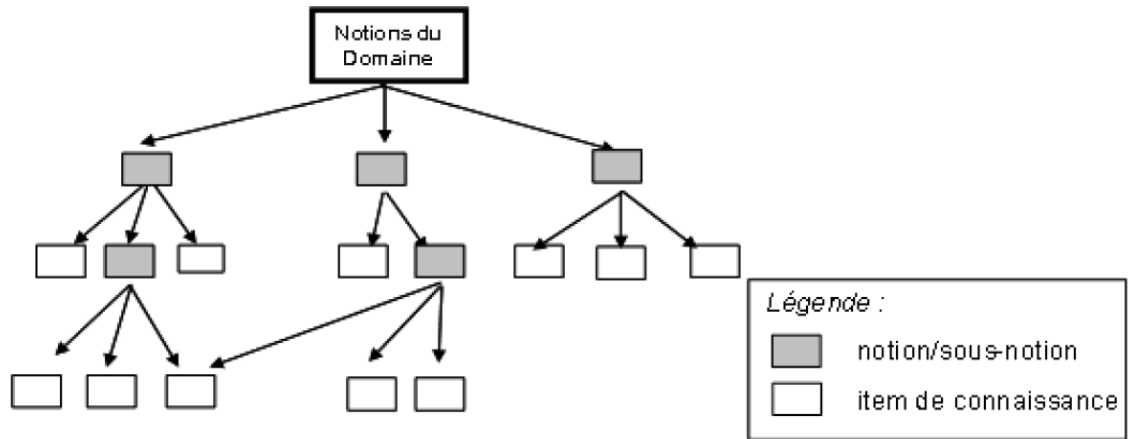


Figure 09 : Décomposition des notions du domaine d'enseignement

Introduction :

Le Linked Open Data consiste à publier des données sur le web de telle sorte qu'elles puissent être interprétées et connectées entre elles. Il est donc vital d'établir les liens entre ces données à la fois pour le web des données et pour le web sémantique qu'il contribue à nourrir. Nous proposons un cadre général dans lequel s'inscrivent les différentes techniques utilisées pour établir ces liens et nous montrons comment elles s'y insèrent. Nous proposons ensuite une architecture permettant d'associer les différents systèmes de liage de données et de les faire collaborer avec les systèmes développés pour la mise en correspondance d'ontologies qui présente de nombreux points communs avec la découverte de liens.

La conception de toute solution logicielle doit être traitée avec précision et détail, précédée d'une analyse profonde et bien réfléchie, car elle est le reflet du futur système avant même sa concrétisation. Dans le but d'avoir une meilleure analyse et rendre la conception de notre plate forme plus complète.

1. Planification stratégique

1.1 Objectif du projet

Notre projet consiste à la création d'une plate forme open data sets, c'est pour cela qu'on doit réaliser une application web sur laquelle on pourra implémenter les meilleures techniques de gestion et de liaison de donnée, des ontologies pour garantir une meilleure représentation de données et le RDFa pour lier les données entre elles.

C'est dans ce sens que nous allons concevoir et réaliser notre plate forme open data sets traitant les données du département informatique de l'Université Mouloud Mammeri de Tizi_Ouzou, en garantissant l'accès ouvert aux données et une bonne structuration. Pour cela nous respectant les différents principes de LOD définie par le W3C.

1.2 Connaitre son environnement

1.2.1 Notre plateforme face aux données du département informatique de l'UMMTO

Les données du département informatique de l'UMMTO sont organisées en plusieurs catégories selon le niveau d'étude et la spécialité de la formation.

Chapitre4 : Implémentation d'un prototype d'un site LOD pour les ressources disciplinaire

Chaque niveau d'étude contient ses propres modules et chaque spécialité d'étude a aussi ses propres modules.

Le département informatique est composé de cinq années d'études :

- Licence 1
- Licence 2
- Licence 3
- Master 1
- Master 2

Et composé de 6 spécialités d'études :

- Licence informatique générale
- Licence 3 Multimédia
- Master système informatique (SI)
- Master ingénierie des systèmes d'informations (ISI)
- Master conduite de projet informatique (CPI)
- Master réseau et système embarquer

Chaque module de chaque spécialité représente un data set dans notre plateforme et le contenu du module sera structuré suivant l'ontologie disciplinaire ONT_TDM et les ressources seront représentées sous forme d'URIs.

1.2.2 Les enseignants du département informatique de L'UMMTO

Le département informatique de l'UMMTO contient plusieurs enseignants et chaque enseignant assure un ou plusieurs modules. Un enseignant possède ses cours.

Sur notre plateforme un enseignant sera le propriétaire du dataset c'est lui qui structure le dataset, qui insère les ressources relatives au cours. Les enseignants de l'université Mouloud Mammeri de Tizi-Ouzou sont classifiés selon leur niveau.

Les niveaux existant :

- Vacataire
- Maître assistant A
- Maître assistant B
- Maître de conférences A
- Maître de conférences B
- Professeur

2. Conception

2.1 Présentation des fonctions de base de la plateforme

- Créer un compte enseignant
- Création des data sets (par l'enseignant)
- Gestion des data sets (par L'enseignant)
- Création des ressources (par l'enseignant)
- Consultation des ressources (par les différents utilisateurs)

2.2 Architecture de l'information :

2.2.1 organisation du contenu :

