

Chapitre 2 : Web de données

1. Introduction

Le Web de Données, également connu sous le nom de Linked Data (Données Liées), représente une révolution majeure dans la manière dont nous concevons, partageons et interagissons avec les informations sur Internet. Contrairement au Web traditionnel, qui est principalement composé de documents hypertextes et de médias, le Web de Données s'appuie sur un modèle de données sémantique qui vise à rendre les données aussi interconnectées et significantes que les pages web l'ont été pour les documents. Dans ce chapitre nous allons découvrir toutes les notions relatives au web de données.

2. Séparation du fond et de la forme

La fin des années 1990 marque le début d'une évolution vers un Web davantage axé sur la structuration. Le W3C a joué un rôle clé dans cette transition en promouvant la séparation du contenu et de la présentation. Au fil du temps, ses directives ont contribué à façonner le Web tel que nous le connaissons aujourd'hui.

En janvier 1997, la recommandation HTML 3.2 a été publiée, normalisant les extensions les plus courantes du HTML. Cela incluait la gestion des tableaux, l'agencement du texte autour des images, l'intégration d'applets (petites applications Java) dans les pages, ainsi que la possibilité d'utiliser des indices et des exposants dans les textes. Cette recommandation a également posé les bases pour les évolutions ultérieures, notamment l'introduction des scripts, ainsi que les feuilles de style, marquant ainsi le début de la séparation du contenu et de la présentation (les feuilles de style CSS datent de 1995, avec leur première recommandation en 1996).

En 1999, la recommandation HTML 4.01 a été adoptée, généralisant l'intégration d'objets multimédias et de scripts dans les pages, élargissant les fonctionnalités des tableaux et des formulaires, etc. L'accent a été mis sur la séparation plus poussée du contenu et de la présentation, en encourageant l'utilisation des feuilles de style CSS pour définir la mise en forme du contenu HTML.

HTML continue son évolution, avec la mise en place d'un groupe de travail dédié à HTML 5 dès mars 2007. Parmi les nouvelles fonctionnalités envisagées, on trouve des interfaces de programmation unifiées qui permettent d'accéder au code de la page HTML de manière standard. Cela autorise, par exemple, un script inclus dans une page à interagir de manière cohérente avec le code de cette page lorsqu'il s'exécute dans le navigateur. D'autres fonctionnalités incluent l'intégration et la gestion de contenu audio et vidéo. HTML 5 normalise également des éléments de page courants, tels que <footer> pour le bas de page, <nav> pour les éléments de navigation, et <figure> pour attribuer des légendes.

Cette pratique avancée de séparer le contenu de la présentation a pris une nouvelle dimension avec l'émergence d'une nouvelle famille de langages. En 1998, la première recommandation XML (Extensible Markup Language) a été publiée. Alors qu'HTML fournissait un langage textuel pour l'échange de documents semi-structurés, XML a introduit un format textuel pour l'échange de données structurées. Ce nouveau standard a permis la définition de langages balisés pour l'échange de données structurées, notamment sur le Web, sous forme d'arborescences. XML se distingue de son prédécesseur SGML par sa simplicité [1].

1.1 XML (Extensible Markup Language)

XML est un langage qui va permettre de structurer des données en choisissant les balises que l'on veut et non plus en suivant uniquement les balises proposées par le langage HTML et ceci dans un format textuel, ce qui permet d'écrire des documents XML en choisissant ses propres balises pour structurer ses données et ses documents.

Principe :

- Une et une seule racine <racine>...</racine>
- Utilisation d'autant de balise que l'on veut <x>...</x> ou <x/>
- Respecter l'emboîtement des balises <a>
- Sensibilité à la casse <x> ≠ <X>
- Respect de certaines syntaxes : des noms de balises qui commencent par un chiffre, par xml, ou de mettre des espaces, etc. comme <4x>, <xlma>, <nom prénom>.
- Utilisation d'attributs dans les balises :

```
<pays> #racine
  <nom>Algérie</nom> #balise ouvrante et balise fermante
  <capitale>Alger</capitale>
  <superficie unite='km²'>2 381 741</superficie> #unité est un attribut
  <population>43 851 044 habitants (estimation 2021)</population>
  <monnaie>Dinar algérien (DZD)</monnaie>
  <gouvernement type='République'></gouvernement> #balise auto fermante
</pays>
```

L'intérêt de cette catégorie de langages ne se limite pas uniquement à l'établissement d'une méthode standardisée pour échanger des données, mais englobe également la normalisation des outils qui les entourent. Cela inclut ce que l'on appelle communément les 'parseurs', des outils conçus pour lire ces données. Ces parseurs permettent, par exemple, de fournir des bibliothèques standardisées dans tous les langages de programmation, facilitant ainsi la lecture et l'écriture de données au format XML. De plus, il devient possible de standardiser la validation de ces données

en fournissant des schémas qui garantissent, par exemple, que tout document XML reçu respecte la structure attendue.

XML Schema (XMLSchema) peut être qualifié de langage de schéma pour la définition de structures de données XML. Il permet de définir la structure, les contraintes et les types de données autorisés dans un document XML. XML Schema est utilisé pour valider et garantir la conformité des documents XML à une structure spécifique [2].

