

Les données et métadonnées sur le Web

Du web de documents au web de données

ATED Tunis

Juin 2015

Nicolas Larrousse



Avant le commencement ...

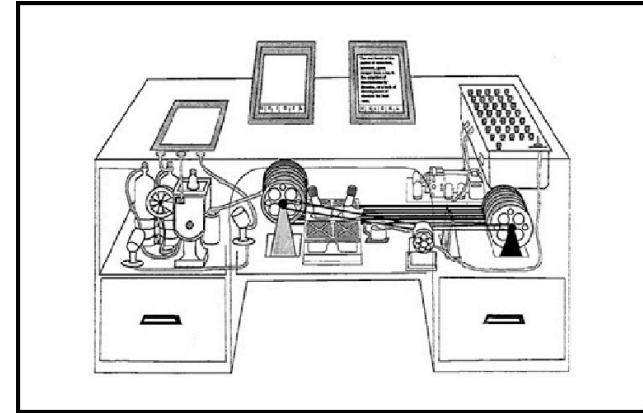
1945 : Vannevar Bush

Article « As we may think »

Comment gérer les montagnes de documents (livres, publications ...)

Retrouver ceux qui sont utiles et ceux qui leur sont liés

MemEx (MemoryExpander)



1965 : Ted Nelson

Article « Complex information processing: a file structure for the complex, the changing and the indeterminate »

Gérer l'évolution constante de l'information

ELF (Executable and Linking Format)

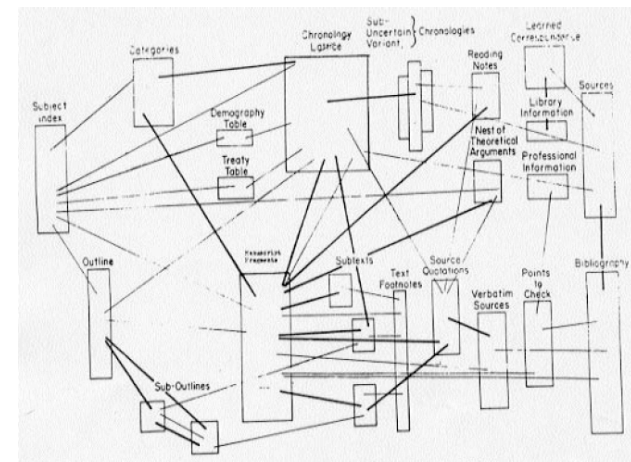


Figure 11.4. ELF's capacity for total filing: hypothetical use by historian. (A thin line indicates the presence of links; a heavy line indicates that some linked entries are identical.)

Au commencement ... Internet

- Adressage IP & Routage par paquets
- Protocoles de transport ... TCP/IP (1978)
- DNS (1983)
- Protocoles pour les applications ... Mail, FTP (1974/75)

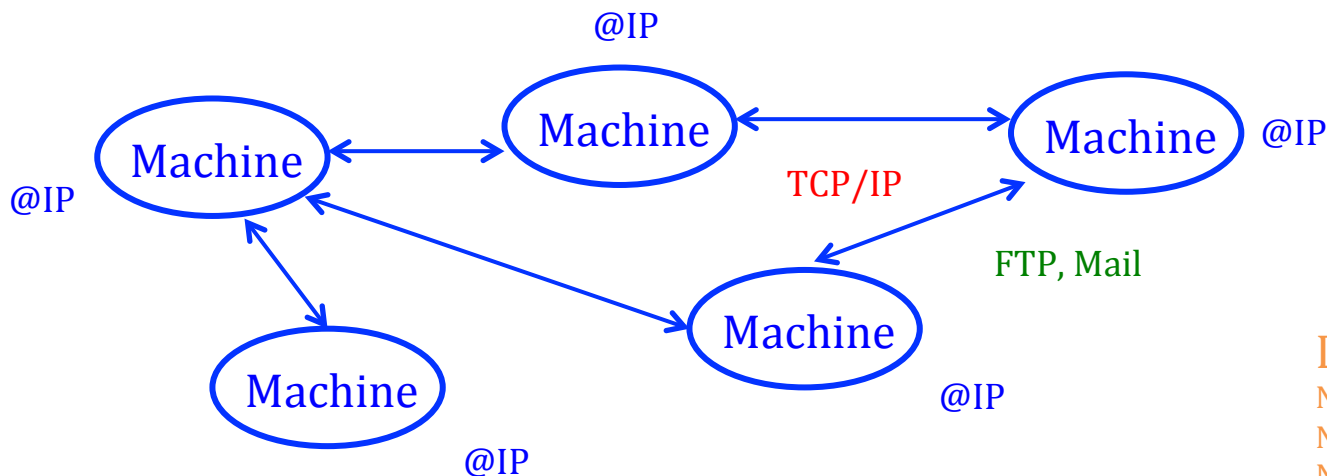
V. Cerf,

L. Pouzin,

P. Mokapetris

...