a. Enumération du contenu

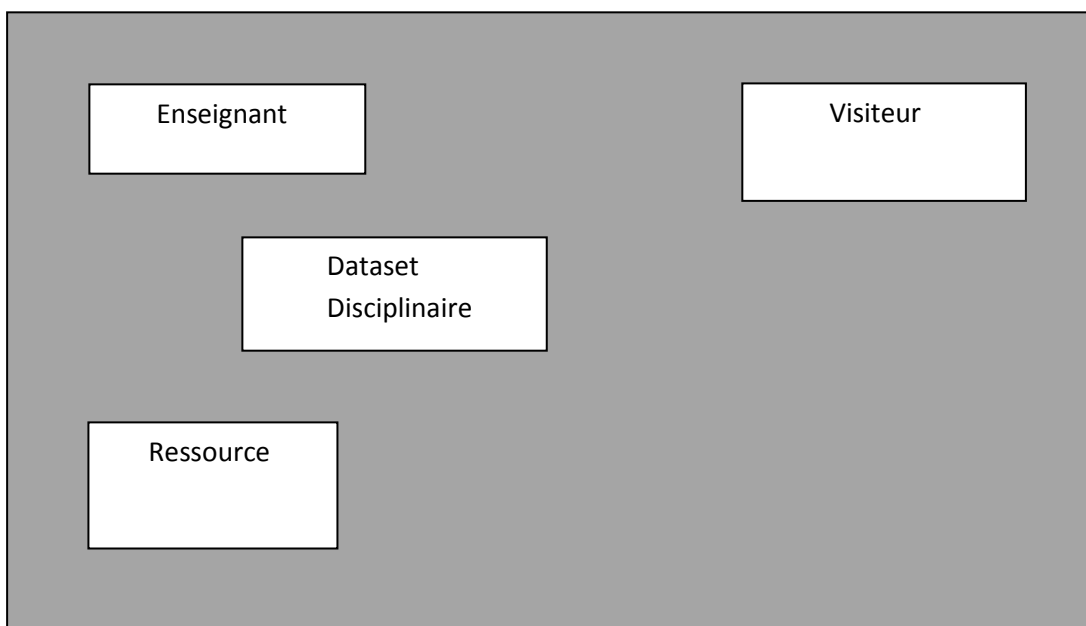


Figure 10 : Enumération du contenu

b. Structure de catégorisation

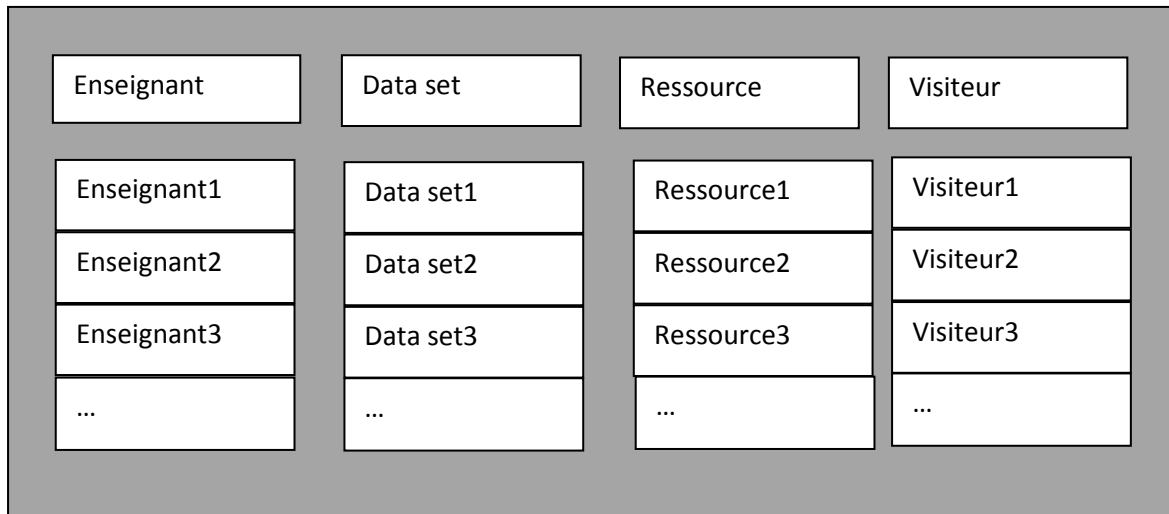


Figure 11 : Catégorisation du contenu

c. Structure Métadonnées du Dataset

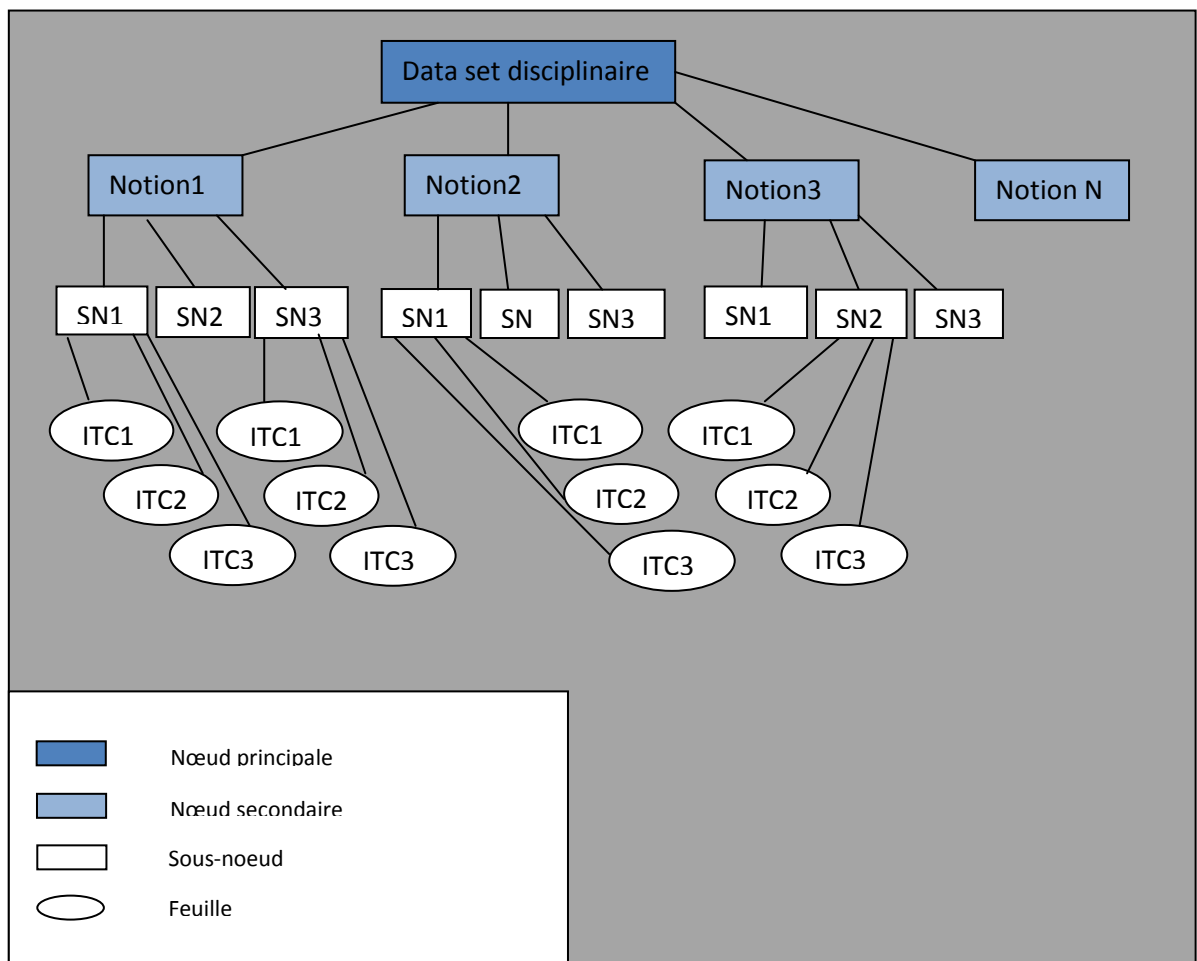


Figure 12 : structure méta data du data set

d. Structure externe du contenu

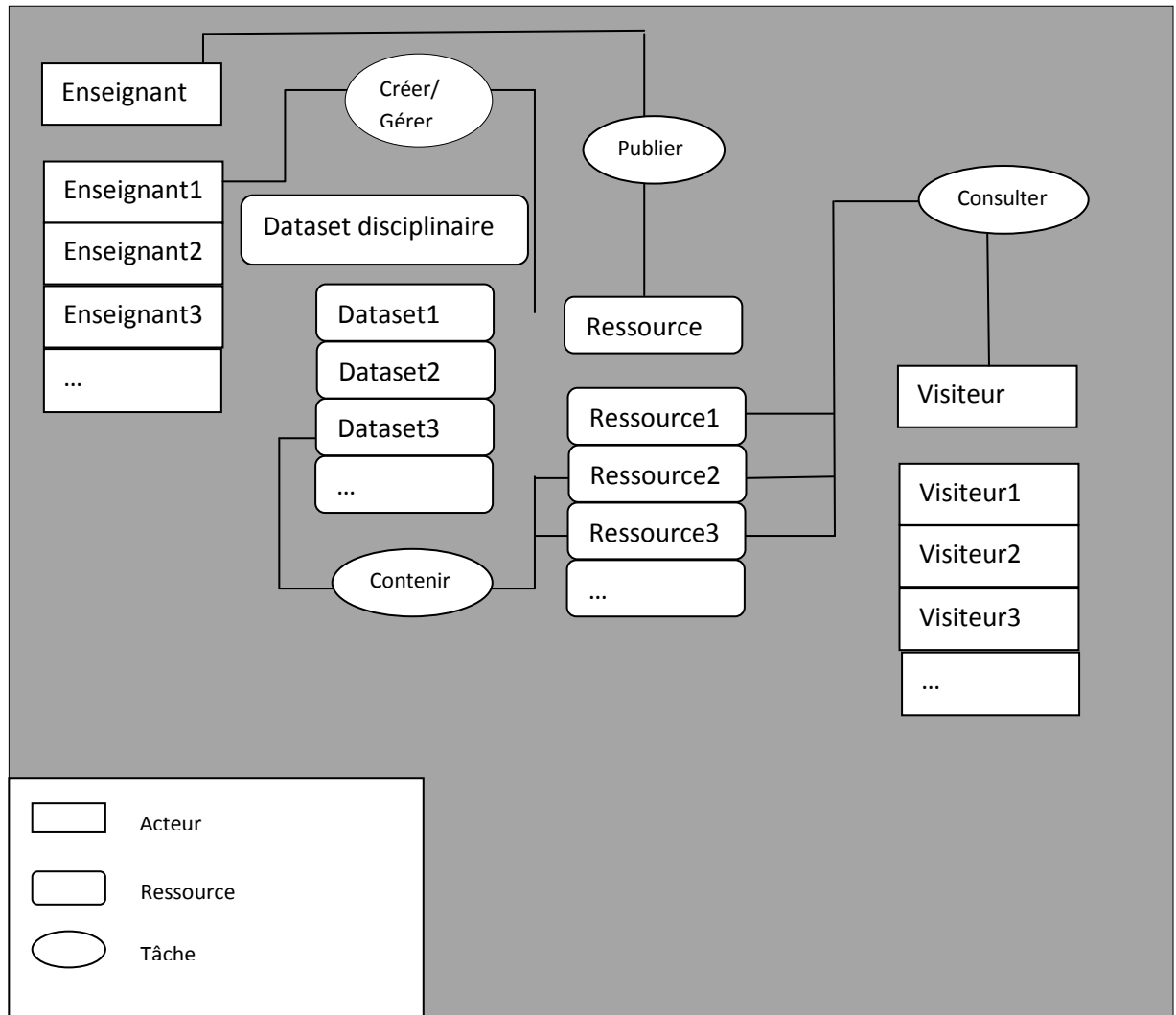


Figure 13 : Structure externe du contenu

Chapitre4 : Implémentation d'un prototype d'un site LOD pour les ressources disciplinaire