3. Définition du Web des données

Le Web des données, également connu sous le nom de "Linked Data" en anglais, est une initiative du W3C qui vise à transformer la manière dont les données sont publiées sur Internet. Au lieu de les isoler dans des silos de données distincts, l'objectif est de les connecter entre elles pour former un vaste graphe de données interconnectées.

Cette approche repose sur l'utilisation des normes du Web telles que HTTP. Au lieu de simplement utiliser ces normes pour faciliter la navigation humaine sur le Web, le Web des données les élargit pour permettre le partage automatique d'informations entre les machines. Cette évolution permet d'interroger les données de manière automatisée, quel que soit leur emplacement, sans nécessité de les dupliquer.

Tim Berners-Lee, le directeur du W3C et inventeur du World Wide Web, a introduit le terme "Linked Data" ou "données liées" ainsi que son synonyme "web of data" dans un ouvrage consacré à l'avenir du Web sémantique [3].

Les données liées consistent simplement à utiliser le Web pour **créer des liens typés** entre des données provenant de sources différentes. Ces sources peuvent être aussi diverses que des bases de données entretenues par deux organisations dans des emplacements géographiques différents, ou simplement des systèmes hétérogènes au sein d'une seule organisation qui, historiquement, n'ont pas facilement interopéré au niveau des données. Techniquement, **les données liées font référence aux données publiées sur le Web de manière à ce qu'elles soient lisibles par les machines**, que leur signification soit explicitement définie, qu'elles soient liées à d'autres ensembles de données externes, et qu'elles puissent à leur tour être liées à partir d'ensembles de données externes [4].

Alors que les principales unités du Web hypertexte sont des documents HTML (HyperText Markup Language) reliés par des hyperliens non typés, les données liées reposent sur des documents contenant des données au format **RDF (Resource Description Framework)** [5]. Cependant, au lieu de simplement relier ces documents, les données liées utilisent RDF pour établir des déclarations typées qui relient des éléments arbitraires dans le monde. Le résultat, que nous appellerons le Web de données, peut-être plus précisément décrit comme un réseau d'éléments dans le monde, décrits par des données sur le Web [4].

4. Principes de données liées

Berners-Lee [4] a énoncé un ensemble de principes pour publier des données sur le Web de manière à ce que toutes les données publiées fassent partie d'un espace de données global unique :

1. Utilisez des **URI** comme noms pour les éléments.
2. Utilisez des **URI HTTP** afin que les personnes puissent rechercher ces noms.
3. Lorsque quelqu'un recherche une **URI**, fournir des informations utiles en utilisant les normes (**RDF, SPARQL**).
4. Inclure des liens vers d'autres **URI**, afin que plus d'éléments puissent être découverts.

Ces principes sont désormais connus sous le nom de "principes des données liées" et fournissent une recette de base pour publier et relier des données en utilisant l'infrastructure du Web tout en respectant son architecture et ses normes.

5. Vers un graphe de données

La nécessité d'enrichir le Web de documents par des données structurées et sémantiques était essentielle pour rendre le Web plus utile pour les machines, améliorer la recherche d'informations, faciliter l'interopérabilité des données et ouvrir la voie à de nouvelles applications basées sur l'intelligence artificielle et l'Internet des objets.

Dans le contexte du Web des données, une **ressource** fait référence à une entité identifiée et définie de manière unique qui peut être représentée, consultée ou manipulée sur le Web. Ce qui permet de faciliter leur intégration, leur recherche et leur utilisation par des machines et des applications pour diverses finalités, notamment la recherche d'informations, l'analyse de données et la création de connaissances interconnectées.

2.1 RDF (Resource Framework Description)

RDF (Resource Framework Description) a été spécifié pour la première fois en 1999 par le World Wide Web Consortium (W3C) en tant que modèle de données pour représenter des données sémantiques. RDF a été conçu pour être un modèle simple mais puissant, basé sur des triplets (sujet-prédicat-objet), qui permettent de représenter des assertions sur des ressources et leurs relations de manière formelle et extensible. Cette structure triplet a été choisie pour sa simplicité et sa capacité à représenter efficacement des connaissances. RDF s'insère dans l'architecture classique du Web en réutilisant notamment le mécanisme des **URI (Uniform Resource Identifier)** pour identifier les ressources et **représenter explicitement toute relation entre deux ressources**.

3.1 Triplet

Le modèle RDF permet d'établir des graphes étiquetés orientés en exploitant le triplet [6] dont la structure est la suivante :

<ressource -propriété -valeur>

ou

< sujet -prédicat -objet >

- Ressource

Les ressources sont des concepts de base sur le web sémantique, tout à quoi on peut se référer est considéré comme une ressource, que ce soit un concept physique ou abstrait, notamment une page web, une image, une vidéo, mais aussi un lieu, une personne, une organisation, un évènement, un produit, un service, etc.

- Propriété

C'est la description d'une ressource qui peut être défini comme un aspect, une caractéristique, un attribut ou une relation spécifique pour décrire une ressource. Chaque ressource possède une signification spécifique, définit ses valeurs permises, les types de ressources qu'elle peut décrire et les relations qu'elle entretient avec les autres ressources.

