WEB SÉMANTIQUE ET ONTOLOGIES

WEB DES DONNÉES

DONNÉES LIÉES (LINKED DATA)

Philippe GENOUD – Danielle ZIEBELIN - LIG-Steamer

Prénom.Nom@imag.fr

GI analyses des données, web des données, web sémantique

on pourrait ajouter : connaissances et raisonnement

- Partie 1 (5 à 6 séances de 3h, incluant TP):

  Interrogation du Web des données et Web sémantique

  - Modèles de données et de connaissances standardisés par le W3C (URIs, namespaces, RDF, RDFS, OWL)
  - Langage de requêtes SPARQL
  - Interrogation de données en présence d'ontologies
    - raisonnement automatique combiné avec SPARQL
  - Ontologie
  - Prenez votre propre ordinateur
  - Évaluation est un QCM avec des exercices : 40% de la note, en conditions d'examen, sans réseau, document autorisé : 1 feuille A4 manuscrite recto-verso.

- Partie 2 (5 à 4 séances de 3h, projet soit 14h): 7 trinômes, 2 binômes

  Conception d'un graphe de connaissances

  - Collecte de données pour établir un graphe de connaissances
  - Extraction d'information
  - Représentation connaissances : conception d'une ontologie, peuplement
  - Analyse sémantique des informations et raisonnement et proposition de requêtes

Make our datasets accessible and interoperable on the Web…

- For what ?
  - Not only because of H2020 requirement
  - Linking data increases its value
    - Mash up with related data
    - Produce new knowledge
    - Opportunity for new (unexpected) usage
  - Citizenship demand for access to public
  - data (scientific, government…)
  - …

Make our datasets accessible and interoperable on the Web…

- Publication/interlinking of open datasets
  - Publish heterogeneous data in a **common format**
  - Using **common vocabularies**
- Driven by major initiatives, e.g.:
  - Linking Open Data
  - W3C Data Activity
  - Open Data hosting services...
    - OpenAIRE, Zenodo...
- As well as other domain-specific projects
  - Bio2RDF, AgroPortal, BioPortal …

Make our datasets accessible and interoperable on the Web…

- Which data ?
  - All (values, map, images, sound ….) with Heterogeneous data models with Heterogeneous query capabilities

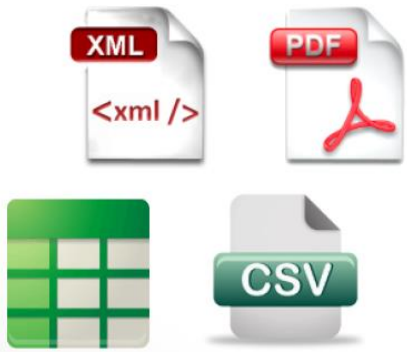
XML Native DBs,
Document stores…

```
<Books>
    <Book ISBN="0553212419">
        <title>Sherlock Holmes: Complete Novels...
        <author>Sir Arthur Conan Doyle</author>
    </Book>
    <Book ISBN="0743273567">
        <title>The Great Gatsby</title>
        <author>F. Scott Fitzgerald</author>
    </Book>
    <Book ISBN="0684826976">
        <title>Undaunted Courage</title>
        <author>Stephen E. Ambrose</author>
    </Book>
    <Book ISBN="0743203178">
        <title>Nothing Like It In the World</title>
        <author>Stephen E. Ambrose</author>
    </Book>
</Books>
```

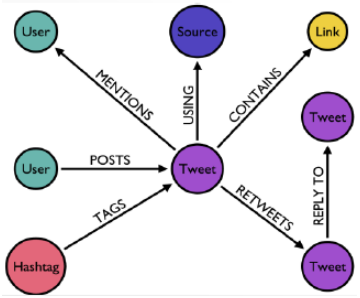
Tabular

ID	NAME	

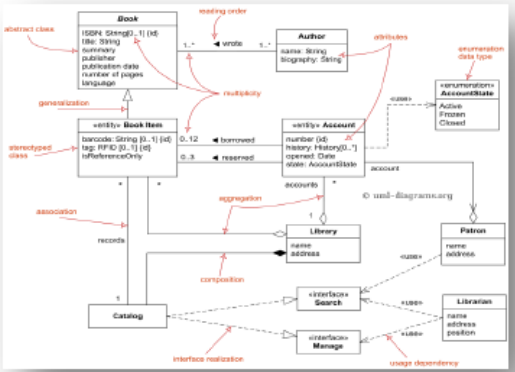
Documents



Graph



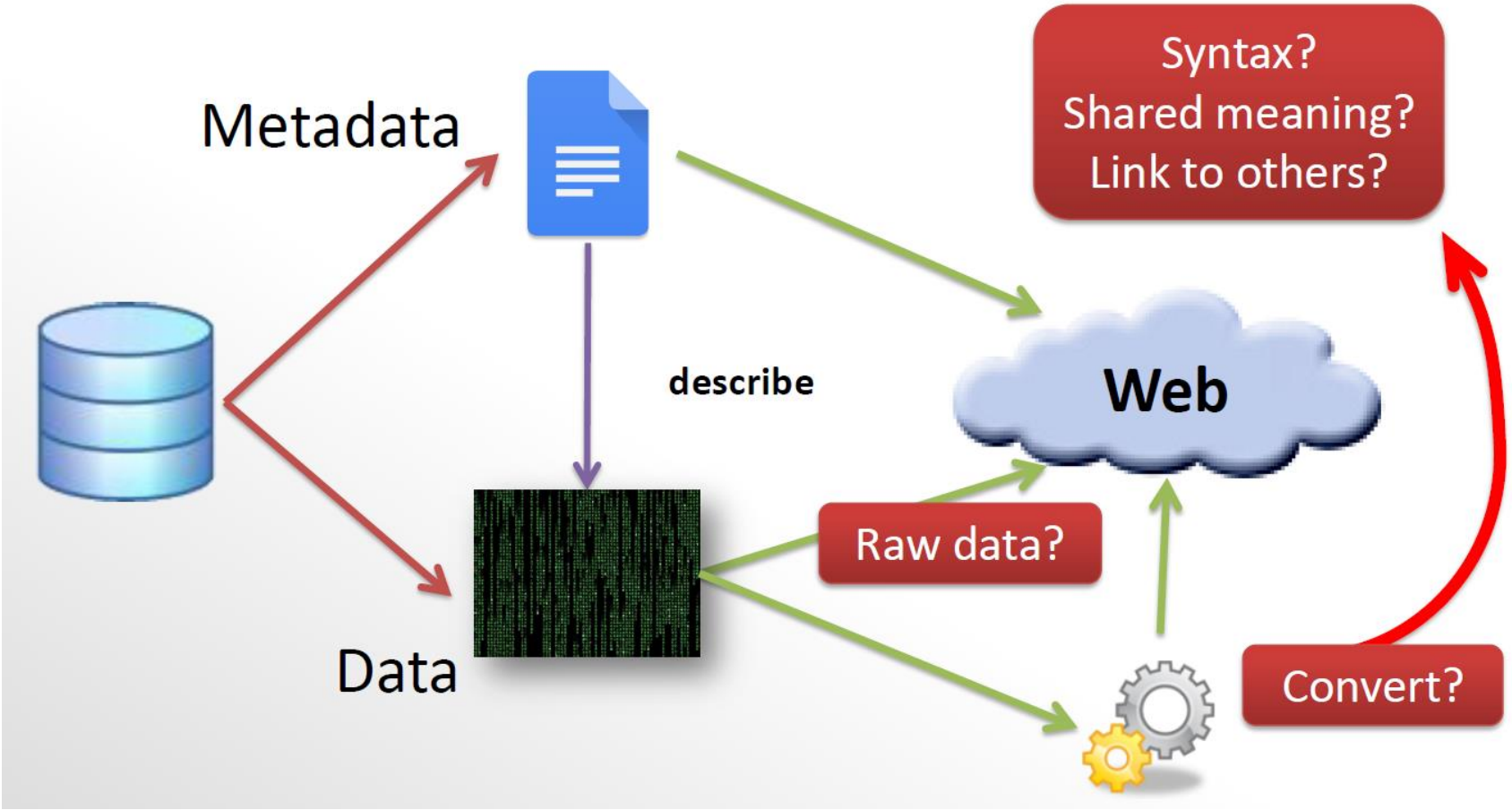
Object-Oriented



The key is Metadata

- Context: identification, authors, dates, license, version,
- reference articles
- Access: format, structure, location (dwld), query method
- Meaning:
  - What does it represent? What concepts, entities, semantics?
  - Interpretation: units (cm or inches, left/right)…
  - Provenance:
  - Acquired with what equipment? Parameters, protocols?
  - Derived from what dataset? With what processing?
  - Dataset-level or entity-level provenance
- Statistics
- Etc.

Challenges of publishing Metadata and/or Data ?



Crédit Franck Michel CNRS Université Côte d'Azur.