a. Structure interne du contenu

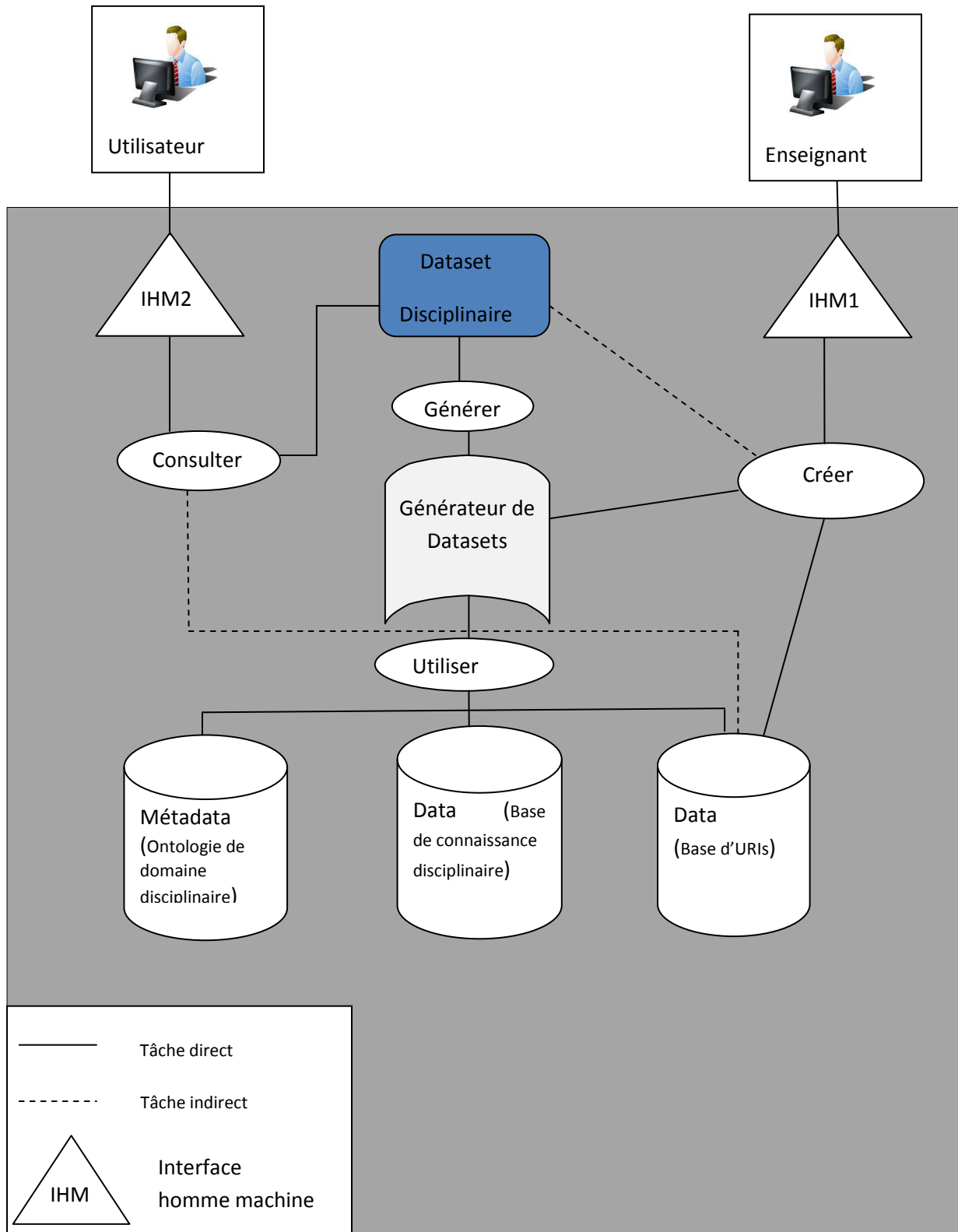


Figure 14 : Structure interne du contenu

2.2.2 Plan de la plate forme :

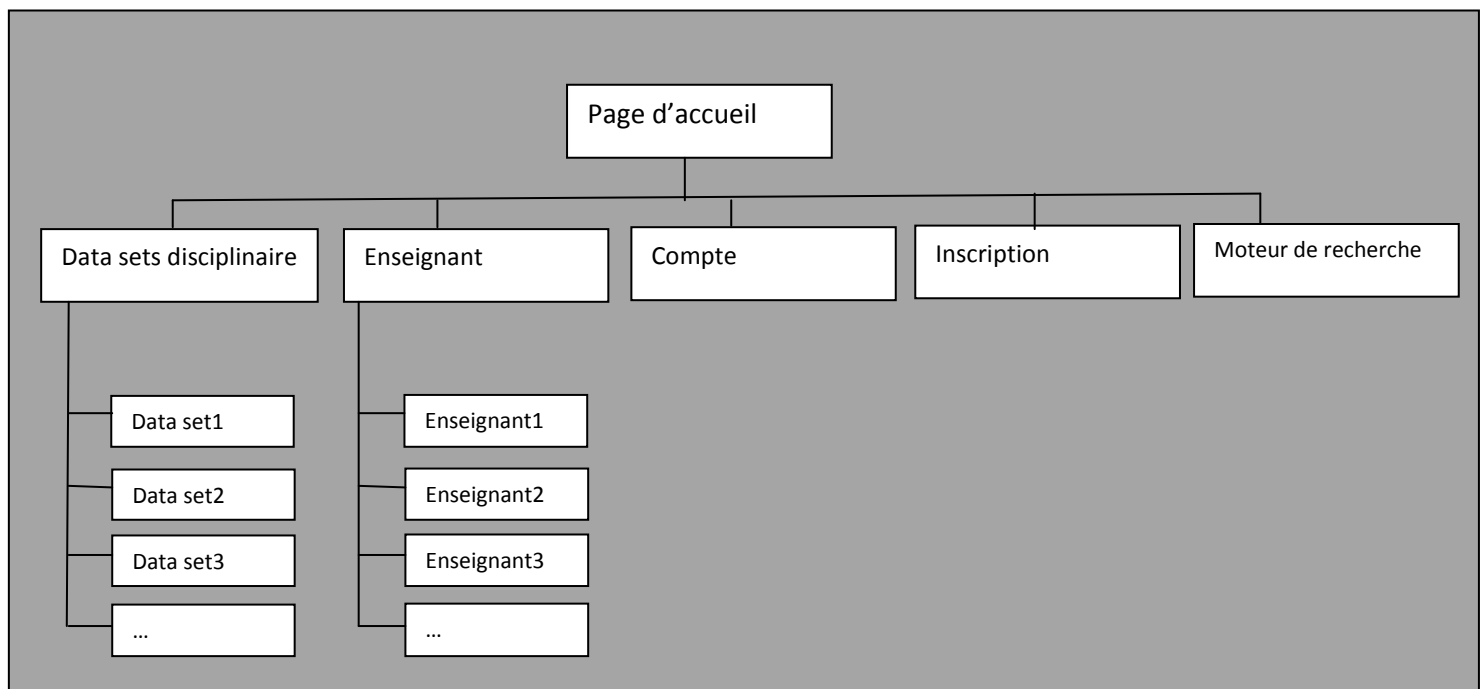


Figure 15 : Plan de la plateforme

2.2.3 conception de la base de données

a. Présentation du dictionnaire de données

Désignation des rubriques	Code	Type	Taille
Identifiant du dataset	ID_ds	N	20
Identifiant d'enseignant	ID_en	N	20
Identifiant de la ressource	ID_res	N	20
Intitulé du dataset	Intitule_ds	AN	50
Nom utilisateur d'enseignant	User_ens	AN	50
Nom de l'enseignant	Nom_ens	A	50
Prénom de l'enseignant	Prénom_ens	A	50
Date de naissance de l'enseignant	DateN_ens	Date	
Lieu de naissance de l'enseignant	LieuN_ens	AN	50
Etat civile de l'enseignant	E_civile_ens	A	50
Adresse de l'enseignant	Adresse_ens	AN	100
E-mail de l'enseignant	E_mail_ens	AN	50

Chapitre4 : Implémentation d'un prototype d'un site LOD pour les ressources disciplinaire

Téléphone de l'enseignant	Telephone_ens	N	20
Grade de l'enseignant	Grade_ens	A	20
Mot de passe du compte enseignant	Password_ens	AN	50
URI de la ressource	URI_res	AN	50

Tableau 01 : Présentation du dictionnaire de données

a. Présentation du dictionnaire de métadonnées

Désignation des rubriques	Code	Type	Taille
Identifiant de la notion	ID_n	N	20
Identifiant de la sous-notion	ID_sn	N	20
Identifiant d'item de connaissance	ID_itm	N	20
Identifiant du dataset	ID_ds	N	20
Notion	Notion	AN	50
Sous-notion	SousNotion	AN	50
Item de connaissance	ITconnai	AN	50

Tableau 02 : Présentation du dictionnaire de métadonnées

a. Règle de gestion

Nous donnerons dans ce qui suit quelques règles des gestions liées aux données.

- Une ressource doit avoir son identifiant, son URI et l'identifiant de son créateur.
- Une ressource doit être créée par un seul enseignant.
- Un dataset est créé par un seul enseignant.
- Un dataset est visible par les différents visiteurs.
- Un dataset contient au minimum une ressource.

b. Schéma de la base de données

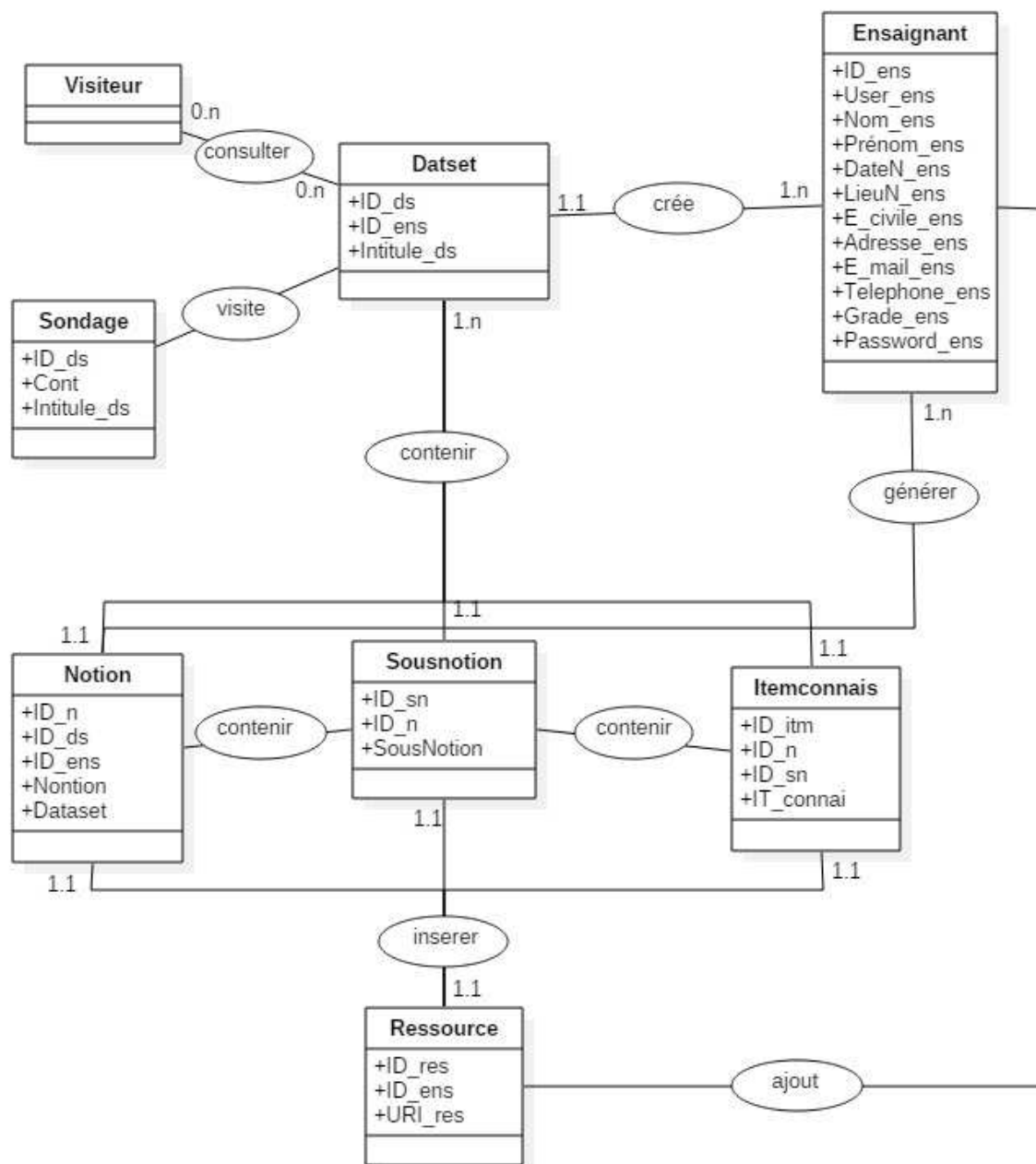


Figure 16 : Entité Association

2.2.4. Conception du générateur de dataset disciplinaire

Le générateur de datasets est un système intelligent pour la création des dataset. Il reçoit comment entrés les ressources relative au dataset, créer par l'enseignant et en se basant sur l'ontologie du domaine disciplinaire et sur les base de connaissances du domaine éducatifs, le générateur fait des traitements et créer le schéma interne du dataset.

Le générateur de datasets créer dynamiquement les structures des datasets et cela grâce à la bonne utilisation des métadonnées.

➤ Création des métadonnées

Lors de la création d'un dataset par un enseignant le générateur récolte le maximum d'information sur le dataset (métadonnées) et en se basant sur l'ontologie de domaine disciplinaire le système génère un schéma à ce dataset.

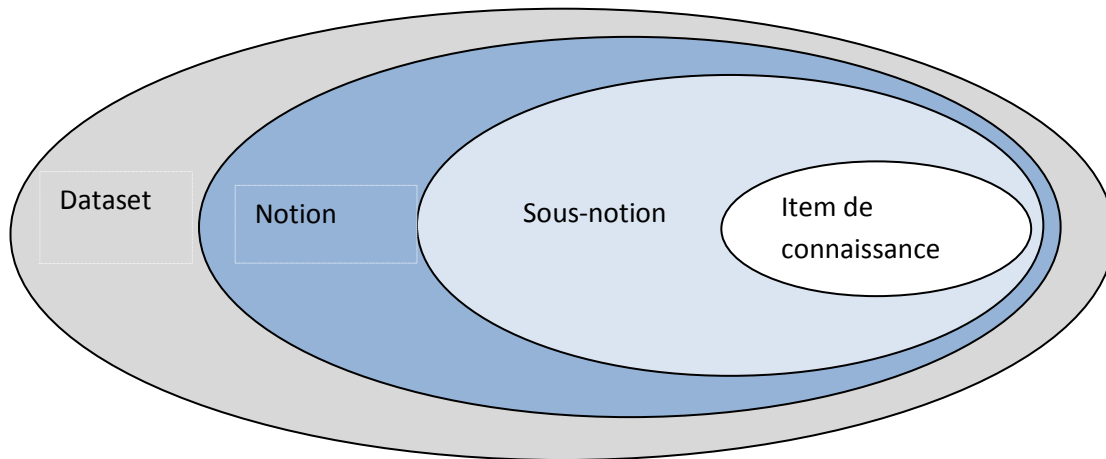


Figure 17 : Représentation des métadonnées d'un dataset

➤ Fonctionnement de générateur de dataset disciplinaire

Le fonctionnement du générateur de dataset se base essentiellement sur la base des connaissances de domaine disciplinaire, sur l'ontologie de domaine disciplinaire et sur les informations délivrées par le créateur. Le système après avoir récolté les métadonnées sur le dataset à créer en se basant sur l'ontologie, le système interroge la base des connaissances pour bien identifier les différentes composantes du dataset et finira par créer une structure valide de ce dernier.