1970-1980



DNS

Nom -> @IP

Nom -> @IP

Nom -> @IP

Au commencement ... le Web

1990 : Tim Berners Lee

« Information management: a proposal »

- Notion d' Hypertexte
- Extension à internet
- Définition du langage HTML, du protocole HTTP
- Navigateurs (Mosaic, Netscape, IE)
- Recherche (Altavista ... fin de Gopher, Wais)

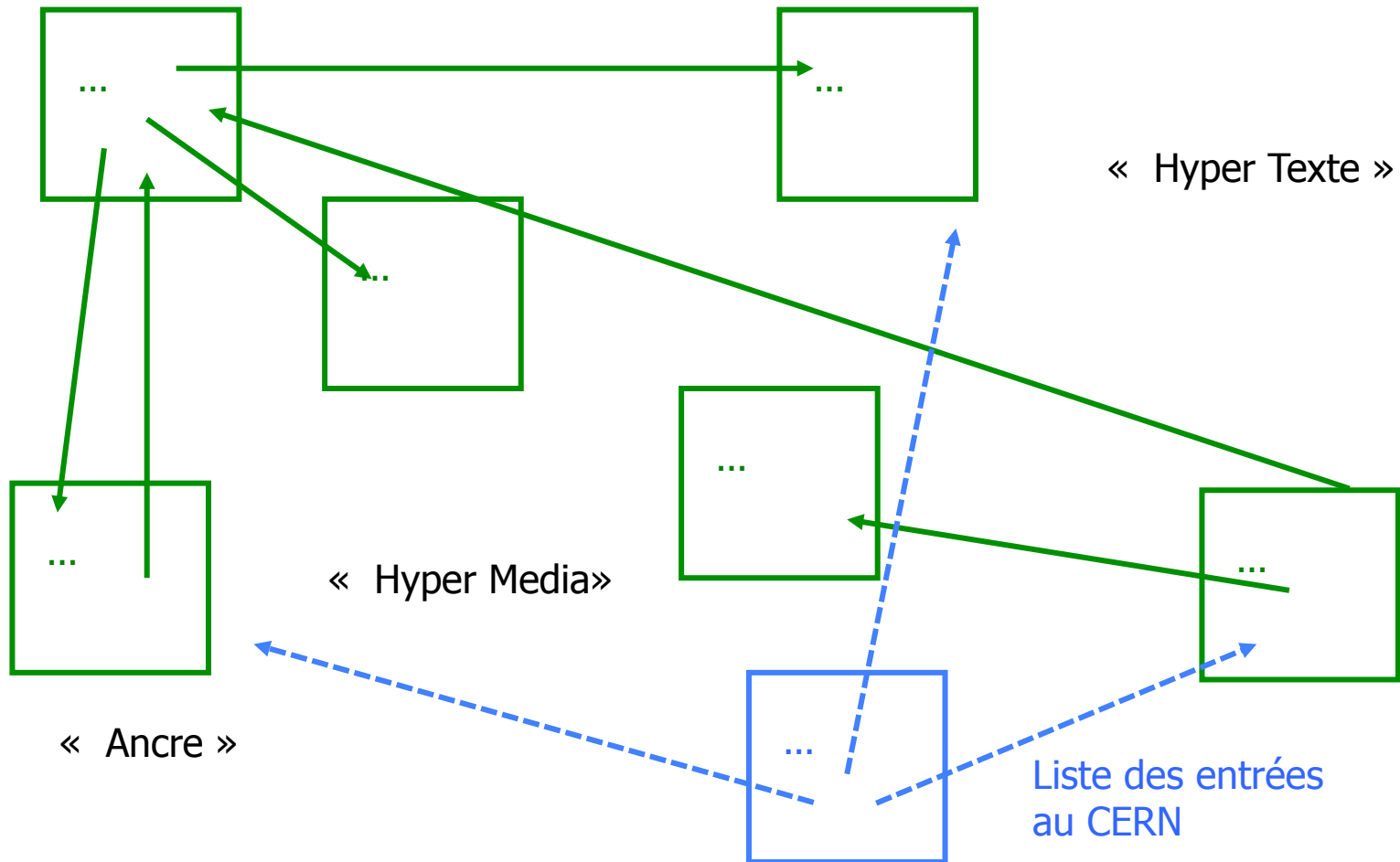


« The Mesh »