Make our datasets accessible and interoperable on the Web…

- How ?
  - Have a common representation format
    - => Structural heterogeneity
  - Have common ways to describe the data
    - Vocabularies, ontologies, thesaurus…
    - => Semantic heterogeneity
  - Have common ways to query the data
  - The Web of Data and the Semantic Web
    - Create, reuse and link vocabularies
    - Populate the Web of Data
    - Publish Linked Open Data on the Web

The Semantic Web

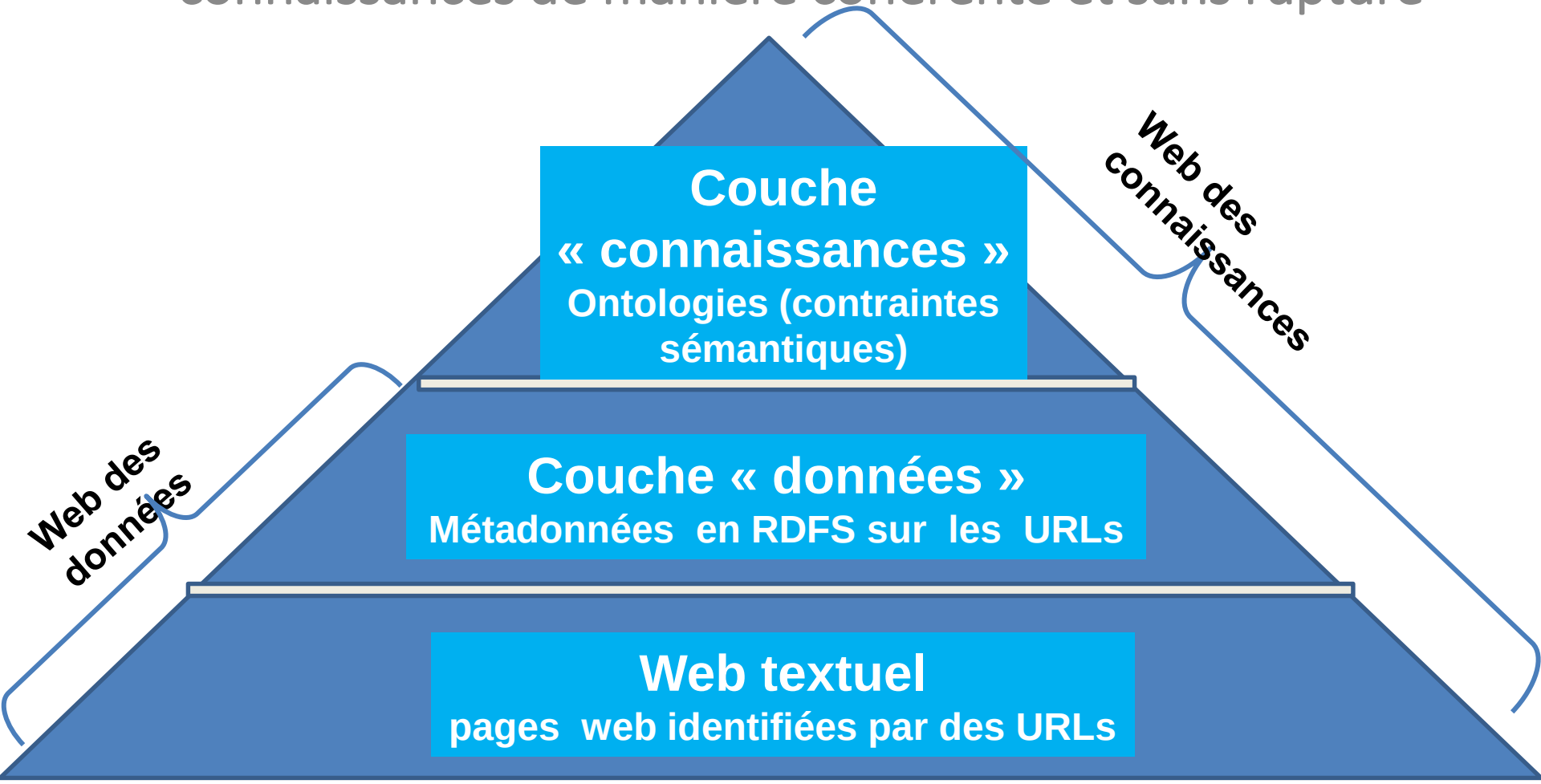
- Tim Berners-Lee's description :

" *The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The first steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will usher in significant new functionality as machines become much better able to process and "understand" the data that they merely display at present."*
Berners-Lee et al. 2001. The Semantic Web. Scientific American

- Automating tasks with the Web can be done by:
  - Making more powerful intelligent software agent on the client side. This is part of Artificial Intelligence.
  - Making a **more intelligible Web**: better standards, better practices, on both client and server sides. This is what the Semantic Web is about.

Une vision globale et cohérente

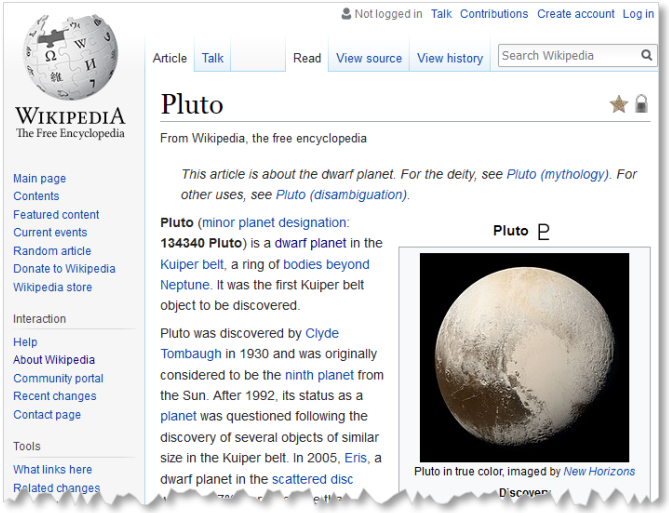
Le web est en train d'évoluer d'un web textuel à un web de connaissances de manière cohérente et sans rupture



From Web of Documents to Web of Data

"The **web of human-readable document** is being merged with a **web of machine understandable data.**"

- The traditional web (Web of Documents) is for humans
  - based on the HTML markup language
  - HTML describes
    - what information is presented + how it's presented in conjunction with CSS
    - how information is linked
    - but not what the information means
  - meaning (semantics) of information is derived from available information



Tim Berners-Lee, James Hendler and Ora Lassila , ***The Semantic Web***
Scientific American, 284(5), pp 34-43 (2001)

From Web of Documents to Web of Data

- How a machine knows?
- Understanding people's documents (text, image, sound, video) is **too difficult** for a machine (at the moment)
- Machines understand **formal languages** (programming languages, data formats, etc.)
- Make **knowledge** accessible to, and processable by machines in a formal language

From Web of Documents to Web of Data

"The **web of human-readable document** is being merged with a **web of machine understandable data.**"

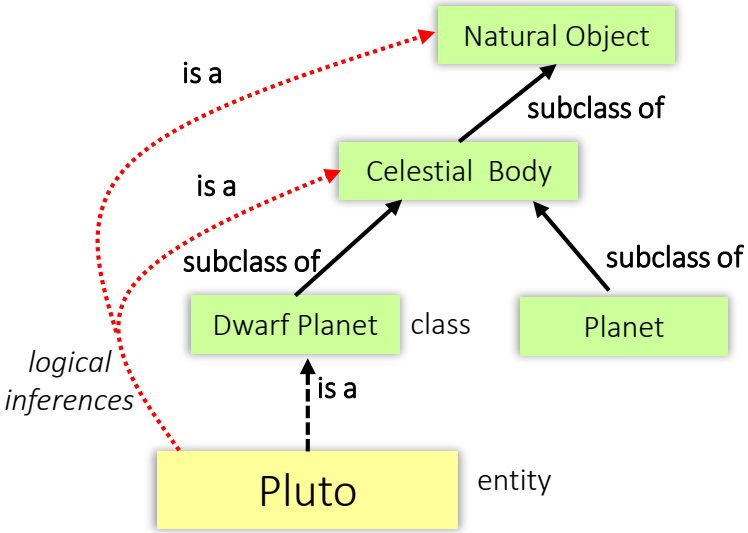
"The **semantic web is** an extension of the current web **in which information is given well-defined meaning**, better enabling computers and people to work in cooperation"

Tim Berners-Lee, James Hendler and Ora Lassila , ***The Semantic Web***
Scientific American, 284(5), pp 34-43 (2001)

- The Semantic Web (Web of Data) is for machines
  - meaning of entities is defined explicitly using formal (structured) and standardized knowledge representations (ontologies)

➔possibility to

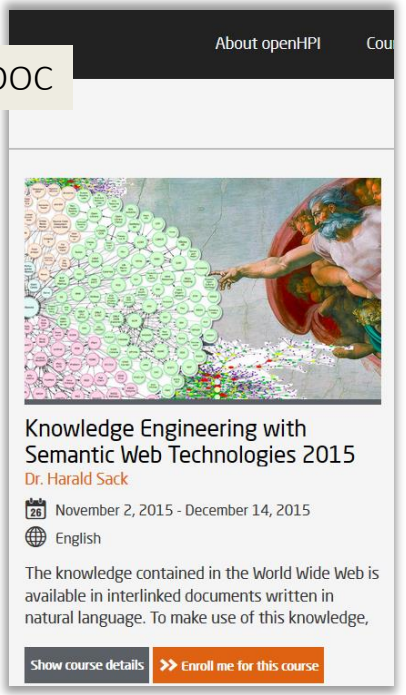
- process the meaning of information automatically
- automatically deduce implicit information from existing data
- relate and integrate heterogeneous data

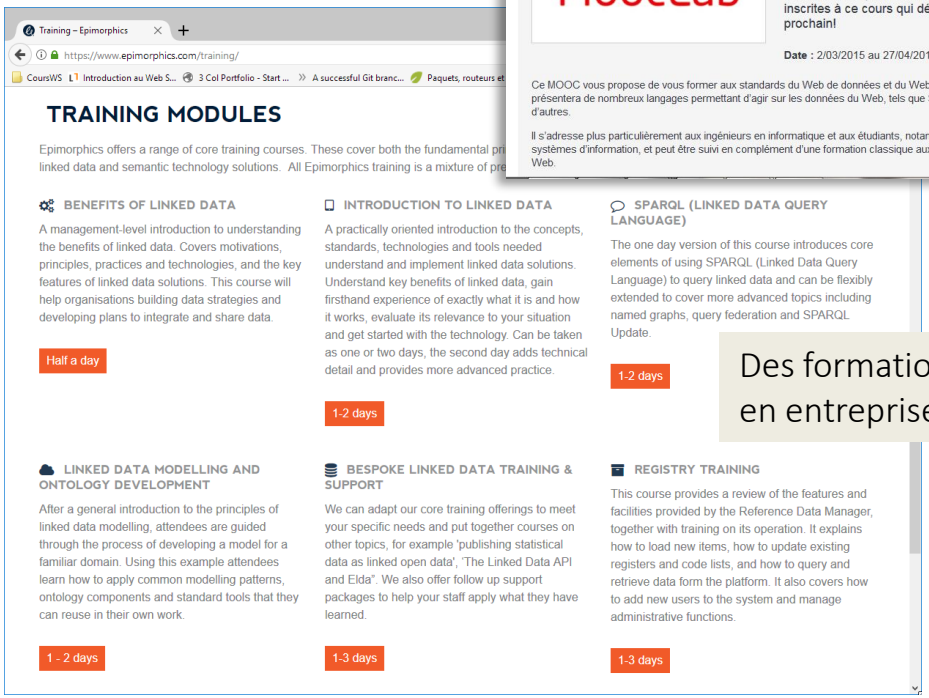


Pourquoi un cours web sémantique ?