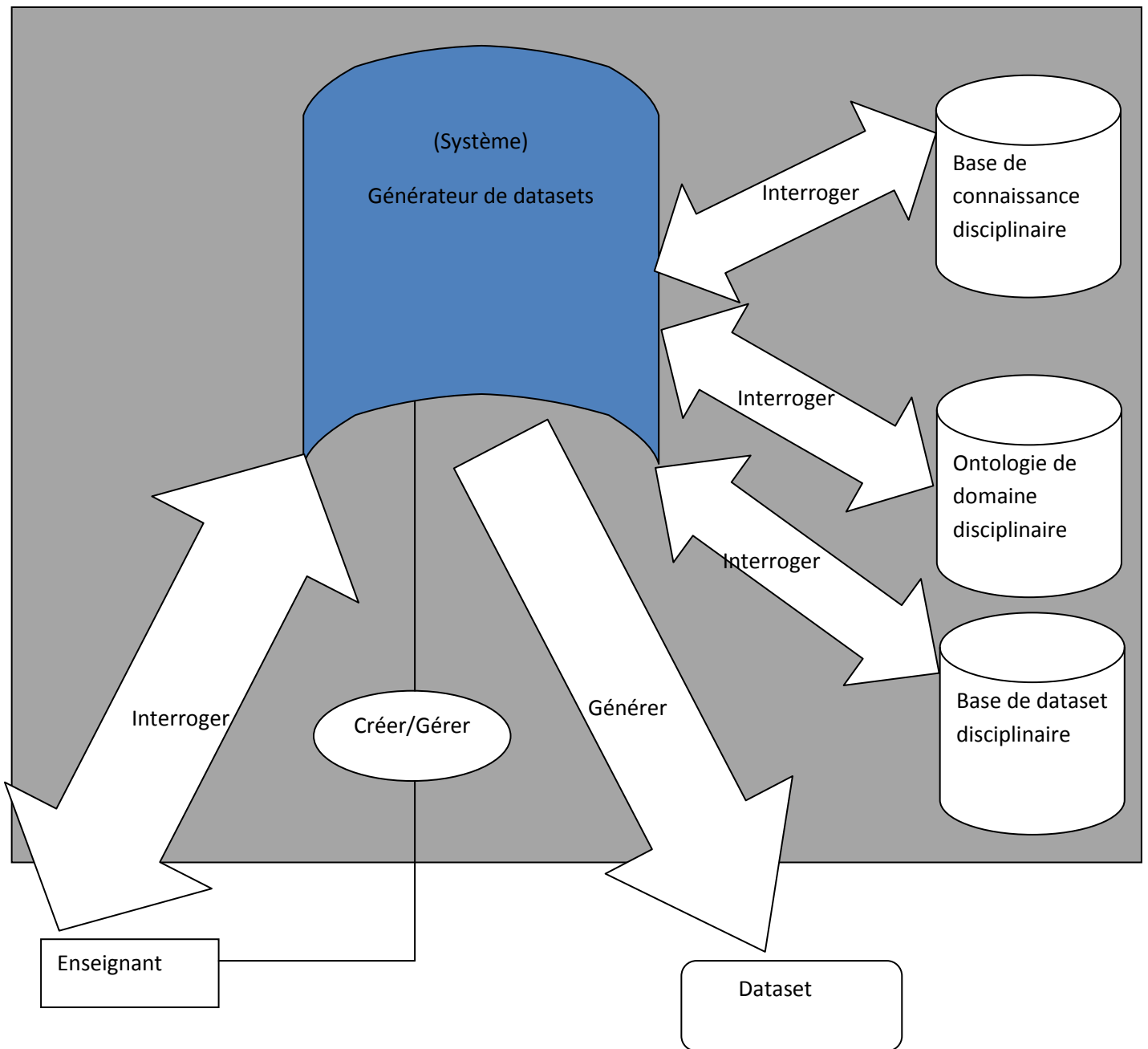


Figure 18 : Fonctionnement de générateur de dataset

3. Le Zoning de la plateforme

3.1 Zoning de la première page maître

Cette étape est cruciale car elle permet de décider l'organisation générale des pages relatives à l'affichage des datasets et à la consultation des ressources. Et cela permet de positionner les zones de navigation, les contenus, le logo et le moteur de recherche. Nous réalisons ici le zoning de la première page maître.

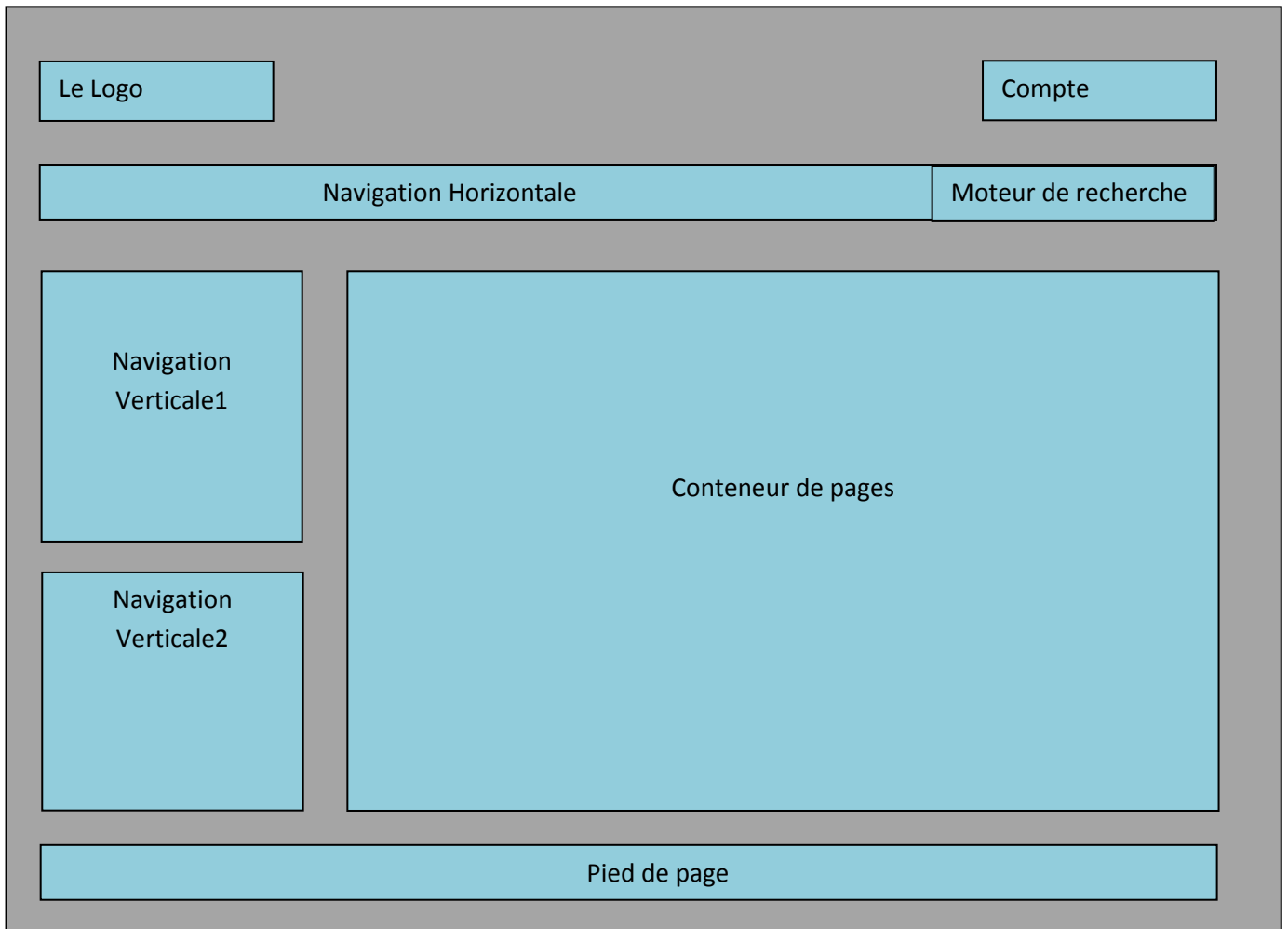


Figure 19 : Le zoning de la première page maître

Le conteneur de pages est un espace pour afficher les différents pages telle que la page d'accueil, la page qui affiche les datasets...

3.2 Zoning de la deuxième page maître

Le zoning de la deuxième page maître permet d'organiser les différentes tâches de création et de gestion des datasets par l'enseignant. Cette page n'est accessible qu'après authentification. Cela permet aussi de positionner les différentes zones de navigation.