1994 - W3C - <http://www.w3.org>

Au commencement ... le Web

Un web de documents liés par des ancres



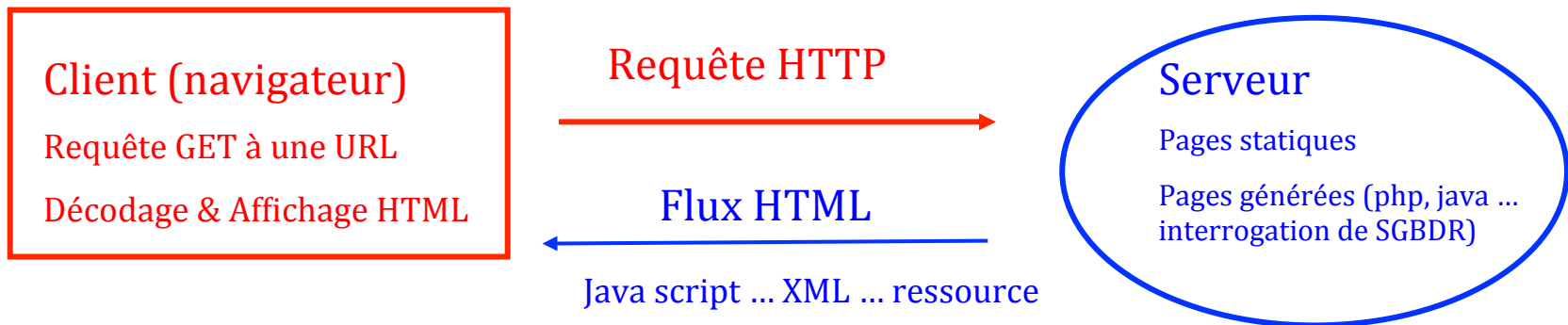
Au commencement ... le Web

Un mécanisme simple

La page devient

« calcul »

Notion d'URL



- 1995 HTML 2 (première évolution, formulaire ...)
- 1997 HTML 3 (applets, scripts, feuilles de style ...)
- 1999 HTML 4 (internationalisation, multimedia, scripts ...)
- > 2010 HTML5

Le web devient

« inscriptible »

Le contenu du Web

Riche en nombre mais pauvre en description

Pages HTML & autres ressources

- Peu ou pas de description normalisée du contenu
 - Peu ou pas de description structurée
 - Peu exploitable par une machine
- > Les moteurs de recherche fonctionnent « en aveugle »

Bases de données

- Les bases de données ne sont pas accessibles de manière normalisée mais par le programme qui les propose
- Il est nécessaire de connaître la structure et l'organisation des données pour les utiliser

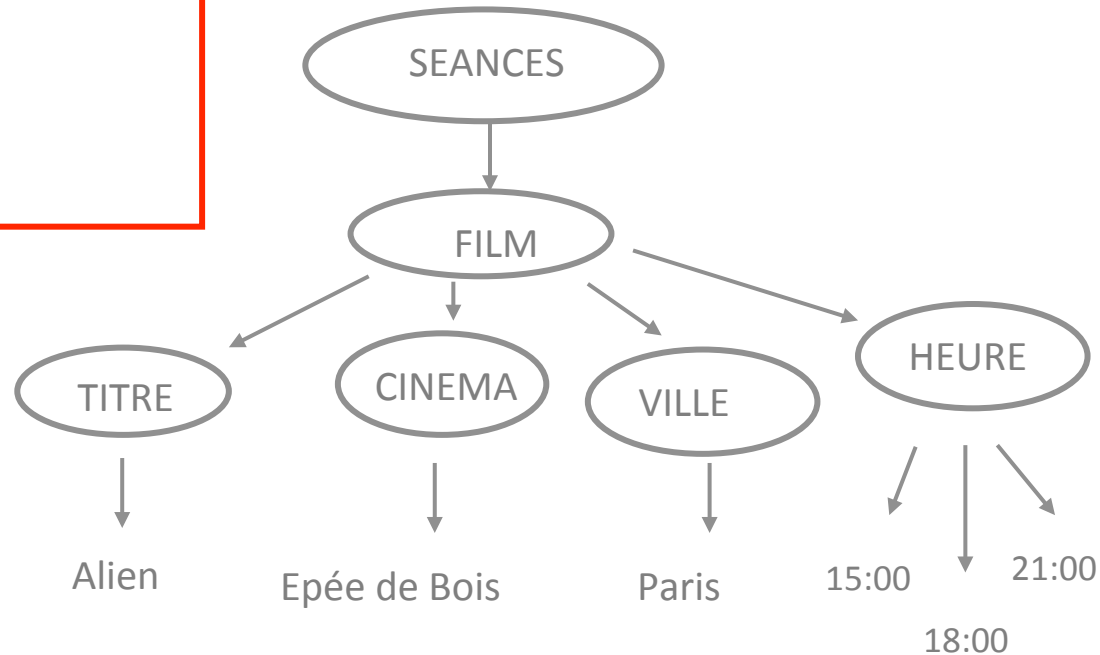
XML - une proposition de structuration

Fichier XML

DTD/Schema

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE SEANCES SYSTEM "seances.dtd">
<SEANCES>
<FILM>
<TITRE>Alien</TITRE>
<CINEMA>Epée de Bois</CINEMA>
<VILLE>Paris</VILLE>
<HEURE>15:00</HEURE>
<HEURE>18:00</HEURE>
<HEURE>21:00</HEURE>
</FILM>
</SEANCES>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT FILM ( TITRE, CINEMA, VILLE, URL?, HEURE+ )
>
<!ELEMENT TITRE ( #PCDATA ) >
<!ELEMENT CINEMA ( #PCDATA ) >
<!ELEMENT VILLE ( #PCDATA ) >
<!ELEMENT URL ( #PCDATA ) >
<!ELEMENT HEURE ( #PCDATA ) >
```



Un langage de structuration
&
Des outils
(contrôle, recherche ...)

XML - Séparer le fond et la forme

<?xml?>

Fichier XML

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE SEANCES SYSTEM "seances.dtd">
<SEANCES>
<FILM>
<TITRE>Alien</TITRE>
<CINEMA>Epée de Bois</CINEMA>
<VILLE>Paris</VILLE>
<HEURE>15:00</HEURE>
<HEURE>18:00</HEURE>
<HEURE>21:00</HEURE>
</FILM>
</SEANCES>
```



XPATH & Xquery
sur le fichier XML « pur »

MACHINE
&
PROGRAMME



HTML,
Pdf,
Epub ...