- *On en parle*

des MOOC

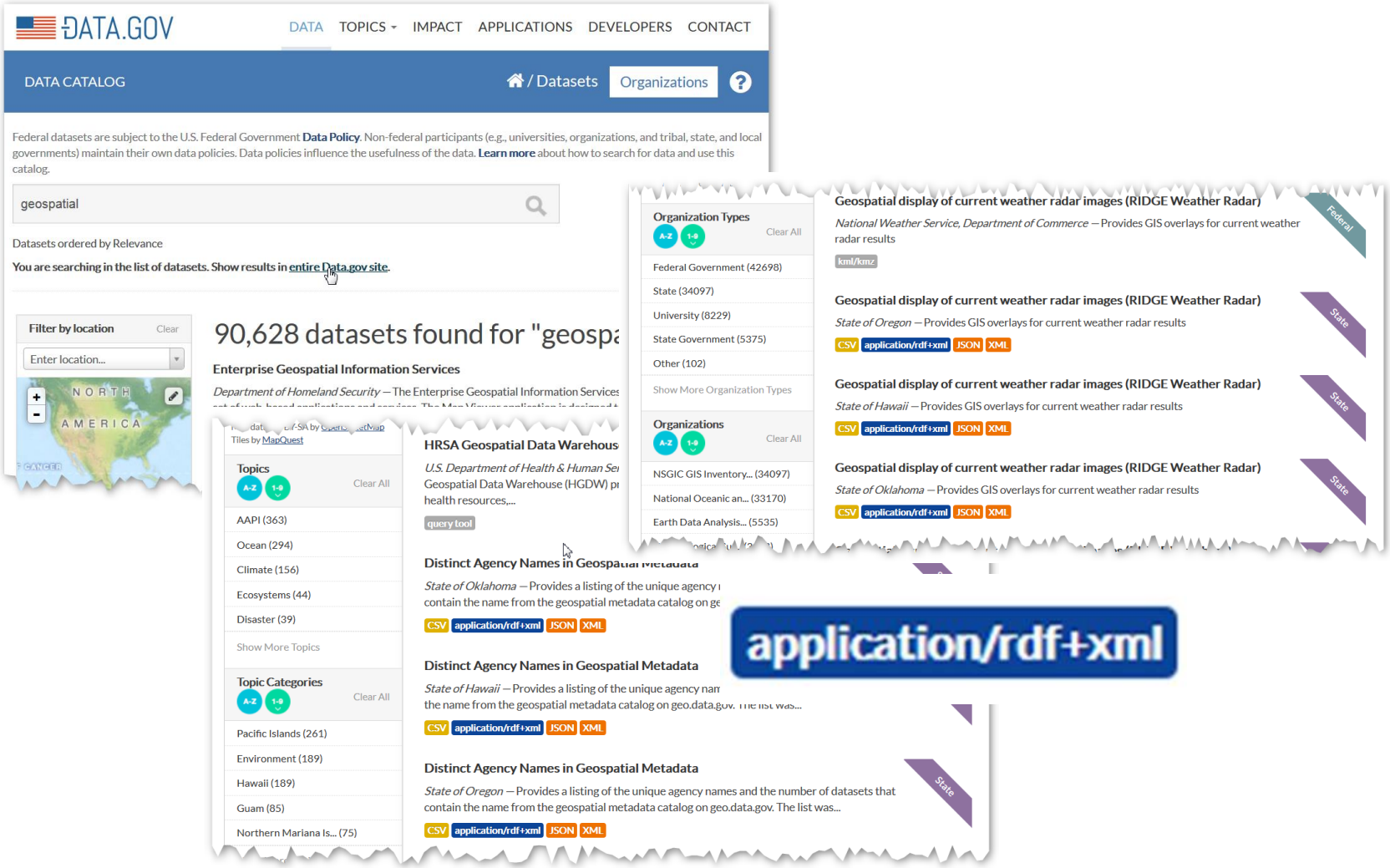




Des formations en entreprise

Pourquoi un cours web sémantique ?

Des technologies adoptées



Pourquoi un cours web sémantique ?

Des technologies adoptees : Plateformes de partage des données

- **Dataverse** : https://dataverse.org/
- **UData** : pour data.gouv.fr. https://github.com/opendatateam/udata
- **Koumoul** qui est un projet français et qui est utilisé notamment pour le site de l'ADEME. https://koumoul.com/
- **CKAN** : https://ckan.org/government pour Etats-Unis (data.gov) et Royaume-Uni (data.gov.uk)
- **Socrata** (Tyler Technologies) leader incontesté des offres payantes avec notamment comme client les villes de New York et San Francisco

Pourquoi un cours web sémantique ?

Des technologies en cours d'adoption

Australian Government Linked Data Working Group

http://www.linked.data.gov.au/

Home | Governance | Assistance | Showcase | Events | Groups | How To | Contact | Join

Australian Government Linked Data Working Group

The Australian Government Linked Data Working Group was established in August 2012 to meet the Linked Data challenges facing the Australian government.

Governance

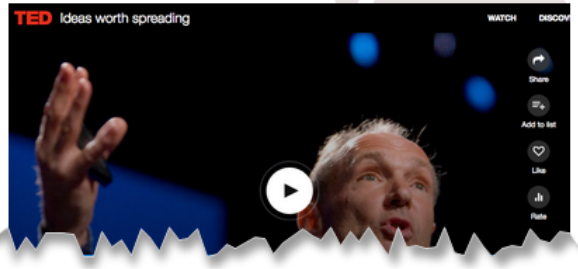
This Working Group is informal but some of its functions are recognised by multiple agencies, in particular a Memorandum of Understanding is in effect between 6 agencies (DTA, BoM, CSIRO, Dept. Finance, GA & NAA) which reserves the use of this web address, `linked.data.gov.au`, for Linked Data resources.

Read the MoU

As Linked Data technologies advance and become commonplace, it will be necessary for Government to become responsive to the demands of its citizens, as well as its own entities. Developing

About Linked Data

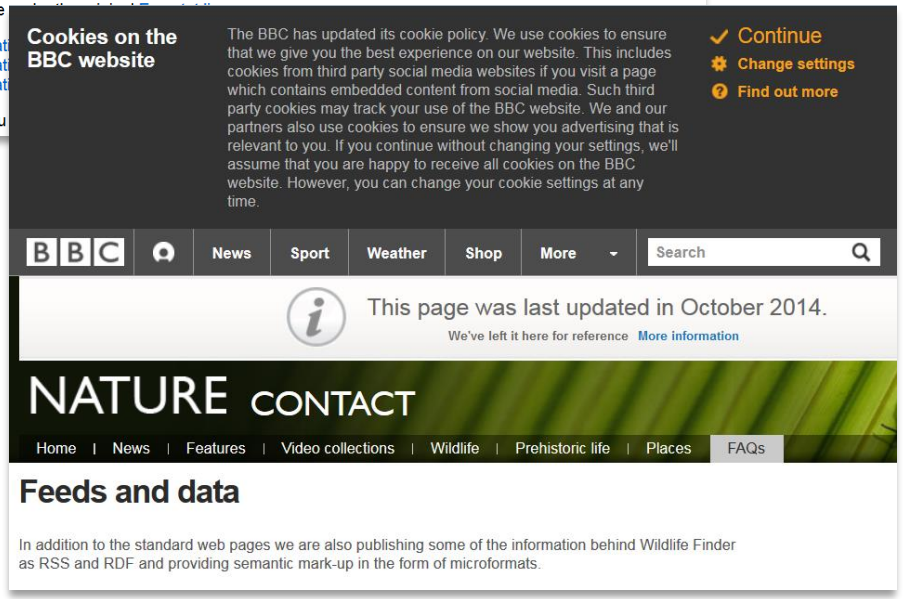
"*Linked Data*" refers to a set of standards, practices, and tools for publishing and linking structured data on the Web. Data that is Linked Data is linked to other data and can in turn be linked from other data. It is data that is published in a machine-readable way because all data is explicitly described in meaning and in format. For data publishers, it aims to efficiently maximise the capacity for interoperability and correct interpretation of published data. For data consumers it aims to maximise the efficient and correct re-use of data.



https://www.insee.fr/fr/information/2410909

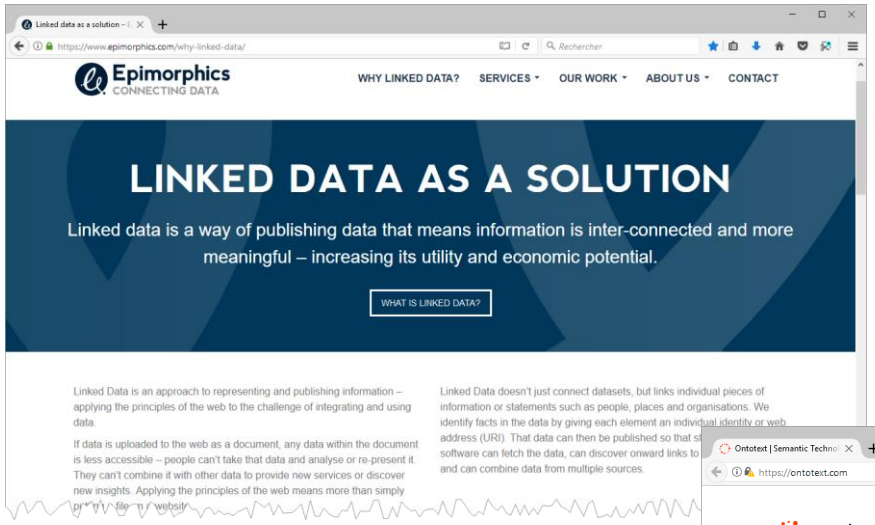
https://cros.ec.europa.eu/book-page/disseminating-official-statistics-linked-open-data-2024