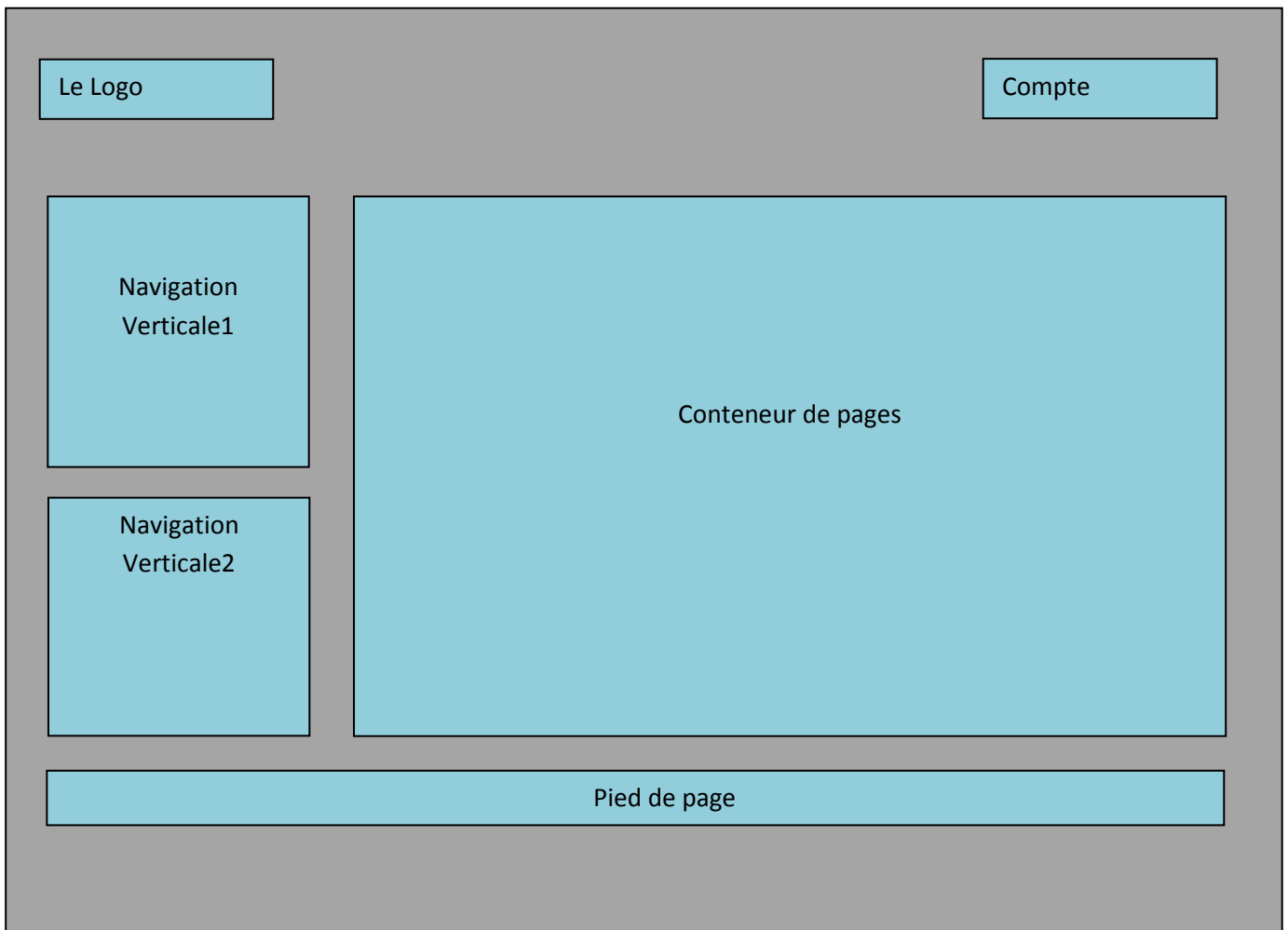


Figure 20 : Le zoning de la deuxième page maître

Le conteneur de pages de cette deuxième page maître permet d'afficher et d'organiser les différentes étapes de création de datasets cela en affichant les pages de créations et de gestions des datasets.

4. Représentation graphique de la démarche de modalisation du site

4.1 Spécification des besoins & cas d'utilisation

4.1.1 Identification des acteurs

Un acteur : représente un rôle que peut jouer l'utilisateur dans le système.

L'acteur est associé à un cas d'utilisation c'est-à-dire qu'il peut interagir avec lui et participer à son scénario, il est représenté par un personnage stylisé.

Les acteurs de notre système sont :

- Le visiteur : toute personne visitant la plateforme.
- L'enseignant : toute personne connectée administrant un dataset.

4.1.2 Identification des cas d'utilisations

Le cas d'utilisation (use case en Anglais) correspond à un objectif du système. Motivé par un besoin d'un ou plusieurs acteurs. L'ensemble des use cases décrit les objectifs du système. Les acteurs définis précédemment effectuent un certain nombre de tâches, ces tâches sont résumées dans le tableau ci-dessous.

Acteur	Tâche
Visiteur	T1 : Se connecter à la plateforme. T2 : Naviguer sur la plateforme. T3 : Consulter les ressources.
Enseignant	T4 : Même opération que T1. T5 : Naviguer sur la plateforme. T6 : S'authentifier pour accéder à l'espace enseignant. T7 : Générer le schéma de dataset. T8 : Créer un dataset. T9 : Publier une ressource. T10 : Supprimer une ressource. T11 : Consulter son profil. T12 : Modifier son profil. T13 : Consulter la liste des datasets. T14 : Consulter les ressources. T15 : Consulter la liste des enseignants. T16 : Se déconnecter.

Tableau 03 : Spécification des tâches.

4.1.3 Identification des scénarios

Un scénario : Une succession particulière d'enchaînements, s'exécutant du début à la fin du cas d'utilisation.

Chacune de tâche effectuée par un acteur est décrite par un ensemble de scénarios, ces scénarios sont illustrés dans les tableaux ci-dessous.

➤ Pour un visiteur

Tâche	Scénarios
T1 : Se connecter à la plateforme	S1 : Saisir L'URL de la plateforme sur un navigateur.
T2 : Naviguer sur la plateforme	S2 : Consulter la liste des datasets. S3 : Consulter la liste des enseignants. S4 : Consulter la liste des ressources relatives aux datasets.
T3 : Consulter les ressources	S5 : Cliquer sur URI d'une ressource et consulter le contenu. S6 : télécharger le contenu d'une ressource.

Tableau 04 : Les scénarios de visiteur

➤ Pour un enseignant

Tâche	Scénarios
T4 : Même opération que T1	S7 : Saisir L'URL de la plateforme sur un navigateur.
T5 : Naviguer sur la plateforme	S8 : Consulter la liste des datasets. S9 : Consulter la liste des enseignants. S10 : Consulter la liste des ressources relatives aux datasets. S11 : Consulter la liste des datasets les plus visités.
T6 : S'authentifier pour accéder à l'espace enseignant.	S12 : Cliquer sur Login. S13 : Saisir le nom utilisateur et le mot de passe. S14 : Cliquer sur « se connecter ».
T7 : Générer le schéma du dataset.	S15 : Cliquer sur « ajouter un dataset ». S16 : Saisir l'intitulé du dataset. S17 : Cliquer sur « Ajouter ». S18 : Cliquer sur « ajouter Notion ». S19 : Saisir les notions.

Chapitre4 : Implémentation d'un prototype d'un site LOD pour les ressources disciplinaire

	<p>S20 : Cliquer sur « Ajouter ».</p> <p>S21 : Cliquer sur « valider les notions ».</p> <p>S22 : Cliquer sur « Ajouter sous-notion ».</p> <p>S23 : Sélectionner une notion.</p> <p>S24 : Saisir les sous-notions.</p> <p>S25 : Cliquer sur « ajouter ».</p> <p>S26 : Cliquer sur « valider les sous notions ».</p> <p>S27 : Cliquer sur « Ajouter les Items de connaissances ».</p> <p>S28 : Sélectionner une notion et une sous-notion.</p> <p>S29 : Saisir les items de connaissances.</p> <p>S30 : Cliquer sur « Ajouter ».</p> <p>S31 : Cliquer sur « valider les items de connaissances ».</p> <p>S32 : Cliquer sur « valider les schémas ».</p>
T8 : Créer un dataset	<p>S34 : Cliquer sur « mon dataset ».</p> <p>S35 : Sélectionner une composante du dataset.</p> <p>S36 : Saisir L'URL de la composante.</p> <p>S37 : Cliquer sur « Valider ».</p>
T9 : Publier une ressource	<p>S38 : Cliquer sur « publier ».</p> <p>S39 : Sélectionner un dataset (si l'enseignant possède plusieurs).</p> <p>S40 : Sélectionner une composante du dataset.</p> <p>S41 : Saisir L'URL de la composante.</p> <p>S42 : Cliquer sur « Valider la ressource ».</p>
T10 : Supprimer une ressource.	<p>S43 : Cliquer sur « Supprimer ».</p> <p>S44 : Sélectionner un dataset (si l'enseignant possède plusieurs).</p> <p>S45 : Sélectionner une composante du dataset.</p> <p>S46 : Cliquer sur « supprimer la ressource ».</p>
T11 : Consulter son profil.	S47 : Cliquer sur « Mon profile ».
T12 : Modifier son profil.	S48 : Modifier les informations du profile.
T13 : Consulter la liste des datasets.	S49 : Cliquer sur « Liste des Datasets »
T14 : Consulter les ressources.	<p>S50 : Cliquer sur un dataset.</p> <p>S51 : Sélectionner une ressource.</p>
T15 : Consulter la liste des enseignants.	S52 : Cliquer sur « Liste des enseignants ».
T16 : Se déconnecter.	S53 : Cliquer sur « Se déconnecter ».

Tableau 05 : Les scénarios de l'enseignant

4.1.4 Diagramme de cas d'utilisation

Le diagramme de cas d'utilisation de chacun des acteurs de notre application est représenté par les figures ci-dessous :

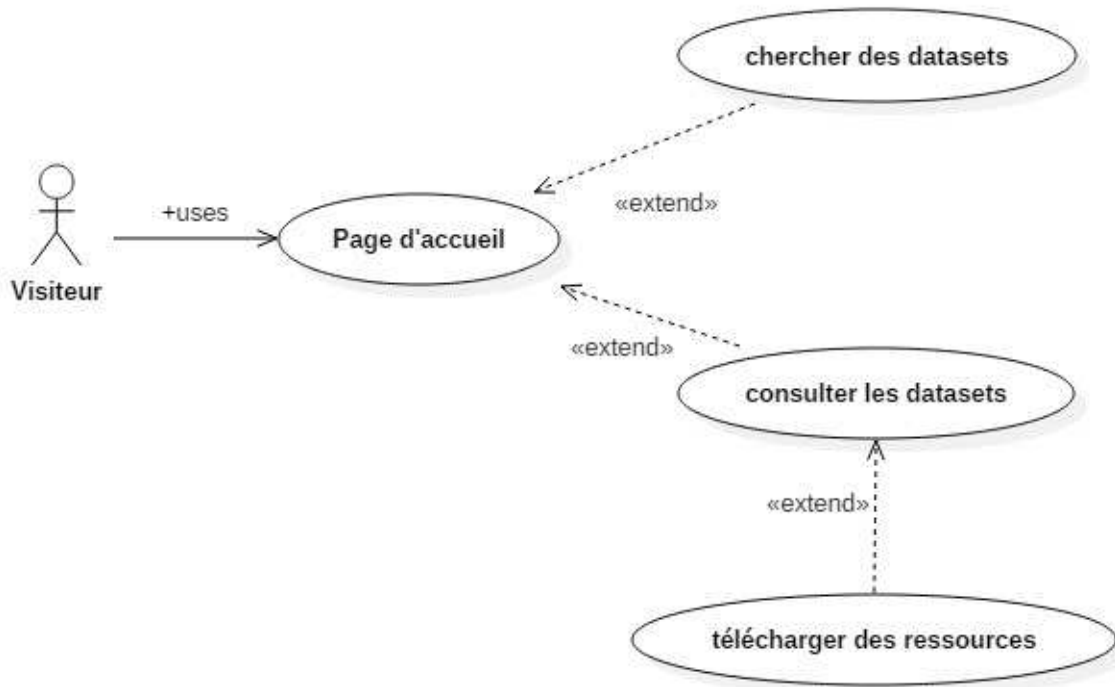


Figure 21 : Diagramme de cas d'utilisation « Visiteur »