HUMAIN



Transformation (XSLT)

Feuille de style



<?xml?>

XML ... suite

- Permet une adaptation (eXtensible) à tout type/modèle de données structurées
- Permet un contrôle syntaxique et dispose d'outils puissants
- Est devenu un standard pour l'échange de données (e.g. Docx)
- Parfois utilisé comme format de base de données (e.g. BaseX)
- Sert de syntaxe à nombre de langages

Est parfois utilisé « curieusement » ...

Pas de description de la sémantique des relations

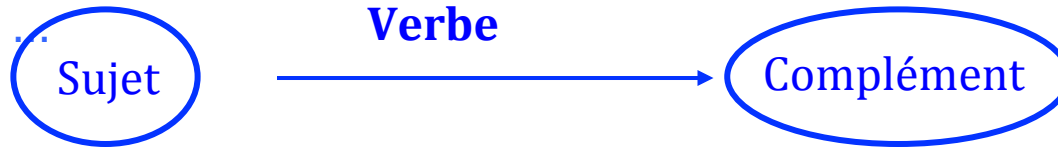
Adapté aux données structurées de manière arborescente ... ne permet pas de former des graphes

Comment représenter la(es) connaissance(s) ?

De la sémantique pour le Web :

Représenter le(s) modèle(s) du monde de chacun et les lier au Web

Un modèle simple ...



Tim Berners Lee

est né à

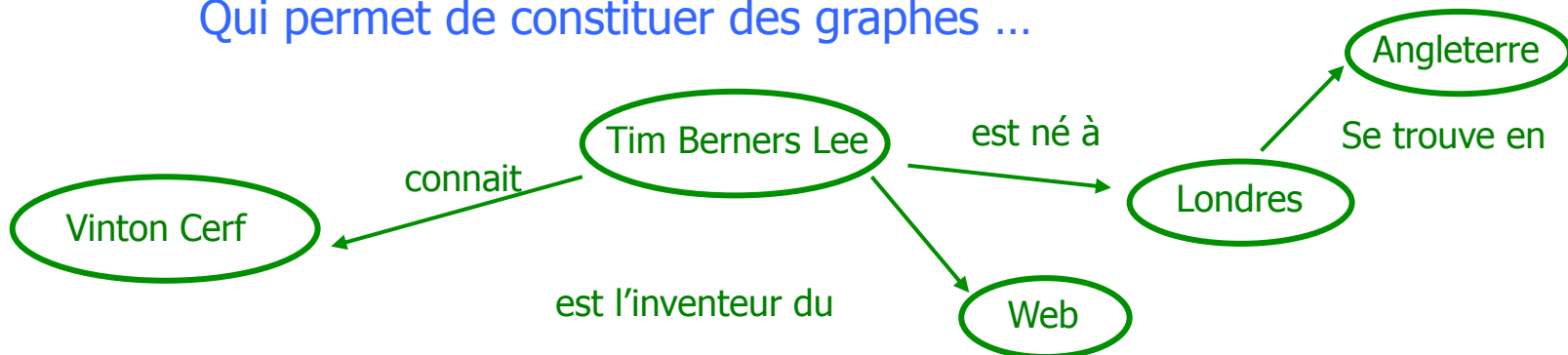
Londres



On constitue ainsi des « triplets » de type (sujet, prédicat, objet)

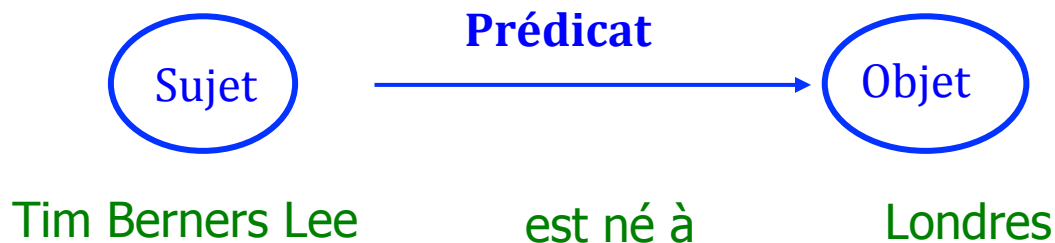
Le modèle proposé par le W3C se nomme **RDF** (Resource Description Framework)

Qui permet de constituer des graphes ...



Mise en œuvre des triplets pour/sur le Web

La notion d'URI



Comment représenter/désigner Tim Berners Lee sur le Web ???