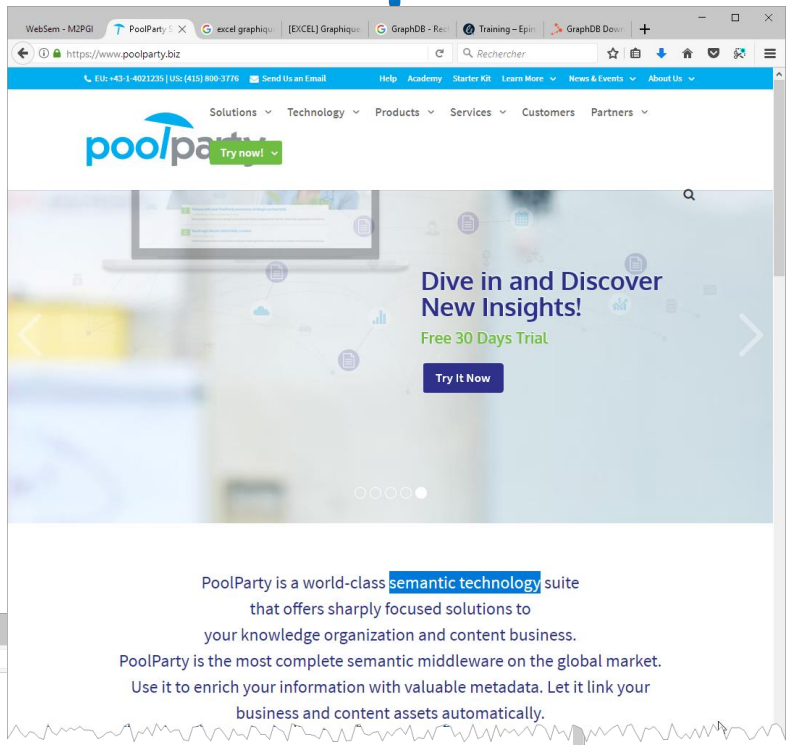
https://www.researchgate.net/publication/220853993_Media_Meets_Semantic_Web_-_How_the_BBC_Uses_DBpedia_and_Linked_Data_to_Make_Connections

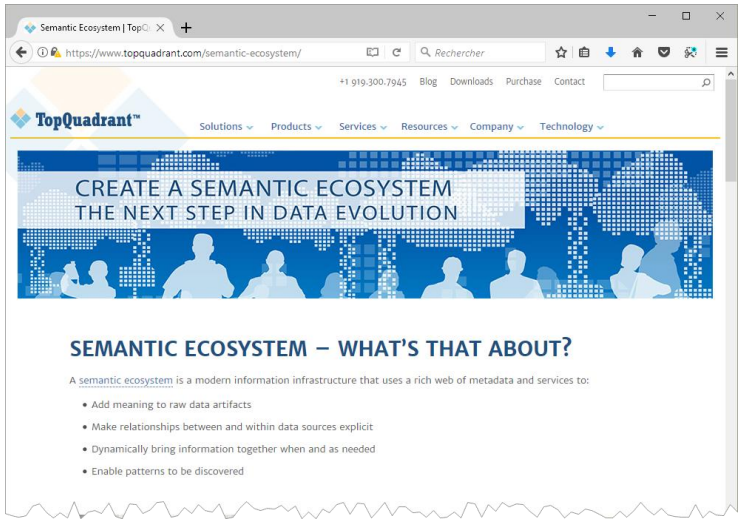


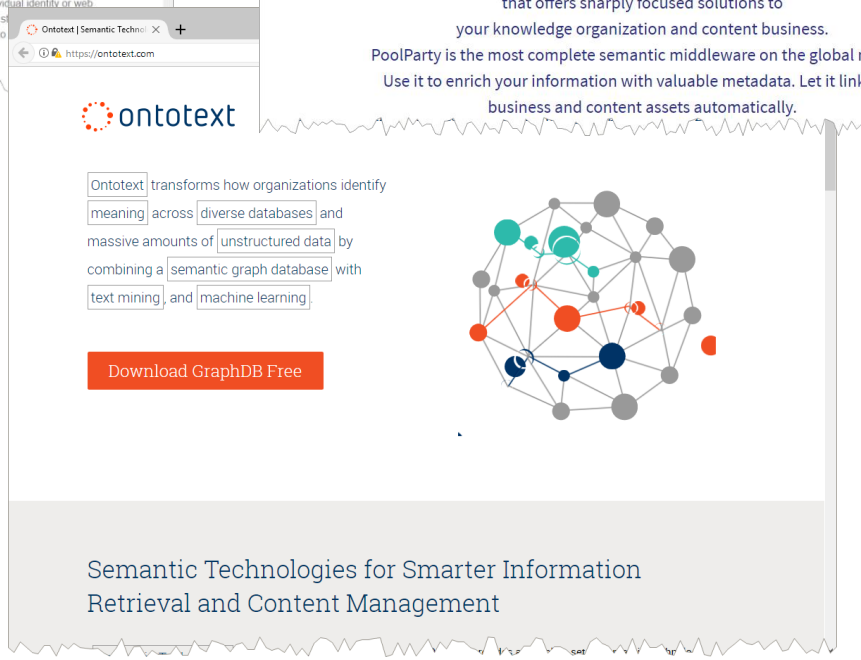
Pourquoi un cours web sémantique ?

- *des entreprises spécialisées*









Pourquoi un cours web sémantique ?

Des offres d'emploi...

Jobs / Job

VISEO

Viseo

France, Grenob

Apply for t

20 days left

19/09/2019

INRA SCIENCE & IMPACT

Description du poste

Ingénieur en développement de systèmes d'information

Contexte :

Le système d'information OpenSILEX (http://www.opensilex.org) est développé à l'INRA. Il aide à répondre aux défis majeurs de l'agriculture : produire des aliments sains, réduire l'impact environnemental ou encore s'adapter au changement climatique.

OpenSILEX est un logiciel « open source » avec une communauté internationale grandissante. Dans ce contexte dynamique, la personne recrutée intégrera l'équipe de développement de l'unité INRA MISTEA (Mathématiques Informatique STatistiques pour l'Environnement et l'Agronomie).

Missions :

La personne recrutée sera responsable de la conception, du développement, des tests et de l'intégration de solutions innovantes dans OpenSILEX.

Plus précisément :

- prendre part à la structuration et la valorisation de données massives (capteur, drones, analyse d'images, etc.)
- élaboration d'outils de visualisat
- élaboration de Web services d'é
- participation à des formations de

Connaissances et compétences atte

- langage Java maîtrisé et langages
- Web services REST appréciés
- langages du Web Sémantique (RD

La personne recrutée sera amenée à tra
acteurs externes du domaine. Un bon rel

Java Software Engineer for Machine Learning and Semantic Technologies

Semantic Web Company | Wien, Austria

€45k - 65k Visa sponsor Paid relocation

Overview Company Developer Culture More Jobs

About this job

Job type: **Full-time**

Experience level: **Junior, Mid-Level, Senior**

Role: **Backend Developer**

Industry: **Artificial Intelligence, Information Technology, Machine Learning**

Company size: **51–200 people**

Company type: **Private**

Technologies

spring-mvc java

ding provider of software and services in the areas
hine Learning, Natural Language Processing, and
PoolParty Software Platform is used in large
s and NGOs around the globe to extract meaning

lligence and innovation and rewards strong
l growth opportunities. We offer great work-life

Create alert

Background More

Sort by: matches

Technologies 6d ago

ntic Technologies 6d ago

artificial-intelligence machine-learning semantic-web spring-mvc java

andromeda

Cognitive Knowledge Platform™

viasema

01/10/2019

Ingénieur en Intelligence Artificielle H/F (Semantic Web)

Descriptif du poste

La société Viasema recherche un ingénieur ayant une parfaite compréhension des enjeux liés à la mise en oeuvre des technologies du Web Sémantique.

Profil recherché

Tous les profils sont susceptibles de nous intéresser mais évidemment, si vous matchez avec les critères suivants, vous partez avec une longueur d'avance :

- Aptitude à apprendre rapidement de nouvelles technologies et à se familiariser avec de nouveaux domaines d'activité.
- Le candidat doit en priorité maîtriser les technologies du Web sémantique (RDF/OWL, SPARQL...), les ontologies classiques comme SKOS, FRBR, Web Annotation, Schema.org...ainsi que les outils associés.
- Au delà des compétences techniques, nous recherchons un candidat très autonome et rigoureux, mais aussi bon vulgarisateur et capable d'être force de proposition auprès des équipes techniques comme des utilisateurs finaux ou de la direction.

pers and sales teams

ou de stages...

Stage Recherche Master Informatique M2
LIUPPA – Equipe T2I
2019-2020

Contexte général

Dans le cadre du projet de recherche FEDER TCVPYR[1] (Inventaire du patrimoine de villégiature dans les Pyrénées), nous av objectif de valoriser le patrimoine culturel matériel et immatériel de la villégiature et du thermalisme inventorié dans la « zone mas Pyrénées françaises. Nous visons d'une part la conception d'applications mobiles qui aideront les utilisateurs (experts et non e accéder aux informations relatives à ce patrimoine culturel. D'autre part, nous envisageons la publication automatique de d'inventaire réalisées sur Wikipedia.

Objectifs

Nous souhaitons concevoir et développer un système logiciel capable de publier automatiquement des articles Wikipédia à partir d d'inventaire collectées par des experts dans nos bases de données. A cette fin, une analyse détaillée des mécanismes de publica en œuvre dans l'univers Wikimédia devra être menée : étude de Wikidata, de WikiCommons et de Wikipédia.