- Valeur

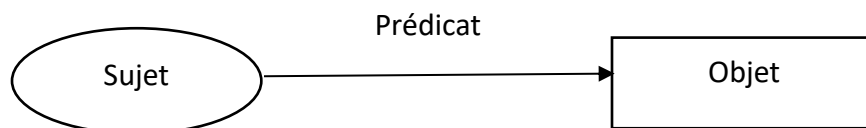
Il représente la valeur ou la cible de la propriété définie par le prédicat pour la ressource spécifiée par le sujet. L'objet peut être une valeur simple (comme une chaîne de caractères, un nombre ou une date) appelée **littérale** ou une autre ressource identifiée par une URI.

Ensemble, les trois éléments forment **une assertion sémantique** qui déclare une relation entre le sujet, le prédicat et l'objet. Les **triplets** RDF peuvent être utilisés pour représenter une grande variété d'informations, de connaissances et de données structurées sur le Web de manière formelle et sémantique

Exemple :

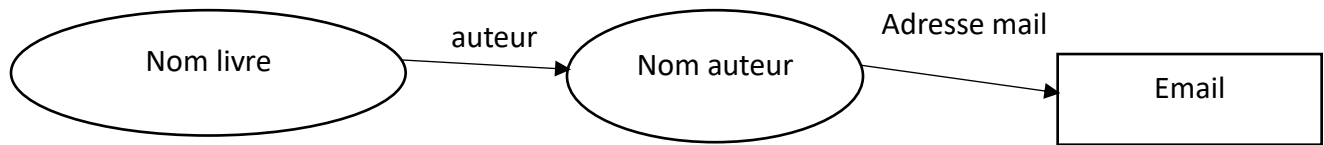
- Sujet/ressource : URI désignant un livre spécifique
- Prédicat/propriété : La propriété "auteur" pour une ressource livre.
- Objet/valeur : URI d'un auteur spécifique ou le titre d'un livre.

Dans le schéma suivant :



- Une ressource est représentée par un cercle (ovale),

- Un prédicat (propriété) est représenté par un arc,
- Un objet est représenté par un rectangle



6. URI (Uniform Resource Identifier)

Cette partie décrit le principe 1 des données liées.

Les URL sont utilisées pour identifier de manière unique les documents dans le Web. Il faut savoir, qu'en plus des URL, le Web définit également la notion d'identifiant uniforme de ressource (URI), qui généralise les URL, qui font référence à la fois à l'identité et à l'emplacement d'une ressource, tandis qu'un URI doit seulement établir **l'identité d'une ressource**. Cela signifie qu'une page Web peut avoir une URL car elle a un emplacement sur le Web (qui sert également d'identifiant : son URL est également son URI), tandis que généralement les personnes, comme l'auteur Pablo Picasso, ne peut pas avoir d'URL car il ne peut pas avoir lui-même d'emplacement sur le Web; mais il peut avoir un URI qui permettra son identification sur le Web. En utilisant des identifiants Web (URIs) tels que https://example.org/Pablo_Picasso pour nommer des éléments du monde réel.



Figure 1 Uniform Resource Identifier [7]

- Une URL, ou Uniform Resource Locator, est une adresse qui est utilisée pour identifier de manière unique une ressource sur Internet. Elle est composée de plusieurs éléments qui décrivent l'emplacement et l'accès à cette ressource. Une URL est le format couramment utilisé pour accéder à des ressources sur le Web.
- URI est une chaîne de caractères qui identifie de manière unique une ressource, qu'elle soit sur Internet ou non. Un URI est composé d'une syntaxe bien définie qui comprend des caractères alphanumériques et certains caractères spéciaux tels que "/", ".", "?", etc.
- IRI (Internationalized Resource Identifier) : Un IRI est une extension de l'URI qui permet d'utiliser une gamme plus large de caractères, y compris des caractères non latins ou non ASCII, pour représenter des ressources. Les IRI sont particulièrement utiles pour prendre en charge des langues et des alphabets non latins dans les noms de domaine, ce qui les rend plus adaptés à la diversité linguistique du Web mondial. Les IRI sont définis par la norme RFC 3987.