Le web dispose de la notion d'**URI** qui permet d'identifier une ressource

*Une **URI** est toujours de la forme
[scheme][autorité][chemin]*

Une **URL** est une **URI** qui permet en plus d'accéder à représentation de la ressource

http://fr.wikipedia.org/wiki/Tim_Berners-Lee

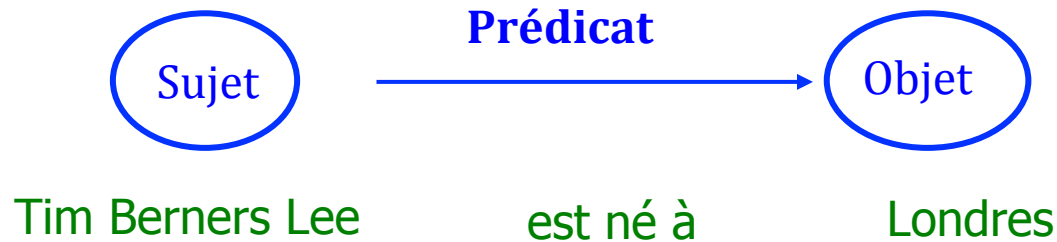
<hdl://11280/c3d77465>

<urn:ietf:rfc:2141>

<ark:/12148/cb34419111x>

Mise en œuvre des triplets sur le Web

*Tout est **URI***



Sujet

Il est **toujours** désigné par un URI

Prédictat

Il est **toujours** désigné par un URI

Il fait partie lui même d'un graphe (ensemble de triplets) qui décrit sa signification/son rôle.

Ce graphe décrivant un ensemble de concepts se nomme une **ontologie** exprimée dans le langage **RDFS** (ou **OWL** plus complet)

<http://xmlns.com/foaf/0.1/Person>

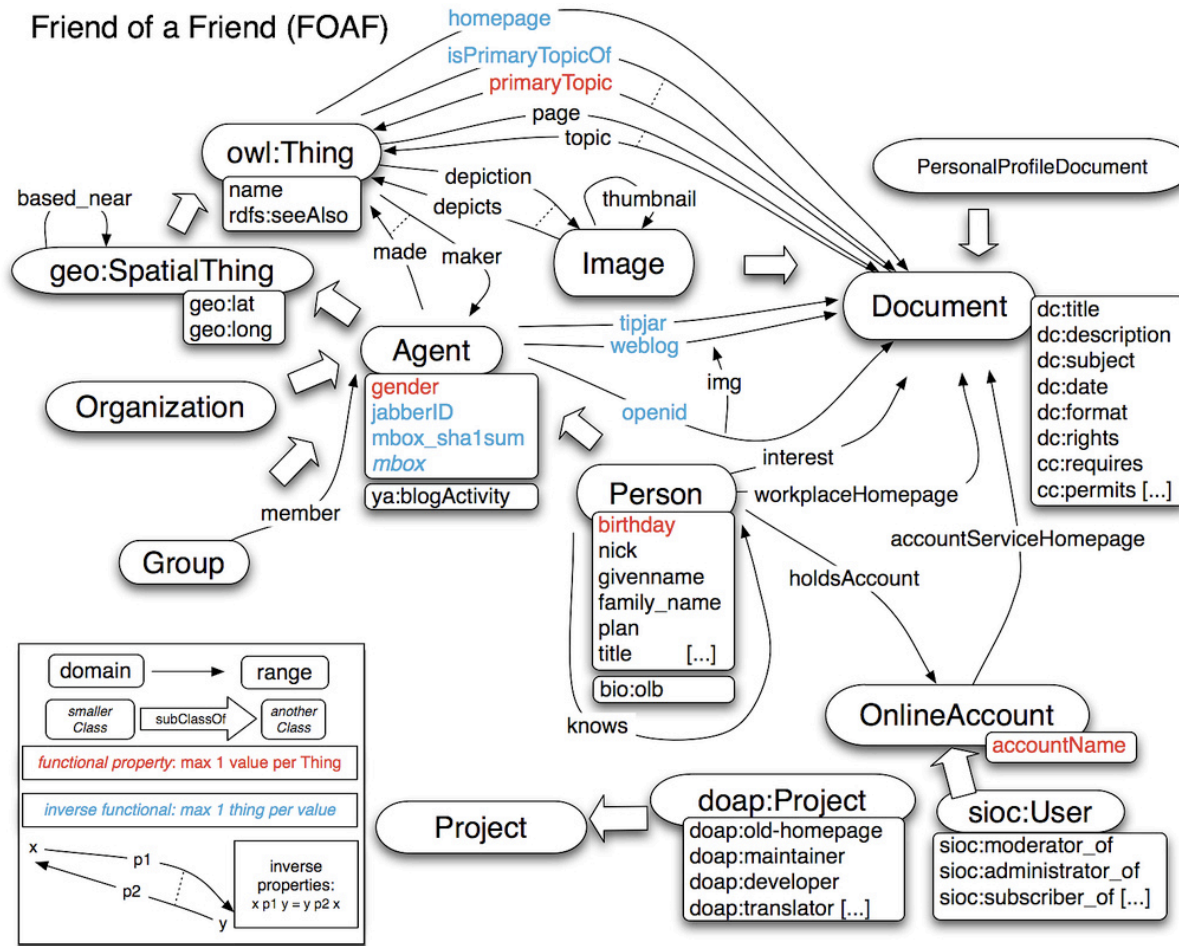
Objet

Peut être un URI ou de texte, nommé dans ce cas « littéral »