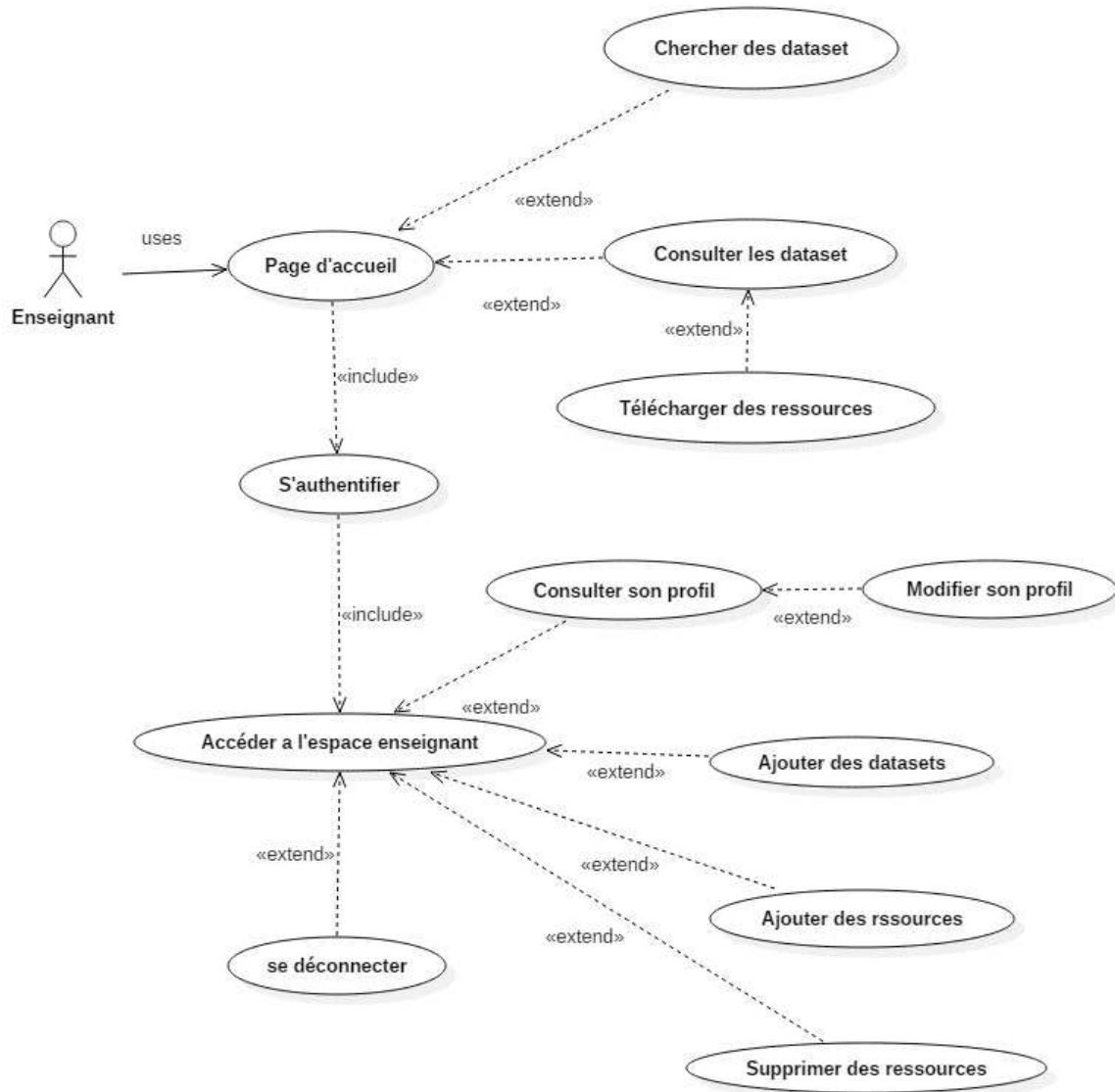


Figure 22 : Diagramme de cas d'utilisation « enseignant »

Introduction

Après avoir présenté dans le chapitre précédent les différentes étapes d'analyse et de la conception, nous allons consacrer ce chapitre à décrire l'environnement et les outils qui ont servi à la mise en œuvre de notre plateforme et enfin, nous terminerons par les différentes fonctionnalités qu'offre notre application web à travers diverses interfaces.

1. Environnement de développement

1.1 Microsoft Visual Studio : est une suite de logiciels de développement pour Windows conçue par Microsoft. La dernière version s'appelle **Visual Studio 2015**.

Visual Studio est un ensemble complet d'outils de développement permettant de générer des applications Web ASP.NET, des Services Web XML, des applications bureautiques et des applications mobiles. Visual Basic, Visual C++, Visual C# et Visual J# utilisent tous le même environnement de développement intégré (IDE, Integrated Development Environment), qui leur permet de partager des outils et facilite la création de solutions faisant appel à plusieurs langages. Par ailleurs, ces langages permettent de mieux tirer parti des fonctionnalités du Framework .NET, qui fournit un accès à des technologies clés simplifiant le développement d'applications Web ASP et de Services Web XML grâce à Visual Web Developer.

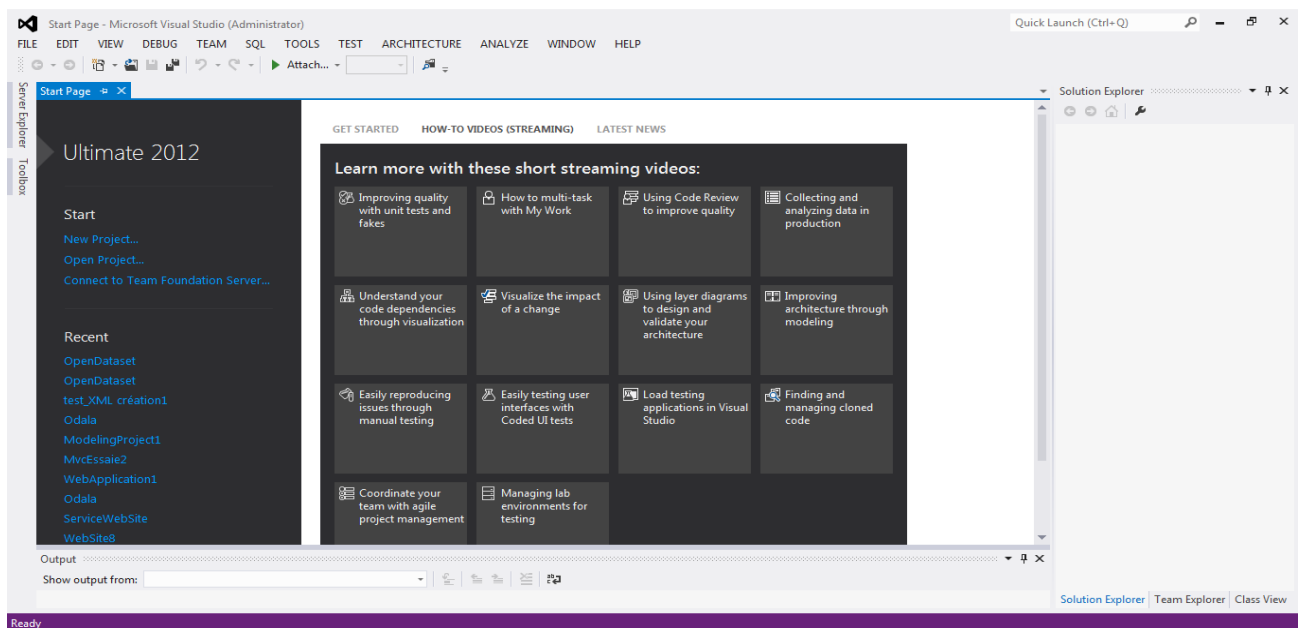


Figure 23 : Visual studio 2012

Nous utiliserons pour notre application le Framework 4.5 et le doctype : XHTML+RDFa 1.0 pour pouvoir générer les triplets RDFa et lier les ressources.

1.2 Internet Information Services : communément appelé **IIS**, est un serveur Web (FTP, SMTP, HTTP etc.) des différents systèmes d'exploitation Windows NT. La dernière version est l'IIS 10.0 utilisé avec Windows Server 2016 et Windows 10

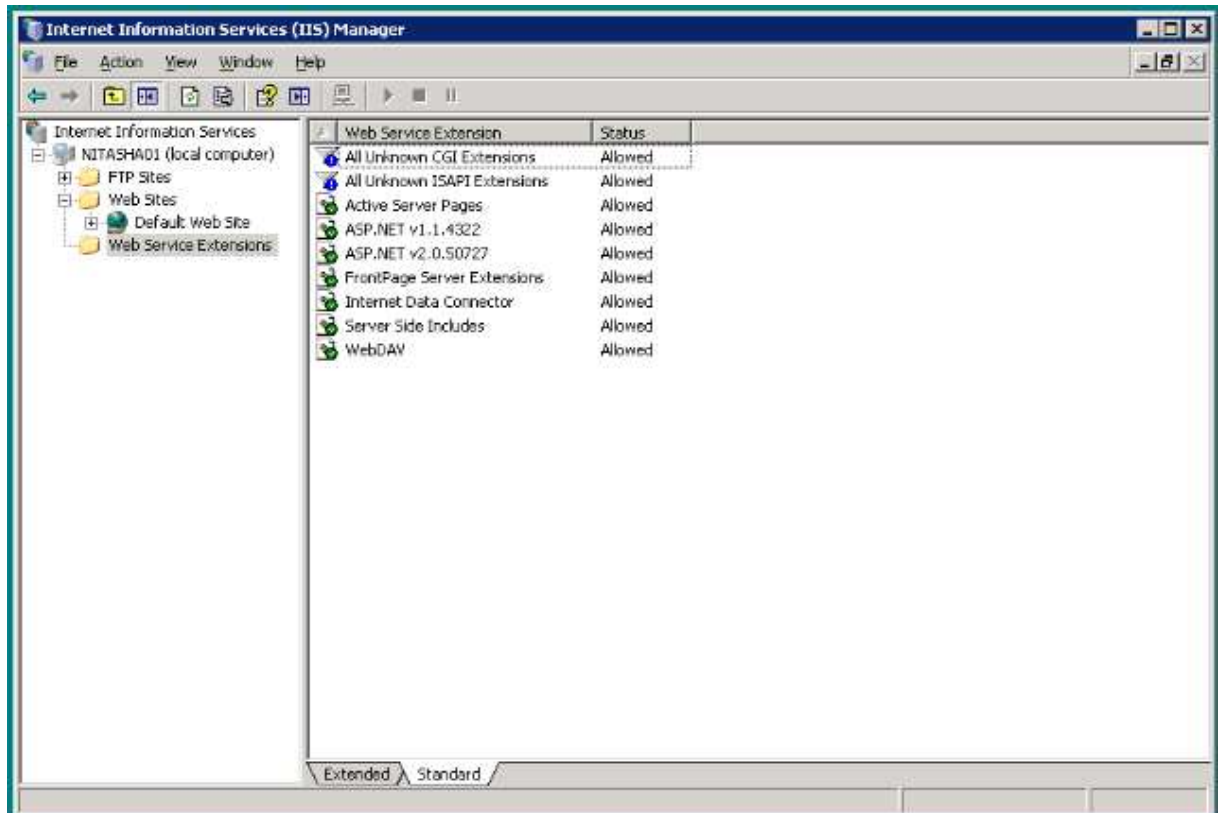


Figure 24 : Serveur web IIS

2. Les langages de programmations

2.1 Le langage XHTML (Extensible HyperText Markup Language)

Le langage XHTML est un langage permettant de décrire la mise en page et la forme d'un contenu rédigé en texte simple. Conçu à l'origine comme un successeur de HTML, XHTML se fonde sur une syntaxe définie par XML.