De façon plus précise ... usieurs étapes :
Etat de l'art sur les ... ata; sur les liens entre les données (Linked Open Data)
Compréhension ... lément sur Wikimedia (Wikidata, WikiCommons, Wikipédia)
Appropriation ...
Développem ... s wikidata, publication automatique d'imag
WikiComm ... a la gestion des liens entre les 3 univers
WikiComm

de Wikipédia.
(Linked Open Data)
, WikiCommons, Wiki
data, publication

Compét

Nous rech ... uivalent) possédant un bon niveau de programmation et un espri et inventif.

Divers

Durée : 5 mois ; ... x légal en vigueur (environ 550€) ; **Localisation :** LIUPPA - Pau

Comment candi

Envoyer un CV ainsi que vos ... es deux dernières années à marie-noelle.bessagnet@univ-pau.fr, christian.sallaberry@univ-pau.fr, philippe.roose@univ-pau.fr;

UGA UFR IM²AG Université Grenoble Alpes

Web evolution

Web 1.0

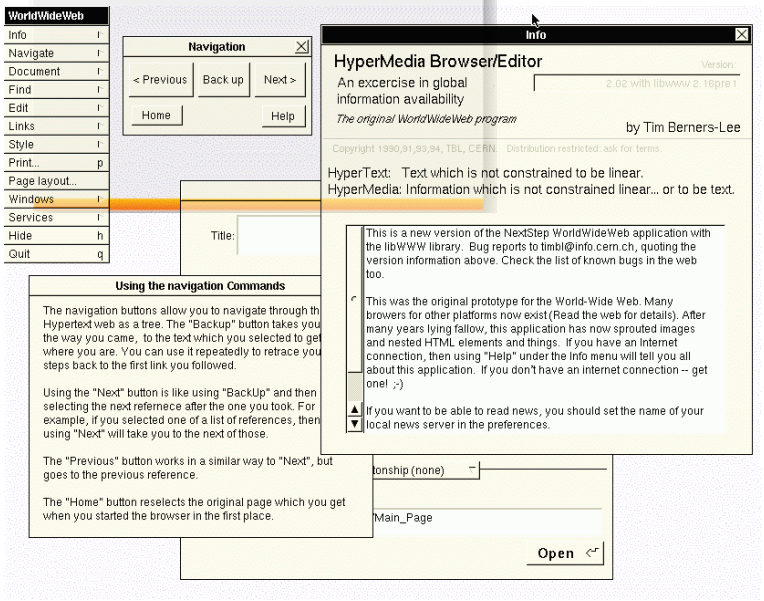
Web 2.0

Web 3.0

Web evolution

Web 1.0 *

- Utilisateur passif
- Essentiellement pro
- Réservé aux experts
- Guerre des navigateurs
- Contenu statique
- Formulaires
- Recherche par mots-clés



Web 2.0 *

Web 3.0

*d'après le cours Technologie du Web – A. Hombiat - Licence Professionnelle Études Statistiques et Systèmes d'Information Géographique (ESSIG) – UPMF - 2015

Web evolution

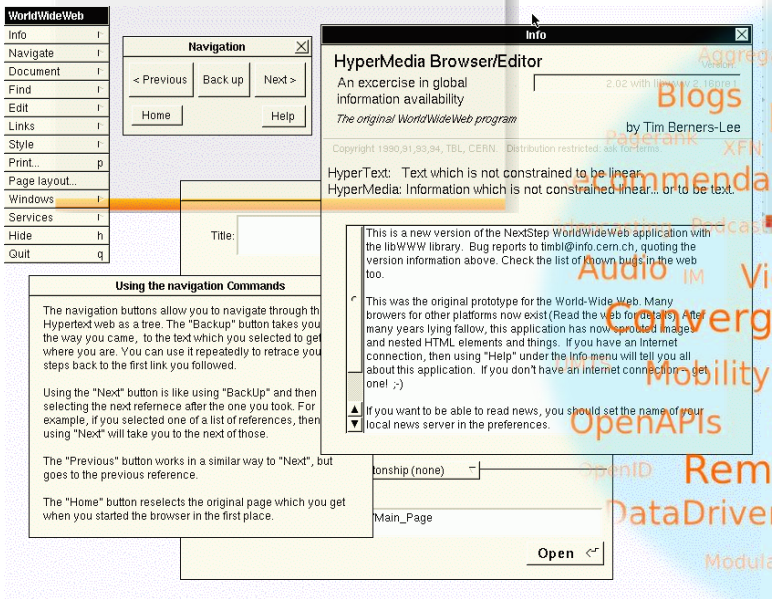
Web 1.0 *

- Utilisateur passif
- Essentiellement pro
- Réservé aux experts
- Guerre des navigateurs
- Contenu statique
- Formulaires
- Recherche par mots-clés

Web 2.0 *

- Utilisateur acteur (Blogs, Wikis, Réseaux sociaux, Podcasts, etc.)
- Ouvert aux profanes (CMS)
- Syndication du contenu (RSS, Atom)
- Tags et folksonomies
- Contenu dynamique
- Collaboration et partage
- Recherche par mots-clés

Web 3.0



*d'après le cours Technologie du Web – A. Hombiat - Licence Professionnelle Études Statistiques et Systèmes d'Information Géographique (ESSIG) – UPMF - 2015

Web evolution

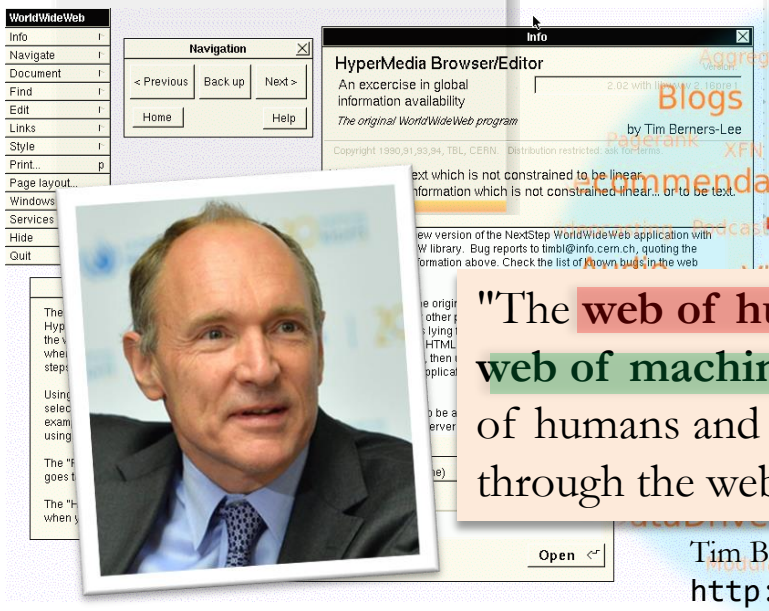
Web 1.0 *

- Utilisateur passif
- Essentiellement pro
- Réservé aux experts
- Guerre des navigateurs
- Contenu statique
- Formulaires
- Recherche par mots-clés

Web 2.0 *

- Utilisateur acteur (Blogs, Wikis, Réseaux sociaux, Podcasts, etc.)
- Ouvert aux profanes (CMS)
- Syndication du contenu (RSS, Atom)
- Tags et folksonomies
- Contenu dynamique
- Collaboration et partage
- Recherche par mots-clés

Web 3.0



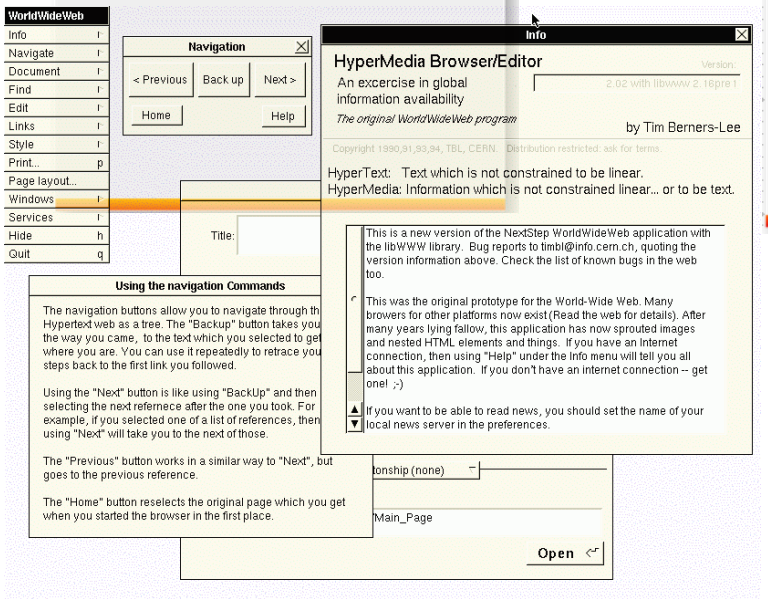
"The **web of human-readable document** is being merged with a **web of machine understandable data.** The potential of the mixture of humans and **machines** working together and communication through the web could be immense."