Un exemple d'Ontologie

Voir Linked Open Vocabularies <http://lov.okfn.org/dataset/lov/>

Foaf : Friend of a Friend



Un exemple de représentation

Tim Berners-Lee est de type "Personne"

Tim Berners-Lee a pour nom "Timothy Berners-Lee"

Tim Berners-Lee a créé "le Web"

Tim Berners-Lee a pour surnom "timbl"

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://www.w3.org/People/Berners-Lee/card#>

est de type

Personne

Tim Berners-Lee

a pour surnom

<http://xmlns.com/foaf/0.1/nick>

"timbl"

a créé

<http://xmlns.com/foaf/0.1/maker>

Le Web

<http://www.w3.org>

a pour nom

<http://xmlns.com/foaf/0.1/name>

"Timothy Berners-Lee"

<http://xmlns.com/foaf/0.1/Person>

Représentation de RDF en XML ou « sérialisation »

```
<rdf:RDF
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:dc="http://purl.org/dc/terms/">

  <rdf:Description rdf:about="http://www.w3.org/People/Berners-Lee/card#i">
    <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
    <foaf:name>Timothy Berners-Lee</foaf:name>
    <foaf:maker rdf:resource="http://www.w3.org"/>
    <foaf:nick>timbl</foaf:nick>
  </rdf:Description>

</rdf:RDF>
```


Le langage d'interrogation SPARQL

Un langage de recherche de triplets

```
SELECT *  
WHERE  
{  
?Un_sujet <URI_Predicat> <URI_objet>  
}
```

Un langage de parcours de graphe

```
SELECT *  
WHERE  
{  
?Un_sujet <URI_Predicat> <URI_objet>  
?Un_autre_sujet <URI_Predicat> ?Un_sujet  
}
```

Le web sémantique et de données

Le Web : URIs, URLs, HTTP etc.

Un modèle de graphe de triplets : RDF

Des langages de description : RDFS, OWL

Un langage d'interrogation : SPARQL

Un modèle auto-descripteur extensible

Totalement intégré au Web

Un entrepôt utile

Le projet « DbPedia »

About: **Mary Shelley**

An Entity of Type : [agent](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