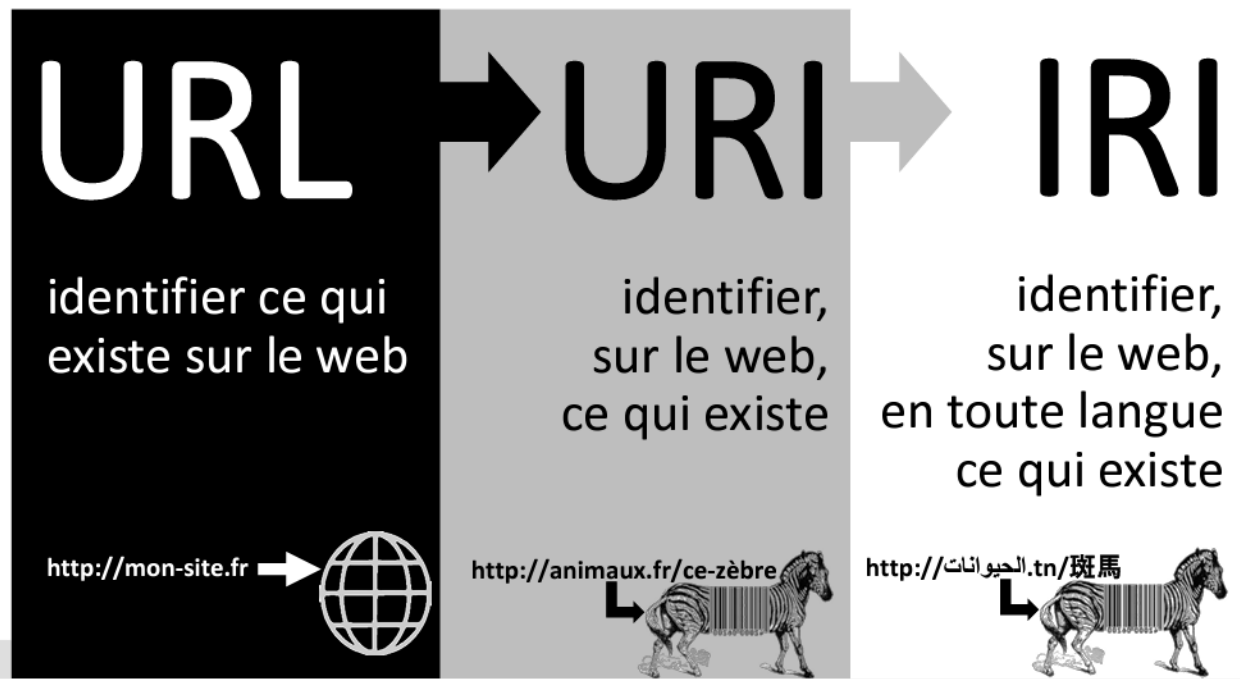


Figure 2 URL, URI et IRI [2]

- Un URN (Uniform Resource Name) est un type d'URI, utilisé pour identifier une ressource par son nom dans un espace de noms particulier de manière unique une ressource sans spécifier son emplacement ou sa méthode d'accès. Les URN se concentrent uniquement sur l'identification de la ressource elle-même de manière durable et persistante. L'URN reste inchangé pendant le cycle de vie de la ressource. Les URN sont définis par la norme

RFC 2141. Un exemple des URN est l'ISSN des livres, qui identifie un livre spécifique, quelle que soit son emplacement ou son format¹ [8].

En résumé, un URI est le terme global qui englobe à la fois les URLs et les URNs. Les URLs sont spécifiquement utilisées pour localiser et accéder à des ressources sur le web, tandis que les URNs sont utilisées pour nommer de manière stable des ressources, souvent indépendamment de leur localisation physique.

1.1 Syntaxe des URIs

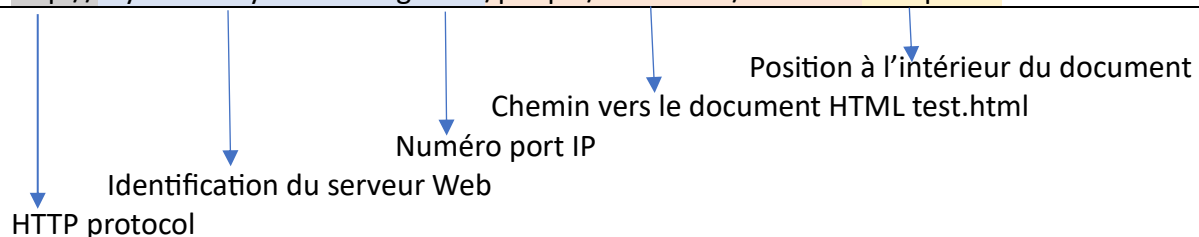
La syntaxe générique des URIs [8] :

```
protocol ":" [ "//" authority "/" ] [ path ] [ "?" query ] [ "#" fragment]
```

- protocol :http, ftp, etc.
- authority: [userinfo@]host[:port]
 - userinfo: authentication section e.g: username:password
 - host: nom de domaine , adresse IP
 - port: port number, ex: 80 for HTTP standard port
- path: une séquence de segments séparés par des barres obliques, par exemple : un chemin dans le système de fichiers hiérarchique du serveur HTTP.
- query: Une chaîne de requête de données non hiérarchiques. (par exemple : une séquence de paires attribut-valeur séparées par un délimiteur (&) pour les requêtes HTTP).
- fragment: un identifiant de fragment fournissant une direction vers une ressource secondaire (par exemple : un identifiant d'ancre dans un document HTML).

Example

http://myserver.mydomain.org:8080/people/mohamed/test.html#chapitre1



¹ URIs, URLs, and URNs: Clarifications and Recommendations 1.0 Report from the joint W3C/IETF URI Planning Interest Group-W3C Note 21 September 2001 <http://www.w3.org/TR/uri-clarification/>

2.1 Exemple d'une ressource

Exemple de la ressource Harry Potter

recherchée sur **le web traditionnel**

Documents

Identifiés par des URLs



Image

https://www.imdb.com/title/tt0241527/mediaviewer/rm3773460480/?ref_=tt_ov_i



Web page (HTML file)