Tim Berners-Lee, The World Wide Web: A very short personal history, May 1998
http://www.w3.org/People/Berners-Lee/ShortHistory.html

Web evolution

Web 1.0 *

- Utilisateur passif
- Essentiellement pro
- Réservé aux experts
- Guerre des navigateurs
- Contenu statique
- Formulaires
- Recherche par mots-clés



Web 2.0 *

- Utilisateur acteur (Blogs, Wikis, Réseaux sociaux, Podcasts, etc.)
- Ouvert aux profanes (CMS)
- Syndication du contenu (RSS, Atom)
- Tags et folksonomies
- Contenu dynamique
- Collaboration et partage
- Recherche par mots-clés

Web 3.0

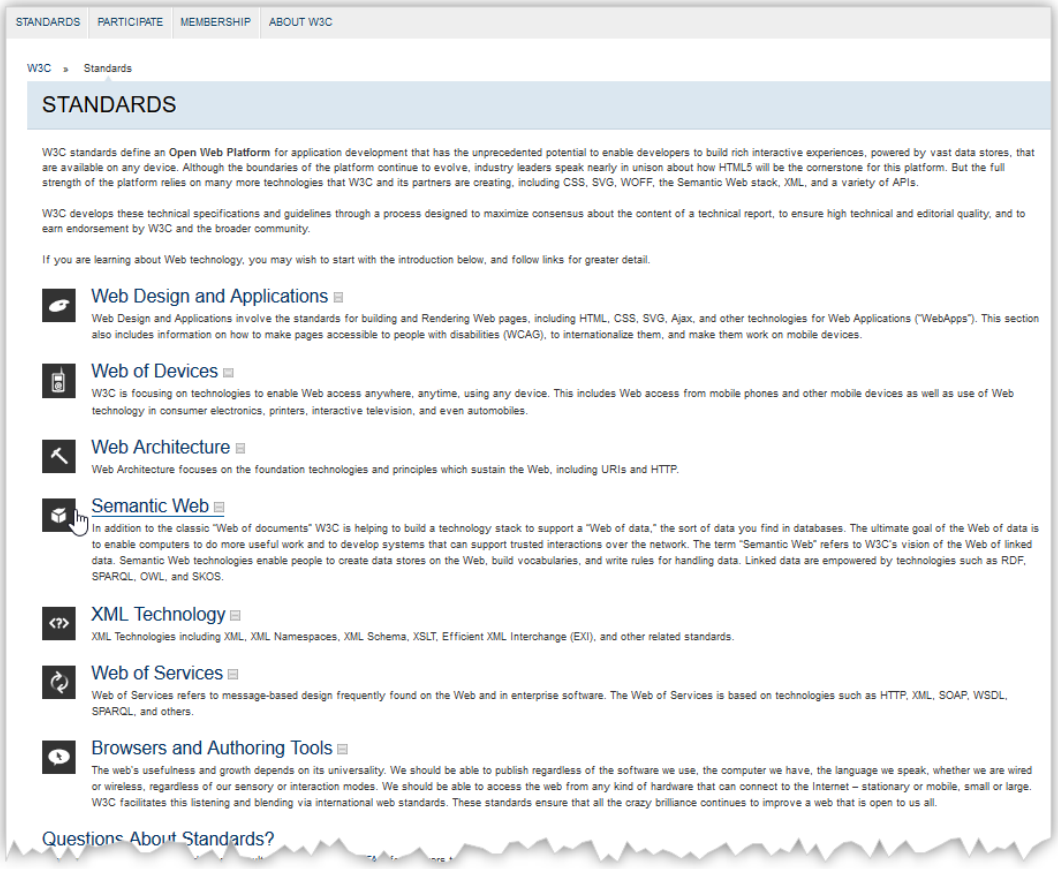
- Web sémantique
- Services internet (registres distribués blockchain),
- utilisateurs contrôlent leurs données (RGPD)
- I.A. exploitation des données
- visualisation

*d'après le cours Technologie du Web – A. Hombiat - Licence Professionnelle Études Statistiques et Systèmes d'Information Géographique (ESSIG) – UPMF - 2015

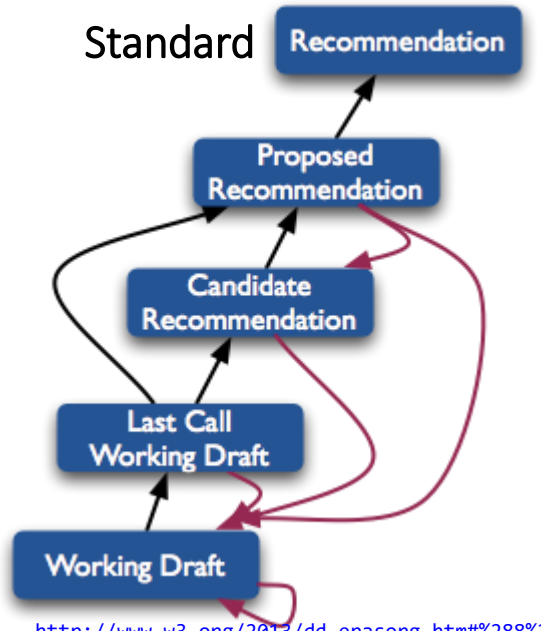
Web standardization

- 1994 creation of the W3C
- International consortium, around 400 Members
- Standardization of core Web technologies, publication of guidelines, technical notes, etc.

Standardization activities https://www.w3.org/standards/



Based on a transparent and public Process
https://www.w3.org/2019/Process-20190301/

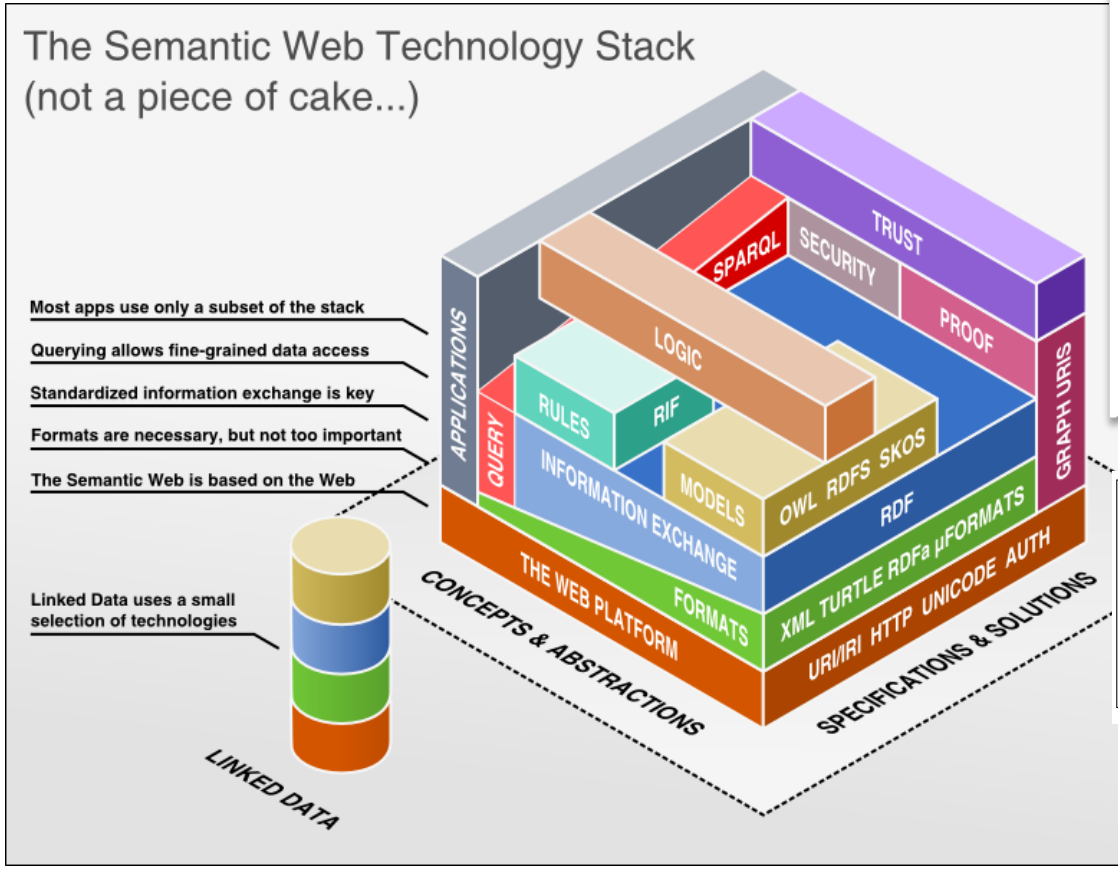


http://www.w3.org/2013/dd-epasorg.htm#%288%29

Technical reports, notes, recommendations
https://www.w3.org/TR/

Web standardization

- Semantic web standards



(Crédit Benjamin Nowack)

W3C SEMANTIC WEB ACTIVITY

On this page → publications, interviews • presentations • active groups • completed groups • past groups

W3C Semantic Web

The Semantic Web is a web of data. There is lots of data we all use every day, and it is not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar?

Why not? Because we don't have a web of data. Because data is controlled by applications, and each application keeps it to itself.

The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.

What is the Semantic Web?

The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF). See also the separate FAQ for further information.

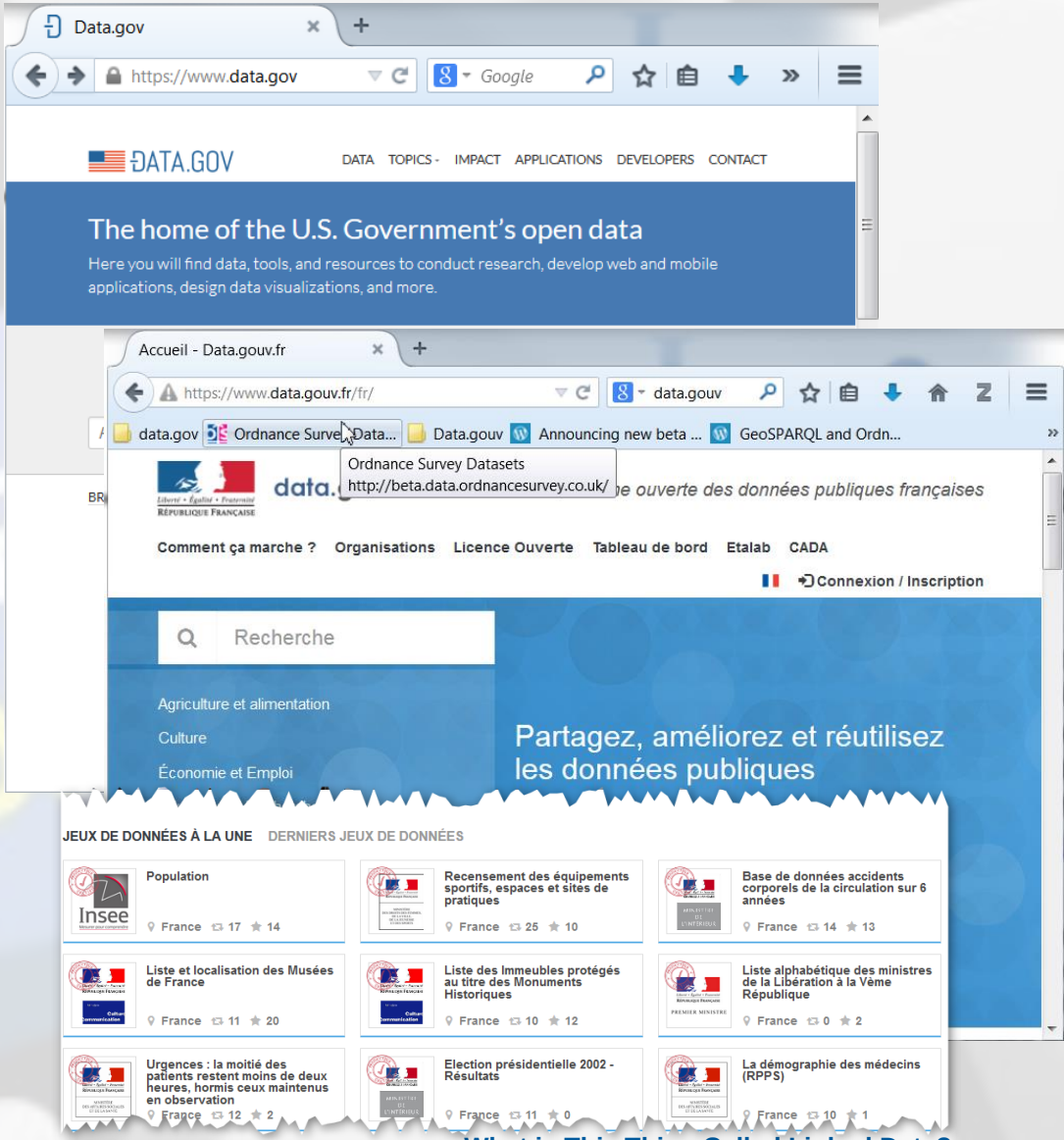
Warning: this Activity has been subsumed, in December 2013, by the W3C Data Activity. That activity has a larger scope; new or current Working and Interest Groups related to "traditional" Semantic Web technologies are now part of that Activity.