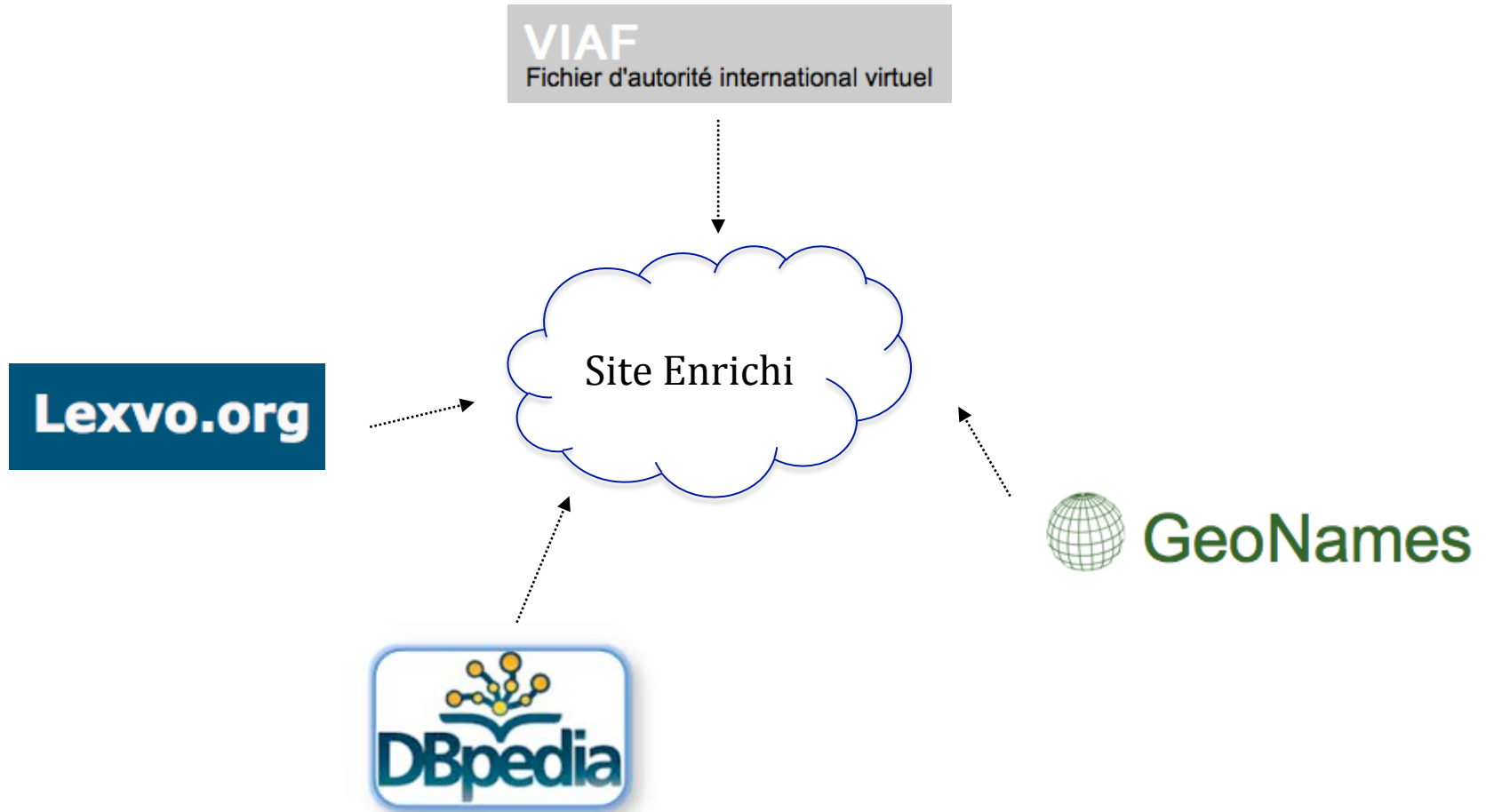


Mary Shelley, née Mary Wollstonecraft Godwin le 30 août 1797 à Somers Town, un faubourg de Londres (aujourd'hui dans le district de Camden), et morte le 1er février 1851 à Belgravia (Londres), est une femme de lettres anglaise, romancière, dramaturge, essayiste, biographe et auteur de récits de voyage.

Property	Value
dbpedia-owl:abstract	<ul style="list-style-type: none">■ Mary Shelley, née Mary Wollstonecraft Godwin le 30 août 1797 à Somers Town, un faubourg de Londres (aujourd'hui dans le district de Camden), et morte le 1er février 1851 à Belgravia (Londres), est une femme de lettres anglaise, romancière, dramaturge, essayiste, biographe et auteur de récits de voyage. Elle est surtout connue pour son roman Frankenstein ou le Prométhée moderne. Fille de la philosophe féministe Mary Wollstonecraft et de l'écrivain politique William Godwin, elle perd sa mère alors qu'elle-même n'est âgée que de onze jours. Son père se remarie quatre ans plus tard. Il offre à sa fille une éducation riche et l'encourage à adhérer à ses théories politiques libérales. En 1814, Mary Godwin entame une liaison avec un homme marié, partisan de son père Percy Bysshe Shelley. Accompagné de Claire Clairmont, la fille de la belle-mère de Mary, le couple voyage à travers l'Europe. Au cours des deux années qui suivent, Mary et Percy affrontent un endettement permanent et la mort de leur fille. Ils se marient en 1816, après le suicide de la première épouse de Percy. En 1816, lors d'un séjour près de Genève, Mary (devenue Mary Shelley) écrit son premier roman, Frankenstein. En 1818, les Shelley quittent la Grande-Bretagne pour l'Italie, où meurent leur deuxième et leur troisième enfant, avant que Mary Shelley ne donne naissance à son fils, Percy Florence Shelley, qui seul survivra. En 1822, son mari se noie dans le golfe de la Spezia, au cours d'une tempête. Un an plus tard, Mary Shelley retourne en Angleterre et, dès lors, se consacre entièrement à l'éducation de son fils et à sa carrière d'auteur. Les dix dernières années de sa vie sont marquées par la maladie. Elle décède d'une tumeur du cerveau le 1er février 1851. Jusqu'aux années 1970, Mary Shelley, outre son Frankenstein, est surtout connue pour les efforts qu'elle fit en vue de faire publier les œuvres de son mari. Les études récentes ont permis une vision plus complète de son œuvre et montré que Mary Shelley est restée toute sa vie une radicale sur le plan politique, soutenant l'idée que la coopération et la solidarité, pratiquées tout naturellement par les femmes au sein de leur famille, sont la voie qui permet de réformer la société civile.■ Mary Shelley (née Mary Wollstonecraft Godwin; 30 August 1797 – 1 February 1851) was an English novelist, short story writer, dramatist, essayist, biographer, and travel writer, best known for her Gothic novel Frankenstein; or, The Modern Prometheus (1818). She also edited and promoted the works of her husband, the Romantic poet and philosopher Percy Bysshe Shelley. Her father was political philosopher William Godwin, and her mother was the philosopher and feminist Mary Wollstonecraft. Mary Godwin's mother died when she was eleven days old; afterwards, she and her older half-sister, Fanny Imlay, were raised by her father. When Mary was four, Godwin married his neighbour, Mary Jane Clairmont. Godwin provided his daughter with a rich, if informal, education, encouraging her to adhere to his liberal political theories. In 1814, Mary Godwin began a romantic relationship with one of her father's political followers, the married Percy Bysshe Shelley. Together with Mary's stepsister, Claire Clairmont, they left for France and travelled through Europe; upon their return to England, Mary was pregnant with Percy's child. Over the next two years, she and Percy faced ostracism, constant debt, and the death of their prematurely born daughter. They married in late 1816 after the suicide of Percy Shelley's first wife, Harriet. In 1816, the couple famously spent summer with Lord Byron, John William Polidori, and Claire Clairmont near Geneva, Switzerland, where Mary conceived the idea for her novel Frankenstein. The Shelleys left Britain in 1818 for Italy where their second and third children died before Mary Shelley gave birth to her last and only surviving child, Percy Florence. In 1822, her husband drowned when his sailing boat sank during a storm near Viareggio. A year later, Mary Shelley returned to England and from then on devoted herself to the upbringing of her son and a career as a professional author. The last decade of her life was dogged by illness, probably caused by the brain tumour that was to kill her at the age of 53. Until the 1970s, Mary Shelley was known mainly for her efforts to publish Percy Shelley's works and for her novel Frankenstein, which remains widely read and has inspired many theatrical and film adaptations. Recent scholarship has yielded a more comprehensive view of Mary Shelley's achievements. Scholars have shown increasing interest in her literary output, particularly in her novels, which include the historical novels Valperga (1823) and Perkin Warbeck (1830), the apocalyptic novel The Last Man (1826), and her final two novels, Lodore (1835) and Falkner (1837). Studies of her lesser-known works such as the travel book Rambles in Germany and Italy (1844) and the biographical article for Dionysius Lardner's Cabinet Cyclopaedia (1829–46) support the growing view that Mary Shelley remained a political radical throughout her life. Mary Shelley's works often argue that cooperation and sympathy, particularly as practised by women in the family, were the ways to reform civil society. This view was a direct challenge to the individualistic Romantic ethos promoted by Percy Shelley and the Enlightenment political theories articulated by her father, William Godwin.
dbpedia-owl:alias	<ul style="list-style-type: none">■ Godwin, Mary Wollstonecraft
dbpedia-owl:birthDate	<ul style="list-style-type: none">■ 1797-08-30 (xsd:date)
dbpedia-owl:birthPlace	<ul style="list-style-type: none">■ dbpedia:England■ dbpedia:London
dbpedia-owl:birthYear	<ul style="list-style-type: none">■ 1797-01-01 (xsd:date)
dbpedia-owl:deathDate	<ul style="list-style-type: none">■ 1851-02-01 (xsd:date)