https://fr.wikipedia.org/wiki/Harry_Potter



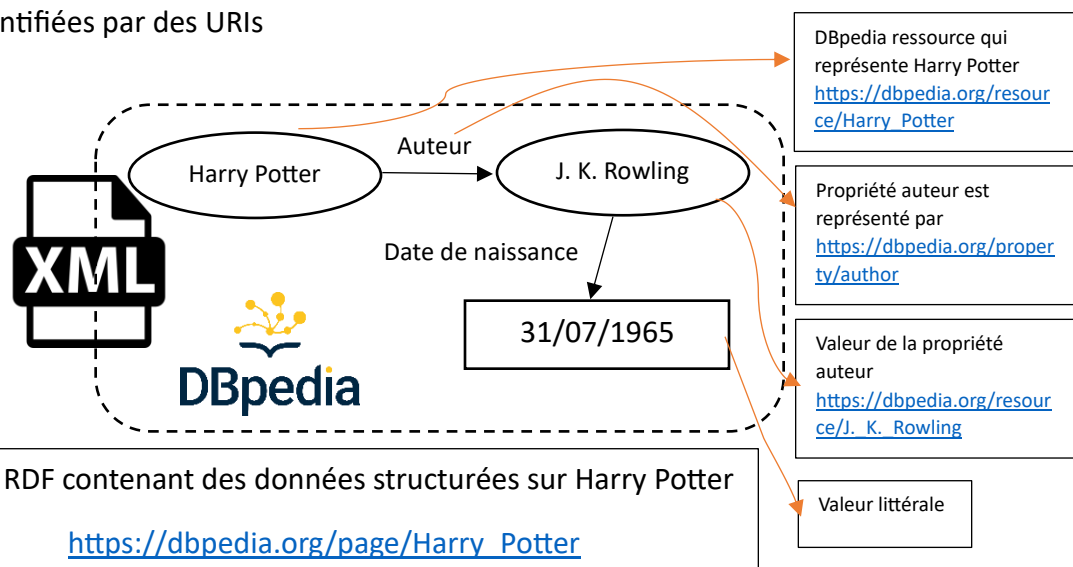
Vidéo

https://www.imdb.com/title/tt0241527/?ref_=tt_mv_close

Le web des données

Ressources

Identifiées par des URIs



Exemple de la ressource Algérie dans deux espaces de noms (name space), DBpedia et Geonames. Différents URIs, dans différents espaces de noms², peuvent représenter la même ressource sur le Web.



Algérie

<https://www.geonames.org/countries/DZ/algeria.html>

L'URI de l'Algérie dans GeoNames

<https://dbpedia.org/page/Algeria>

L'URI de l'Algérie dans DBpedia

GeoNames

Algeria ⓘ
country name : [Algeria](#) [\[info\]](#)
iso code : DZ, 012 and DZA
fips code : AO
capital : Algiers
area : 2,381,740.0 km²
population : 42,228,429
currency : Dinar (DZD)
languages : Arabic (ar-DZ)
neighbours : [Libya](#) [Western Sahara](#) [Libya](#) [Mauritania](#) [Tunisia](#) [Morocco](#) [Mali](#)
postal code format : ##### [\[postal codes\]](#)
national flag :



About: Algérie
An Entry of Type [Geo](#) from Named Graph: [http://dbpedia.org](#) within Data Space: [dbpedia.org](#)
L'Algérie (l'al-ǧazāʾir ; arabe : الجزائر (al-Jazāʾir) ; et arabe algérien : دزير (Dzayer) ou دزائر (Dzair) ; en tamazight : ⵍⵣⵣⵓⵢⵔ (Dzayer) est un pays d'Afrique du Nord faisant partie du Maghreb. Depuis 1962, elle est nommée en forme longue République algérienne démocratique et populaire, abrégée en RADP (en arabe : الجمهورية الجزائرية الديمقراطية الشعبية ; en tamazight : ⵜⴰⴳⴷⵓⴷⴰ ⵜⴰⵎⴻⵔⴰⵏⵜ ⵜⴰⵖⴻⵔⴰⵏⵜ ⵜⴰⵣⴻⵔⴰⵏⵜ (tagduda tameɣdayt tayerfant tazayrit). Sa capitale est Alger, la ville la plus peuplée du pays, dans le Nord, sur la côte méditerranéenne.

Property	Value
dbpedia:PopulatedPlace/area	<ul style="list-style-type: none"> 2381740.116324434 2381741.0
dbpedia:PopulatedPlace/populationDensity	<ul style="list-style-type: none"> 17.7 18.146880145149496
dbpedia:abstract	<ul style="list-style-type: none"> Algeria, officially the People's Democratic Republic of Algeria, is a country in North Africa. Algeria is bordered to the northeast by Tunisia; to the east by Libya; to the southwest by Niger; to the southwest by Mali, Mauritania, and Western Sahara; to the west by Morocco; and to the north by the Mediterranean Sea. It is considered part of the Maghreb region of North Africa. It has a semi-arid <i>steppe</i> with most of the population living in the

DBpedia

3.1 HTTP URIs

Cette partie décrit le principe 2 des données liées.

Les URLs font de bonnes URIs pour plusieurs raisons [8] :

- **Universalité** : Les URLs sont largement reconnues et utilisées dans le monde entier pour identifier et accéder à des ressources sur Internet.
- **Structure standardisée** : Les URLs suivent une structure standardisée.
- **Incorporation de l'emplacement** : Les URLs fournissent également des informations sur l'emplacement ou l'adresse de la ressource sur le Web.
- **Référencement** : Les URLs peuvent être utilisées pour accéder directement à une ressource en utilisant un navigateur Web ou d'autres clients HTTP. Cela les rend pratiques pour partager des liens vers des ressources spécifiques.
- **Interopérabilité** : Les URLs sont conformes à des normes et à des spécifications largement acceptées, ce qui garantit une interopérabilité élevée. Les navigateurs et les applications Web peuvent interpréter et utiliser les URLs de manière cohérente.