2.2 RDFa (Resource Description Framework dans des Attributs)

Le RDFa est une recommandation du W3C définissant une syntaxe permettant d'ajouter des données structurées dans une page HTML, XHTML ou n'importe quel document XML. Ainsi formellement décrites, les données peuvent alors faire l'objet de traitements automatisés complexes, via des outils adaptés. Il permet la génération des triplets RDFa pour la structuration des données (définir des liens entre les données et les relations sémantique).

2.3 Le langage de script C#

Le C# est un langage de programmation orienté objet, commercialisé par Microsoft depuis 2002 et destiné à développer sur la plateforme Microsoft .NET. C# est un langage de script pour le web, il se compile et s'exécute au niveau du serveur web IIS et renvoi les résultats aux navigateurs.

2.4 JavaScript

JavaScript est un langage de script Incorporé dans un document HTML. Historiquement il s'agit même du premier langage de script pour le Web. C'est un langage de programmation qui permet d'apporter des améliorations au langage HTML en permettant d'exécuter des commandes du cotés client, c'est-à-dire au niveau du navigateur et non du serveur web.

2.5 Requêtes SQL

C'est un langage de manipulation de base de données mis au point dans les années 70 par IBM. Il permet notamment :

- La manipulation des tables : création, suppression, modification de la structure des tables.
- La manipulation des bases de données : ajout, sélection, modification, et suppression d'enregistrement.
- La gestion des droits d'accès aux tables : contrôles des données et validation des modifications.

2.5 CSS

Les feuilles de styles simplifient le balisage HTML et soulage grandement ce dernier des responsabilités de la représentation. Elles donnent le contrôle de la représentation des documents, informations sur les polices de caractères, alignement, couleurs etc. l'objectif est de tendre vers la séparation du contenu et de la présentation. Les informations de styles peuvent être spécifiées pour un élément ponctuel ou pour des groupes d'éléments. Elles peuvent se trouver à l'intérieur du document HTML ou dans une feuille de style externe.

3 Présentation des interfaces

Dans ce qui suit nous allons présenter quelques interfaces de notre plateforme.

3.1 Page d'accueil

La page d'accueil est la première page que le visiteur visualisera quand il accédera à la plateforme. D'où une nécessité de clarté et une bonne présentation de la page. le bute premier est facilité au visiteur l'accès aux différents espaces de la plateforme.

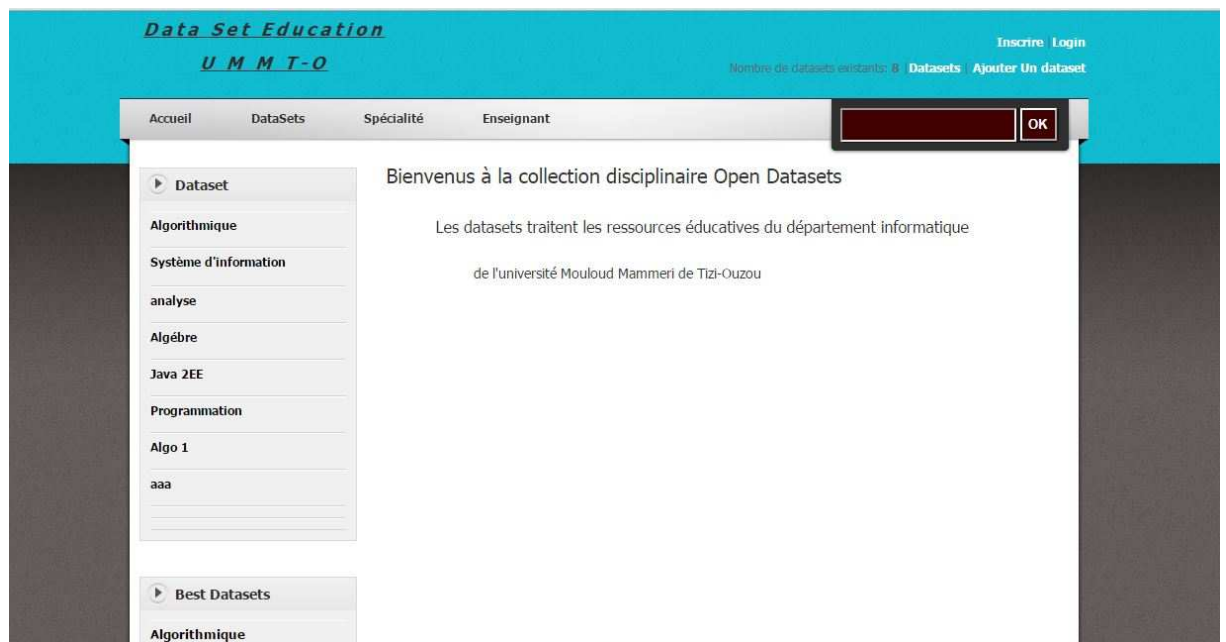


Figure 25 : Page d'accueil

3.5 Page de consultation des datasets

Sur la page de consultation des datasets on trouve la liste des datasets existant dans la plateforme et les informations relatives à chacun. Le visiteur peut même voir les datasets les plus visités.



Figure 26 : Page consultation des datasets

3.6 Page de consultation des ressources

Sur la page de consultation des ressources on trouve les ressources représentant les datasets. L'accès aux ressources est systémique suivant l'ontologie de domaine disciplinaire.

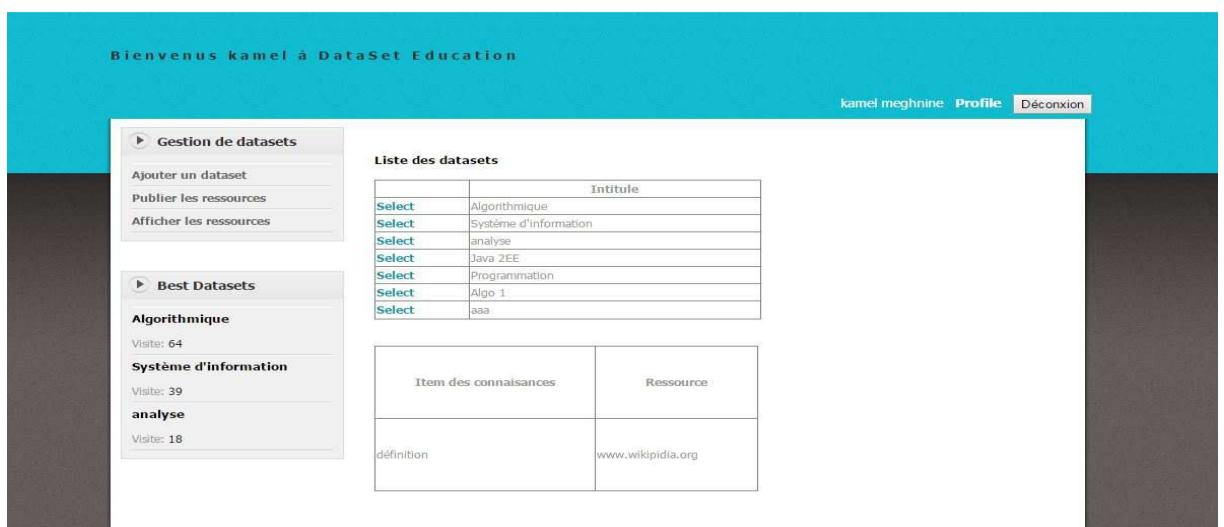


Figure 27 : Page de consultation des ressources

3.7 Page d'authentification

Cette page permet aux enseignants de connecter à leurs comptes. Aussi permet aux visiteurs d'accéder à la page d'inscription.

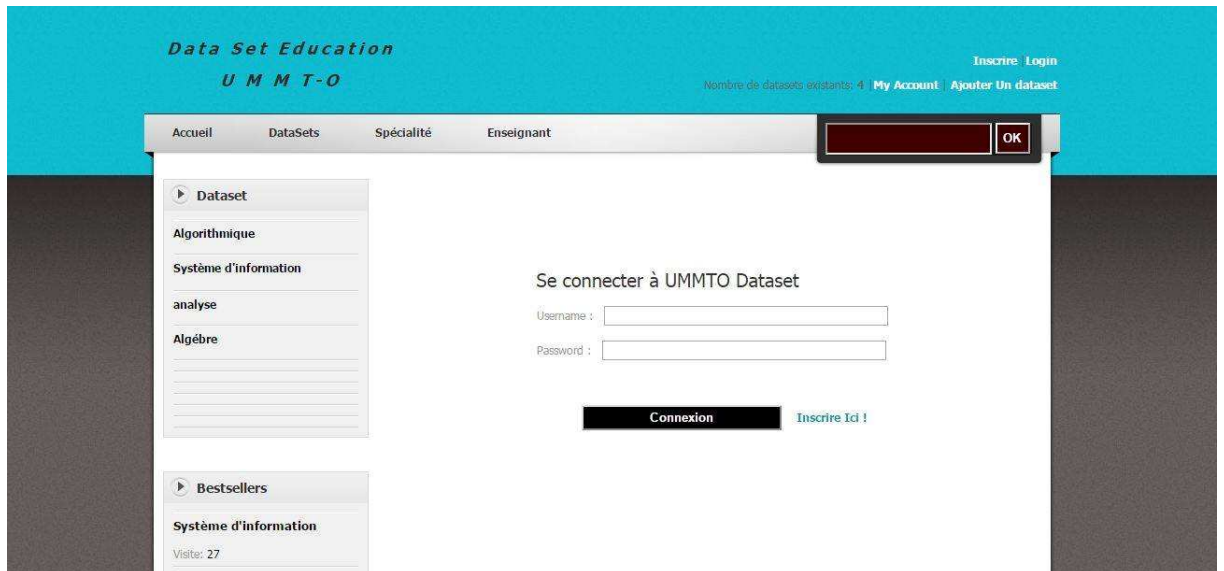


Figure 28 : Page d'authentification

3.8 Page d'inscription

Cette page permet au visiteur de créer un compte.