The current page has been frozen on the 11th December, 2013.

https://www.w3.org/2013/data/

Open Data

- Open Data Movement
  - *"A piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike."* http://opendefinition.org/
  - an old idea but a recent term gaining popularity
    - with the rise of the Internet and World Wide Web
    - with the launch of open-data government initiatives such as Data.gov (USA), data.gouv.fr
      ...

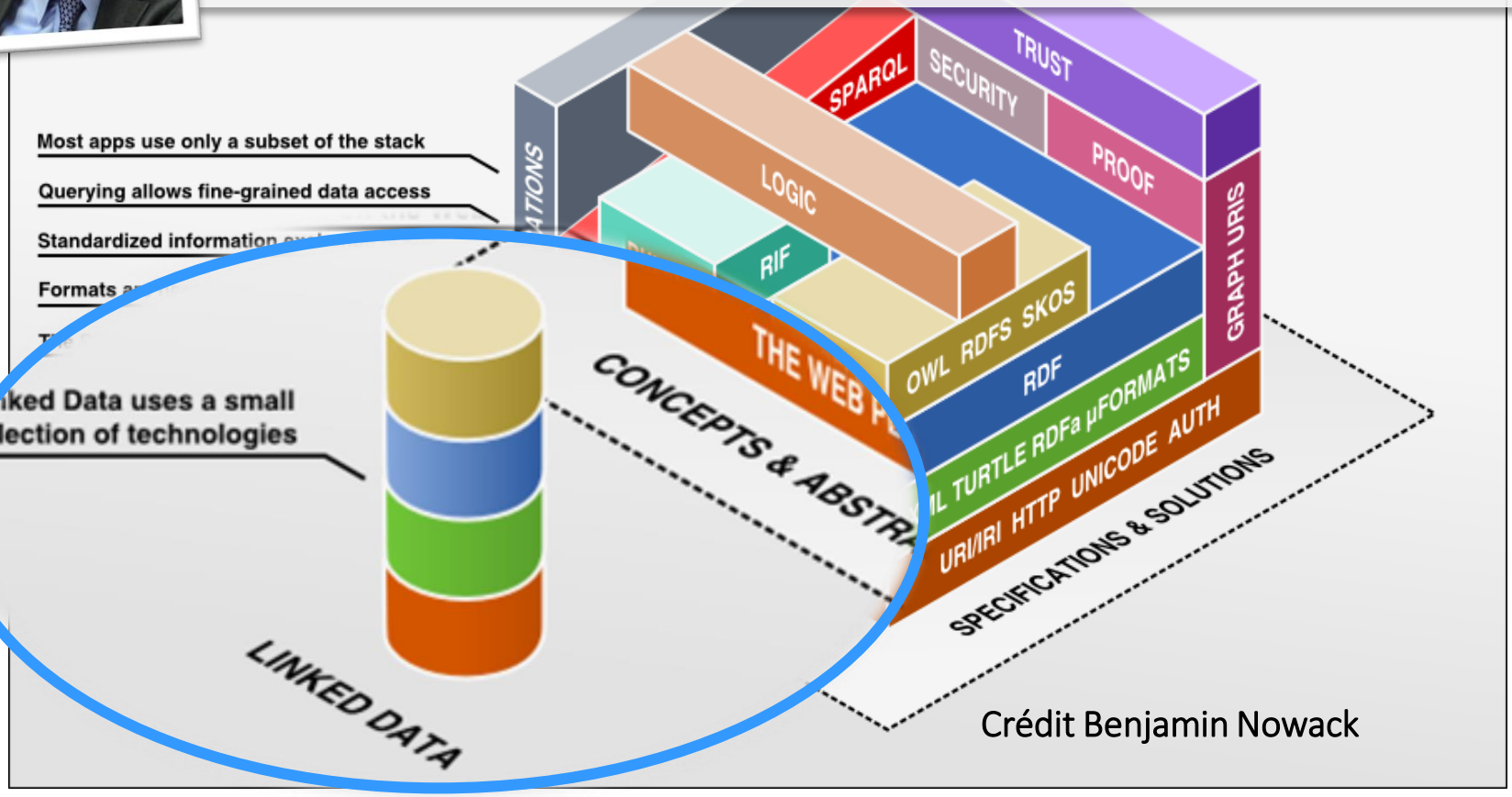


FAIR

From **Open Science overview in Europe,** https://www.openaire.eu

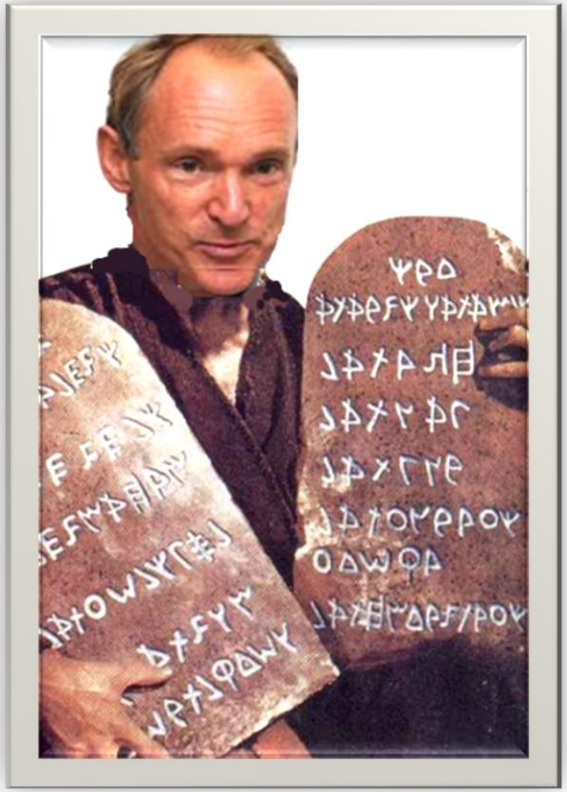
Linked Data

"The Semantic Web isn't just about putting data on the web. **It is about making links, so that a person or machine can explore the web of data**. With **linked data**, when you have some of it, you can find other, related, data."
Tim Berners-Lee - 2006 http://www.w3.org/DesignIssues/LinkedData.html



Crédit Benjamin Nowack

Linked Data Principles



Tim Berners-Lee
http://www.w3.org/DesignIssues/LinkedData.html

a set of best practices for publishing and interlinking structured data on the Web

Basic idea: to apply the general architecture of the World Wide Web to the task of sharing structured data on global scale.

1. Use **URIs** as names for things.
2. Use **HTTP URIs**, so that people can look up those names.
3. When someone looks up a URI, **provide useful information, using the standards** (RDF, SPARQL…).
4. Include **links** to other URIs, so that they **can discover more things**.

From Open Data to Linked Open Data

- 2006: defines basic principles for publishing Linked Data
- 2010: added a 5 star rating system for Linked Open Data (LOD)

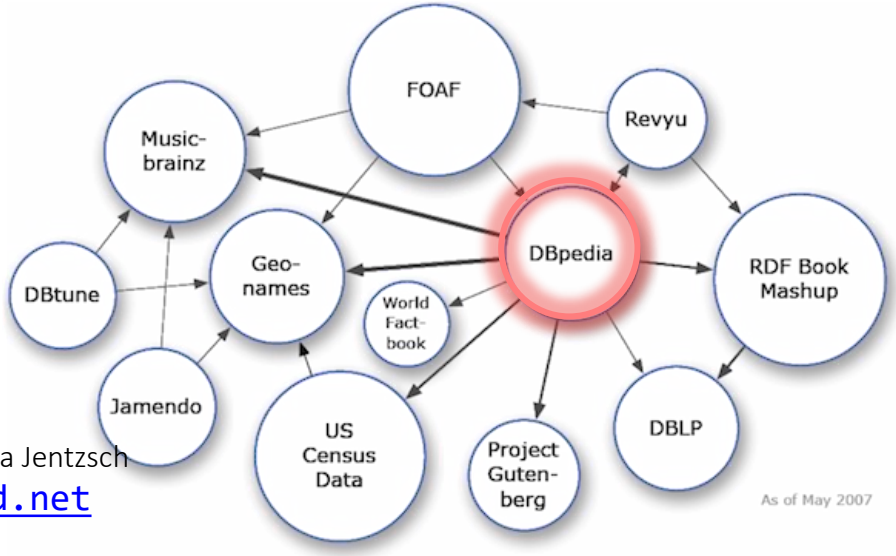
"in order to encourage people -- especially government data owners -- along the road to good linked data…"

Tim Berners-Lee
http://www.w3.org/DesignIssues/LinkedData.html

Linked Open Data (LOD)

The Open Data Movement aims at making data freely available to everyone. There are already various interesting open data sets available on the Web. Examples include Wikipedia, Wikibooks, Geonames, MusicBrainz, WordNet, the DBLP bibliography and many more which are published under Creative Commons or Talis licenses.