² Dans le contexte du Web et de la gestion des ressources, un "espace de noms" (namespace en anglais) est un moyen de garantir que les noms ou les identifiants de ressources sont uniques et ne créent pas de conflits.

- Les clients HTTP peuvent **déréférencer** (c'est-à-dire, rechercher, accéder via web) l'URI en utilisant le protocole HTTP et récupérer une description de la ressource identifiée par l'URI.

Lorsque les URI HTTP identifient des objets du monde réel ou des concepts abstraits, il est essentiel de ne pas confondre les objets ou les concepts eux-mêmes avec les documents Web qui les décrivent. Une ressource peut avoir différentes représentations ce qui permet de faire des déclarations distinctes sur une ressource et sur un document qui décrit cette ressource.

```

<html prefix="
  dbp: http://dbpedia.org/property/
  dso: http://dbpedia.org/ontology/
  dct: http://purl.org/dc/terms/
  dbid: http://dbpedia.org/datype/
  og: https://ogp.me/ns#
">
<!-- header -->
<head => </head>
<body about="http://dbpedia.org/resource/Algeria">
  <!-- navbar -->
  <nav class="navbar navbar-expand-md navbar-light bg-light fixed-top align-items-center"> <!--
  <div class="container-xl"> <!--
    <a class="navbar-brand" href="http://wiki.dbpedia.org/about" title="About Dbpedia" style="color: #2c5078"></a>
    <button class="navbar-toggler" type="button" data-bs-target="#dp-navbar" data-bs-toggle="collapse" data-bs-target="#dp-navbar" aria-controls="dp-navbar" aria-expanded="False" aria-label="toggle navigation"></button>
    <div class="collapse navbar-collapse" id="dp-navbar"> </div> <!--
  </div>
</nav>
</div> </div>
<!-- /navbar -->
<!-- page-header -->
<section>
<div class="container xl">
  <div class="row">

```

URL d'une description de la page HTML
destiné à être lu par un humain

<https://dbpedia.org/page/Algeria>

```

3 <?xml version="1.0" encoding="utf-8" ?>
4
5 <xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
6 <xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
7 <xmlns:foaf="http://xmlns.com/foaf/0.1/"
8 <xmlns:skos="http://www.w3.org/2004/02/skos/core#"
9 <xmlns:dop="http://dpsidia.org/property#"
10 <xmlns:geo="http://www.w3.org/2003/01/geo/geosparse#"
11 <xmlns:dbov="http://dpsidia.org/ontology#"
12 <xmlns:dcterms="http://purl.org/dc/terms/"
13 <xmlns:owl="http://www.w3.org/2002/07/owl#"
14 <xmlns:geonss="http://www.geonss.org/geonss/"
15 <xmlns:gold="http://purl.org/linguistics/gold/"
16 <xmlns:schema="http://schema.org/"
17 <xmlns:prov="http://www.w3.org/ns/prov#"
18 <xmlns:nsl="http://dpsidia.org/ontology/PopulatedPlace/" >
19   <rdf:Description rdf:about="http://dpsidia.org/resource/Cactus">
20     <dbowikiPageWikiLink rdf:resource="http://dpsidia.org/resource/Algeria" />
21   </rdf:Description>
22   <rdf:Description rdf:about="http://dpsidia.org/resource/Caetano_Luis_Pequito_de_Almeida_Sampaio">
23     <dbowikiPageWikiLink rdf:resource="http://dpsidia.org/resource/Algeria" />
24   </rdf:Description>
25   <rdf:Description rdf:about="http://dpsidia.org/resource/Caiazo_massacre">
26     <dbowikiPageWikiLink rdf:resource="http://dpsidia.org/resource/Algeria" />
27   </rdf:Description>
28   <rdf:Description rdf:about="http://dpsidia.org/resource/Cairo-Dakar_Highway">
29     <dbowikiPageWikiLink rdf:resource="http://dpsidia.org/resource/Algeria" />
30   </rdf:Description>
31   <rdf:Description rdf:about="http://dpsidia.org/resource/Calama_(Humidia)" />

```

URL du document RDF/XML Description
destinée à être lue par des machines

<https://dbpedia.org/data/Algeria.xml>

<http://dbpedia.org/resource/Algeria>

C'est un URI qui pointe vers la ressource Algeria dans DBpedia qui met a disposition des données sous forme de données liées RDF.

La figure suivante montre les relations souhaitées entre une ressource et ses documents de représentation.