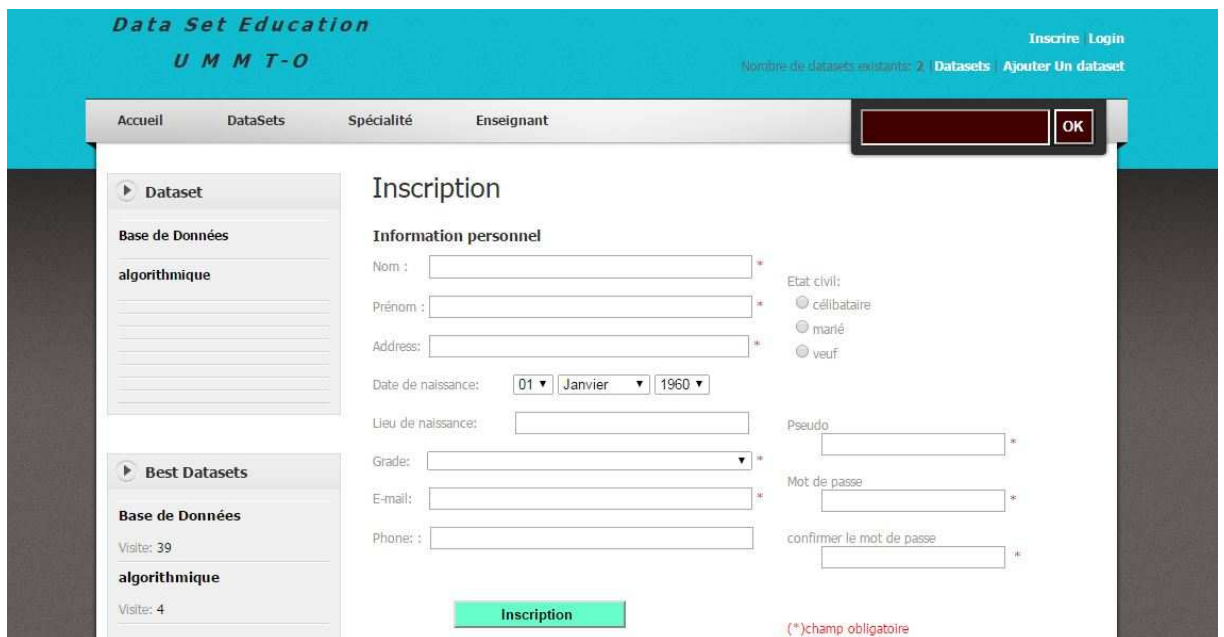


Figure 29 : Page d'inscription

3.9 Page d'accueil dans l'espace enseignant

Cette page est la première page que l'enseignant visualisera après création du compte.



Figure 30 : Page d'accueil dans l'espace enseignant

3.10 Page de création d'un dataset

Sur cette page un enseignant peut créer son dataset en définissant le schéma interne du dataset.

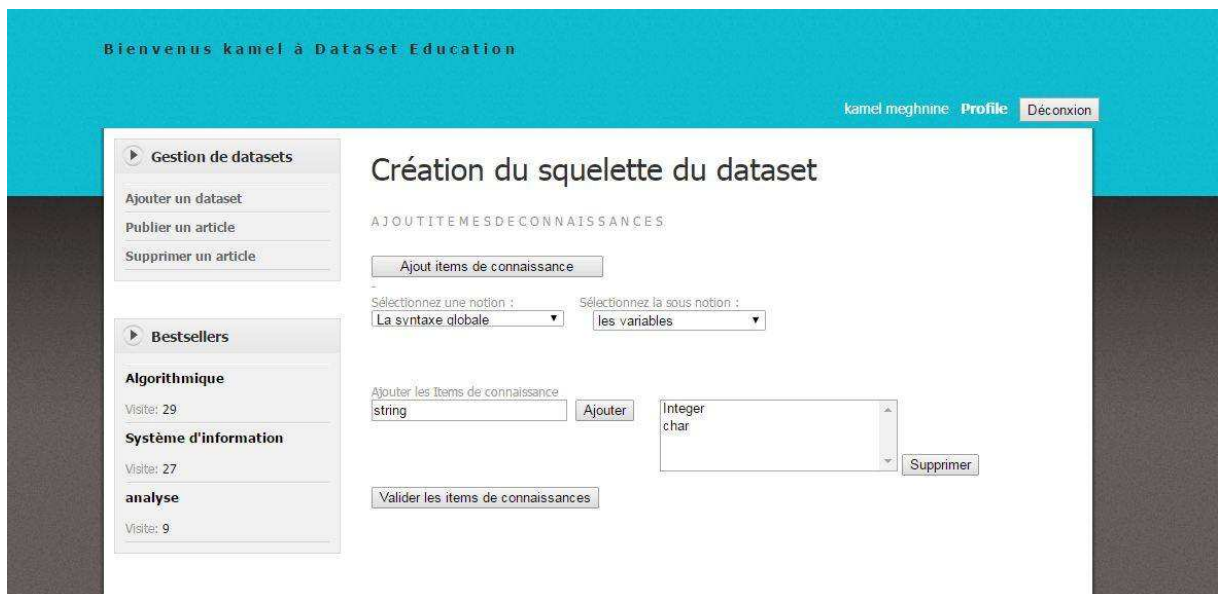


Figure 31 : Page de création d'un dataset

3.11 Page de publication d'une ressource

Sur cette page un enseignant peut publier/supprimer une ressource sur son dataset.

Gestion de datasets

- Ajouter un dataset
- Publier les ressources
- Afficher les ressources

Best Datasets

Algorithmique
Visite: 64

Système d'information
Visite: 39

analyse
Visite: 18

3 Liste des datasets

	Dataset
Détail	Algorithmique
Détail	Système d'information
Détail	analyse
Détail	Java 2EE
Détail	Programmation
Détail	Algo 1
Détail	aaa

Liste des notions

	Notion
Détail	les boucles
Détail	condition

Liste des sous notion

	Sous notion
Détail	for
Détail	while

Liste des items de connaissance

Détail	Item de connaissance
+	définition

Entrer URL :

Figure 32 : Page de Publication d'une ressource

Conclusion Générale :

Le web a connu plusieurs extensions et plusieurs développements, l'un des fruits de ces extensions l'appariation du web sémantique, un web qui se base sur la sémantique. Cette extension est standardisée par le w3c. Ces standards encouragent l'utilisation de format de données et de protocoles normés.

Le web sémantique propose des solutions et des techniques pour atteindre l'objectif de créer un web sophistiqué qui se base sur les données au lieu des documents. Parmi ces solutions : le RDF (ressources description framwork), OWL (Ontologie web language)...

Cette nouvelle vision du web crée des nouveaux concepts et une nouvelle gestion des ressources. Parmi ses nouveaux concept on trouve le web des données (Linked data) et les données ouvertes (open data), ce qui favorise l'utilisation des ressources existantes sur le web.

Notre solution présentée dans ce mémoire se base sur ses deux principes pour la représentions des ressources disciplinaire du département informatique de l'université mouloud Mammeri de Tizi-Ouzou, et créer une plateforme sémantique en respectant les principes du web sémantique, une plate forme interprété par les moteur de recherche du web sémantique.

En vue de la nouveauté et la complexité du thème de notre recherche nous restons ouverts à toutes critiques et nous somme prêt à recevoir toutes les suggestions et remarque tendant à améliorer d'avantage cette étude. Etant donné que tout travail informatique a été l'œuvre d'une équipe.

Références bibliographiques

- [Tom, 2011] TOM HEALTH, CHRISTIAN HEALTH, Linked Data: Evolving the Web into a Global Data Space, 2011
- [Halpin,2010] HARRY HALPIN, PATRICK HAYES, JAMES MCCUSKER, DEBORAH MCGUINNESS, AND HENRY THOMPSON. When owl:sameas isn't the same: An analysis of identity in linked data. In Proceedings of the 9th International Semantic Web Conference, 2010. DOI: 10.1007/978-3-642-17746-0_20 23
- [Lenne, 2005] LENNE D., ABEL M.-H., MOULIN C. ET BENAYACHE A. (2005) Mémoire de formation et apprentissage. Environnements Informatiques pour l'Apprentissage Humain (EIAH 2005), Montpellier, 105-116.
- [Psyche, 2005] PSYCHE V., BOURDEAU J., NKAMBOU R. ET MIZOGUCHI R. (2005) Making Learning Design Standards Work with an Ontology of Educational Theories. Artificial. Intelligence in Education (AIED 2005), 725-731.
- [Abel , 2003] ABEL M.-H., LENNE D., MOULIN C. ET BENAYACHE A. (2003) Gestion des ressources pédagogiques d'une e-formation. Document Numérique 7(1-2): 111-128.
- [Studer, 1998] STUDER R., BENJAMINS VR. ET FENSEL D. (1998) Knowledge Engineering : Principles and Methods. Data and Knowledge Engineering (DKE), 25(1-2):161-197.

Références bibliographiques

- [Jacobs,2004] IAN JACOBS, NORMAN WALSH. Architecture of the World Wide Web, Volume One, 2004. <http://www.w3.org/TR/webarch/> . 7, 9
- [Chris,2016] CHRIS BIZER, RICHARD CYGANIAK, [http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedData Tutorial/](http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/), 7/4/2016.
- [Tim Berners-Lee,1998] TIM BERNERS-LEE, R. FIELDING, AND L. MASINTER. RFC 2396 - Uniform Resource Identifiers (URI): Generic Syntax . <http://www.isi.edu/in-notes/rfc2396.txt>, August 1998. 7
- [Raggett,1999] DAVE RAGGETT, ARNAUD LE HORS, AND IAN JACOBS. Html 4.01 specification - w3c recommendation. <http://www.w3.org/TR/html401/>, 1999. 7
- [Klyne,2004] GRAHAM KLYNE AND JEREMY J. CARROLL. Resource Description Framework (RDF): Concepts and Abstract Syntax - W3C Recommendation, 2004. <http://www.w3.org/TR/rdf-concepts/> .8, 15, 17
- [Moats,1997] R. MOATS. Rfc 2141: Urn syntax. <http://tools.ietf.org/html/rfc2141> , 1997. 1
- [Paskin,2006] NORMAN PASKIN. The doi handbook - edition 4.4.1. http://www.doi.org/handbook_2000/DOIHandbook-v4-4.1.pdf , 2006. 10
- [Walsh,2006] NORMAN WALSH. Names and Addresses. <http://norman.walsh.name/2006/07/25/namesAndAddresses> , 2006. 10

- [Thompson,2006]** HENRY THOMPSON, DAVID ORCHARD. Urns, namespaces and registries
<http://www.w3.org/2001/tag/doc/URNsAndRegistries-50>, 2006. 10
- [Sauermann,2008]** LEO SAUERMAN, RICHARD CYGANIAK. Cool uris for the semantic web - w3c interest group note.<http://www.w3.org/TR/cooluris/> , 2008. 11, 13, 14, 15, 43, 46
- [Boutmedjet, 2004]** BOUTEMEDJET S. (2004) Web Sémantique et e-Learning. Cours FT6261.
- [Gasevic, 2005]** GASEVIC D. ET HATALA M. (2005) Searching context relevant learning resource using ontology mappings. International Workshop on Applications of Semantic Web Technologies for E-learning (SW-EL), Winston-Salem State University, <http://www.win.tue.nl/SW-EL/2005/swel05-kcap05/proceedings>.
- [IMS LD, 2003]** IMS LD (2003) <http://www.imsglobal.org/learningdesign/>
- [SCORM, 2004]** SCORM. Le modèle SCORM (2004). <http://www.adlnet.org>
- [LOM, 2002]** LOM (2002). LOM standard, document IEEE 1484.12.1-2002.
- [Koper, 2001]** KOPER R. (2001) EML-OUNL (Open University of the Netherlands' Educational Modeling Language), Modeling Units of Study from a Pedagogical Perspective. <http://eml.ou.nl/introduction/docs/ped-metamodel.pdf>
- [Flory,2004]** LAURENT FLORY (2004), <http://www.enssib.fr/bibliotheque-numerique/documents/1234-les-caracteristiques-d-une-ressource-pedagogique-et-les-besoins-d-indexation-qui-en-resultent.pdf>