Relier les données de différents entrepôts

Le « Linked Open Data »

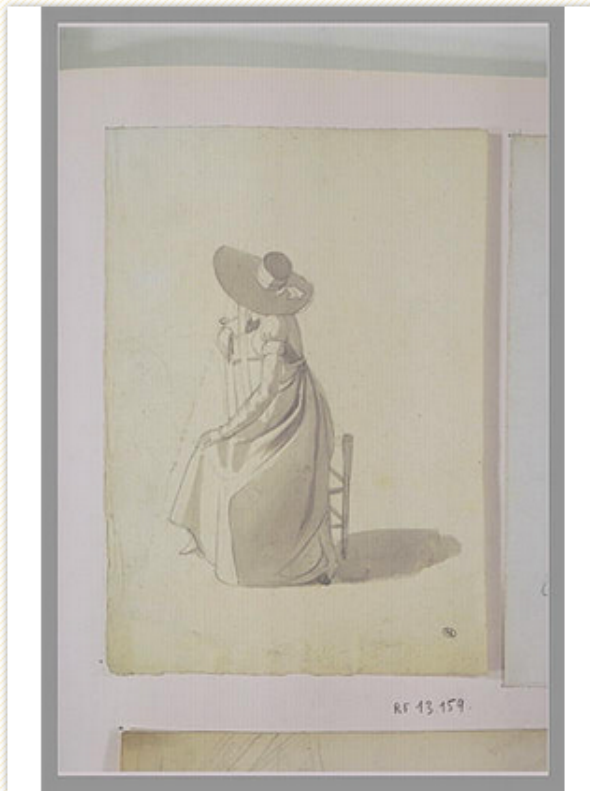


L'exemple de JocondeLab

[Français](#) [English](#) [Italiano](#) [Español](#) [Deutsch](#) [Português](#) [العربية](#) [Русский](#) [简体中文](#) [日本語](#) [Català](#) [Euskara](#) [Brezhoneg](#) [Occitan](#)

JocondeLab » Détail d'une notice

[Accueil](#) [À propos](#) [Carte du monde](#) [Chronologie](#) [Mots-clés](#) [Contribuer](#) [Galerie de créations](#)



Artiste(s) : [Jean-Augustin Franquelin](#)

École : [France](#)

Titre : **Femme assise sur une chaise, coiffée d'un chapeau, tournée vers la gauche**

Domaine(s) : [Dessin](#)

Désignation : **album**

Datation : [XIXe siècle](#)

Techniques : mine de plomb ; lavis gris

Mesures : H. en m 0,227 ; L. en m 0,160

Conservé à : Paris ; musée du Louvre département des Arts graphiques

Copyright notice : © musée du Louvre département des Arts graphiques, © Direction des Musées de France, 1997

Crédits photographiques : © Réunion des musées nationaux - utilisation soumise à autorisation

[Cette œuvre sur le portail Joconde](#)