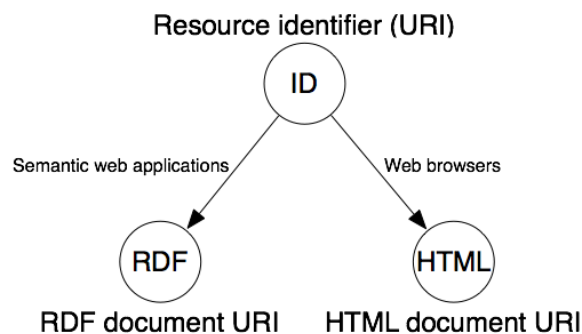


Figure 3 Les relations entre une ressource, ses documents de représentation [9]

7. Vocabulaires

Dans le contexte des données liées, les vocabulaires sont des ensembles structurés de termes, de concepts et de relations qui permettent de décrire des ressources. Ils font référence à des ensembles structurés de termes, d'ontologies, de schémas et de classes utilisés pour décrire des concepts, des entités et des relations entre les données. Ils sont essentiels pour garantir l'interopérabilité et la compréhension des données liées par différentes parties prenantes sur le Web. Voici quelques éléments clés associés aux vocabulaires :

1. Les ontologies sont des modèles de données conceptuels qui définissent des termes, des concepts et des relations spécifiques à un domaine donné. Elles fournissent une structure sémantique pour la description des données et facilitent la compréhension et l'interopérabilité.
2. Les vocabulaires définissent des classes (catégories de choses) et des propriétés (relations entre les choses) qui peuvent être utilisées pour annoter et lier des données. Par exemple, l'ontologie FOAF (Friend of a Friend) définit des classes telles que "Person" et des propriétés telles que "knows" pour représenter des relations sociales.
3. Les vocabulaires LOD utilisent RDF pour exprimer des déclarations sur des ressources, des relations et des métadonnées.
4. Les vocabulaires sont conçus pour être compatibles avec SPARQL, ce qui facilite l'extraction d'informations pertinentes à partir des données liées.
5. RDF Schema (RDFS) et OWL (Web Ontology Language) sont des langages de modélisation et de raisonnement pour définir des ontologies plus complexes. RDFS est utilisé pour des ontologies plus simples, tandis qu'OWL permet une modélisation plus avancée et des raisonnements sémantiques.

Une **ontologie** est une spécification formelle et explicite des concepts, des classes, des propriétés, des relations et des règles qui définissent un domaine de connaissance. Elle vise à capturer la sémantique et la signification des concepts et des relations dans ce domaine.

RDFS (RDF Schema) :

RDFS est utilisé pour définir des schémas de données simples en RDF (Resource Description Framework). Il permet de créer des hiérarchies de classes et de propriétés, et d'établir des relations basiques entre elles. Par exemple, on peut dire que la classe Animal est une sous-classe de la classe Être vivant, et que la propriété aNom a pour domaine la classe Animal et pour valeur une chaîne de caractères.

OWL (Web Ontology Language) :

OWL fournit un langage de modélisation sémantique plus puissant. Il est utilisé pour créer des ontologies riches en sémantique, décrivant en détail les concepts, les relations et les règles dans un domaine spécifique. Par exemple, on peut dire que la classe Chien est équivalente à

l'intersection de la classe Animal et de la classe aNom, et que la propriété estAmiDe est symétrique et transitive.

RDFS et OWL sont donc complémentaires pour décrire des ontologies. RDFS fournit les bases du vocabulaire, tandis qu'OWL permet d'enrichir la sémantique et de faciliter l'inférence de nouvelles connaissances à partir des données existantes.

6. Il existe de nombreux vocabulaires communs pour décrire des domaines spécifiques. Par exemple, DBpedia utilise le vocabulaire DBpedia Ontology pour décrire des concepts liés à l'encyclopédie en ligne.

L'utilisation de ces vocabulaires standardisés permet aux données liées d'être plus interopérables, car elles sont décrites de manière cohérente et partagée sur le Web. Cela facilite la découverte, l'intégration et la réutilisation de données par les chercheurs, les entreprises et d'autres parties prenantes dans le domaine des données ouvertes.

8. Standards

La pile, telle qu'elle est représentée par le W3C, est illustrée la figure suivante et se lit de bas en haut, commençant par les couches inférieures que nous venons d'évoquer, telles que les identifiants URI, jusqu'aux couches supérieures représentant l'utilisateur et l'interaction avec l'utilisateur.

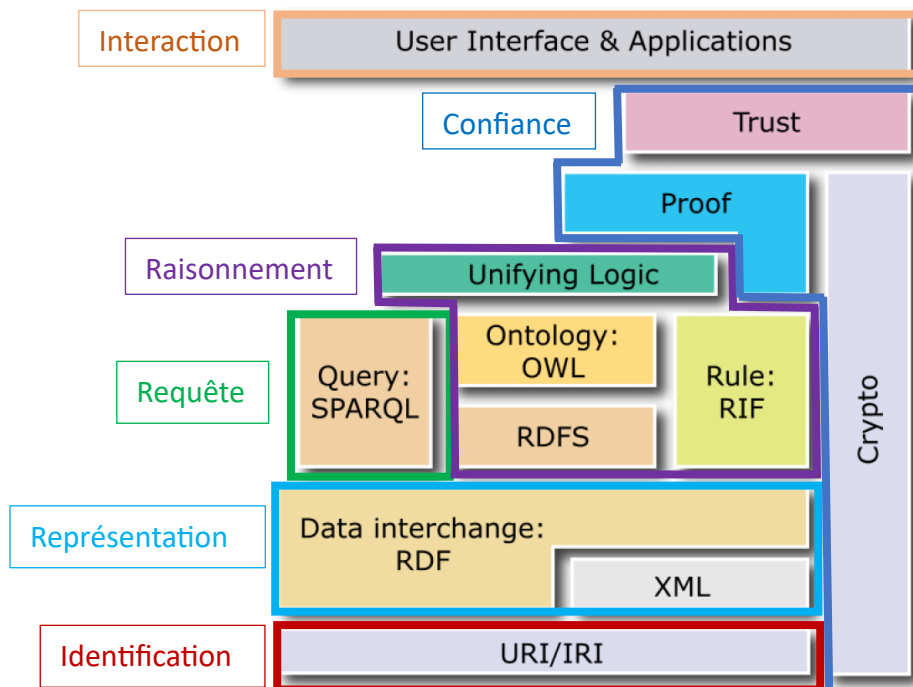


Figure 4 Pile des standards W3C [17]

- **Identification**

La première composante de cette pile concerne donc l'identification des ressources par des URIs.

- **Représentation**

La composante de cette pile concerne la représentation des données échangées sur le Web de données. Cette représentation repose sur le standard RDF qui est à la fois un langage et un modèle de données utilisés pour les échanges sur le Web de données dont l'une des syntaxes est celle basée sur XML.

- **Requête**

Une fois que les données sont mises en ligne, il devient essentiel de les interroger de manière sélective, en extrayant uniquement les informations pertinentes, au lieu de récupérer de vastes ensembles de données. Cela implique la nécessité de formuler des requêtes très précises pour extraire spécifiquement les résultats souhaités. Pour accomplir cela, le langage SPARQL est utilisé et qui permet de cibler de manière très précise des portions spécifiques des données disponibles sur le Web, en fonction des critères que nous définissons.

- **Raisonnement**

Le raisonnement en utilisant RDFS (Resource Description Framework Schema) et OWL (Web Ontology Language) est un élément essentiel dans la pile de standards W3C, car il permet d'exploiter pleinement les avantages de la modélisation sémantique et de l'ontologie pour déduire des informations, inférer des relations et effectuer des opérations logiques sur les données.

- **Confiance**

La pile de standards W3C est un ensemble de spécifications techniques qui définissent les principes et les pratiques pour le développement et l'interopérabilité du Web. Parmi ces spécifications, certaines concernent les domaines de la cryptographie, de la confiance et de la preuve, qui sont essentiels pour assurer la sécurité, la confidentialité et la vérifiabilité des données et des identités sur le Web.

- **Interaction**

Le terme "User Interface and Application" (interface utilisateur et application) dans la pile de standards du W3C fait référence à l'ensemble des normes et des technologies associées à la conception, à l'interaction et au développement d'interfaces utilisateur pour les applications web et les sites web. Cela englobe un large éventail de normes et de spécifications qui visent à améliorer l'expérience de l'utilisateur et à garantir la cohérence, l'accessibilité et la convivialité des interfaces web. Parmi ces normes nous pouvons citer : HTML, CSS, Web APIs, etc.

References

- [1] F. Gandon, "Les débuts du Web... sous l'œil du W3C," [Online]. Available: <https://interstices.info/les-debuts-du-web-sous-loeil-du-w3c/>.
- [2] F. Gandon, "Web sémantique et Web de données," [Online]. Available: <https://lms.fun-mooc.fr/>.
- [3] "Web sémantique," [Online]. Available: https://fr.wikipedia.org/wiki/Web_s%C3%A9mantique.
- [4] T. H. T. B.-L. Christian Bizer, "Linked Data - The Story So Far," *International Journal on Semantic Web and Information Systems (IJSWIS)*, 2009.
- [5] K. Graham, "Resource description framework (RDF): Concepts and abstract syntax," 2004. [Online]. Available: <http://www.w3.org/TR/rdf-concepts/> (2004).
- [6] B. Zizette, "Le Web Sémantique," Constantine.
- [7] H. Sack, "Knowledge Engineering with Semantic Web Technologies," Hasso-Plattner-Institut, Universität Potsdam.
- [8] D. Z. L.-S. Philippe GENOUD, "NOMMER LES DONNEES URIS UNIFORM RESOURCE IDENTIFIERS".
- [9] W3C, "Cool URIs," W3C, 2008. [Online]. Available: <https://www.w3.org/TR/cooluris/>. [Accessed 02 October 2023].
- [10] "Memex: hypothetical proto-hypertext system," [Online]. Available: <https://medium.com/@quartopiano/memex-hypothetical-proto-hypertext-system-bc0c60564eb8>. [Accessed September 2023].
- [11] T. H. Nelson, "A File Structure for the Complex, the Changing and the Indeterminate," *Association for Computing Machinery: Proceedings of the 20th National Conference*, p. 84–100, 1965.
- [12] T. Berners-Lee, "Information Management: A Proposal," *CERN*, March 1989, May 1990.
- [13] "Les 20 ans d'un web libre et gratuit," [Online].
- [14] C. Meilleur, "<https://knowledgeone.ca/du-web-1-0-au-4-0/?lang=fr>," 2021. [Online].
- [15] "W3C Process Flow," 2013. [Online]. Available: [https://www.w3.org/2013/dd-epasorg.htm#\(8\)](https://www.w3.org/2013/dd-epasorg.htm#(8)).
- [16] C. F.-Z. O. C. Fabien Gandon, *Le web sémantique*, Paris: Dunod, 2012.
- [17] F.-R. Chaumartin, "Antelope, a NLP platform for extracting meaning from text: theory and applications of the syntax-semantics interface," Thèse de doctorat. Université Paris-Diderot-Paris VII, Paris, France, 2012.