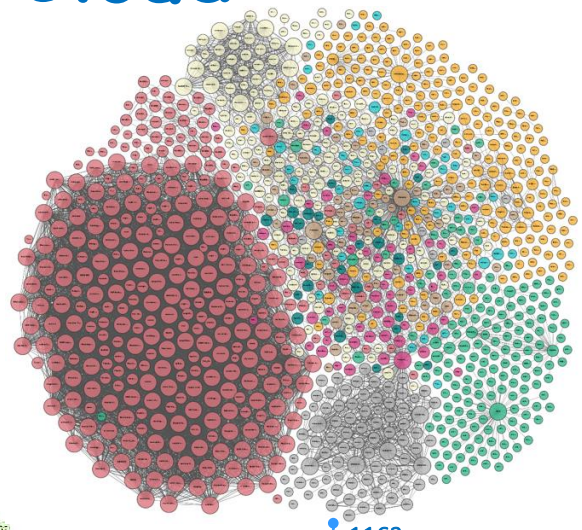
The goal of the W3C SWEO Linking Open Data community project is to extend the Web with a data commons by publishing various open data sets as RDF on the Web and by setting RDF links between data items from different data sources.

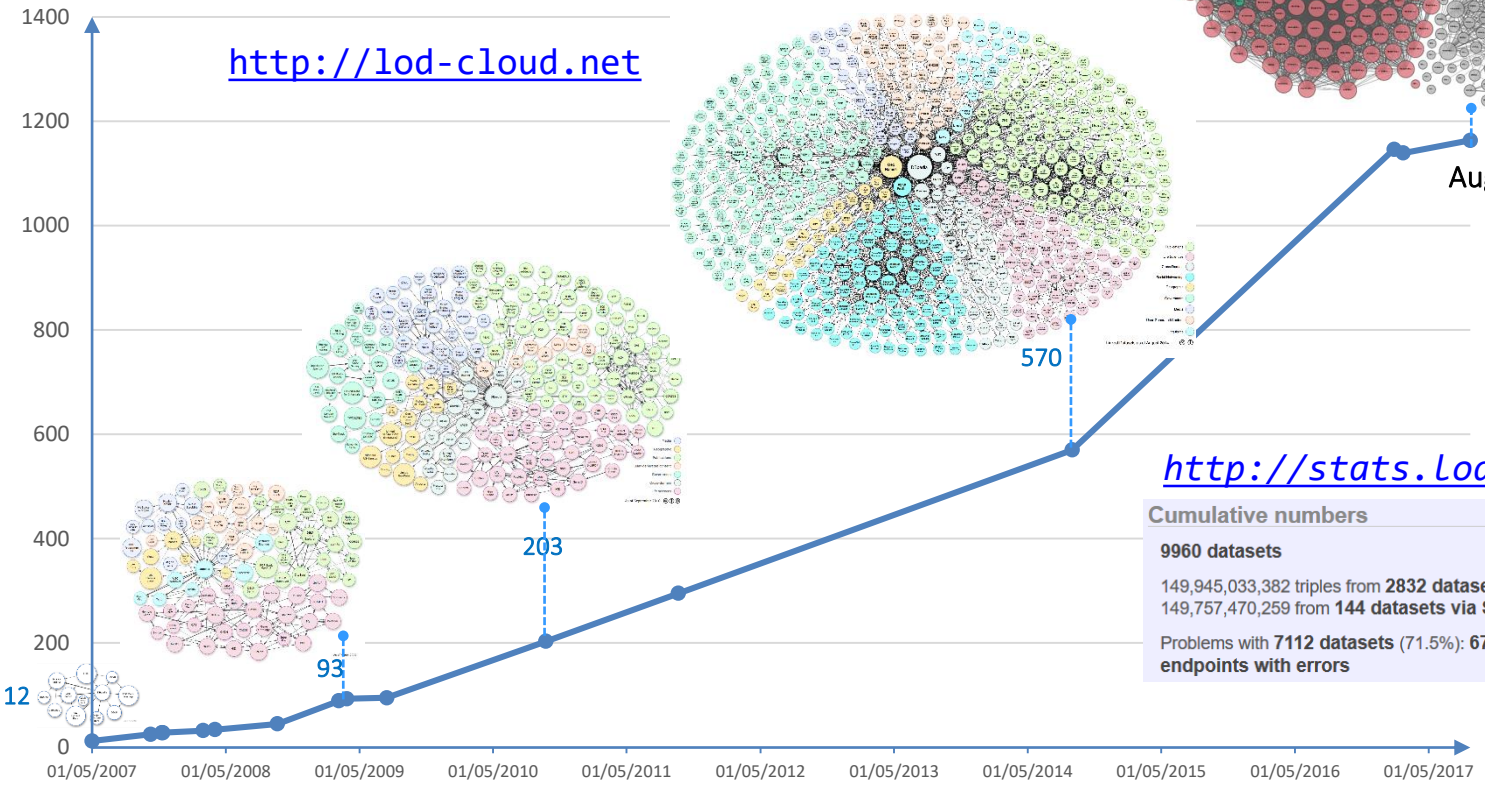


Richard Cyganiak and Anja Jentzsch
http://lod-cloud.net

Linked Open Data Cloud

- Linking open data project
  - goals:
    - Use RDF to “expose” open data sets
    - Create RDF links between these datasets
    - If possible, deploy SPARQL endpoints





LODSTATS

http://stats.lod2.eu/

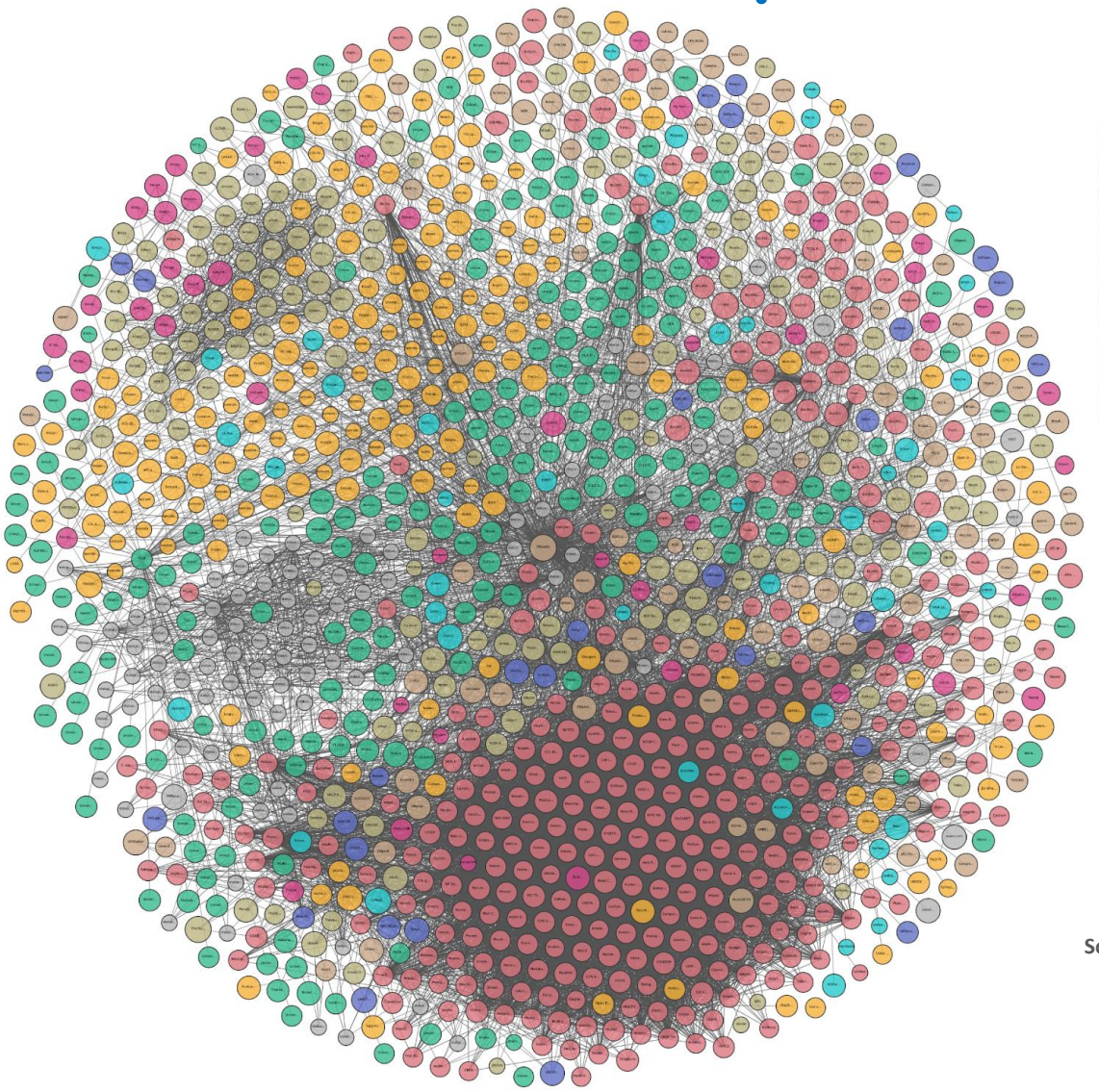
Cumulative numbers

9960 datasets

149,945,033,382 triples from **2832 datasets** (187,563,123 triples from **2704 dumps**, 149,757,470,259 from **144 datasets via SPARQL**)

Problems with **7112 datasets** (71.5%): **6712 dumps having errors**, **400 SPARQL endpoints with errors**

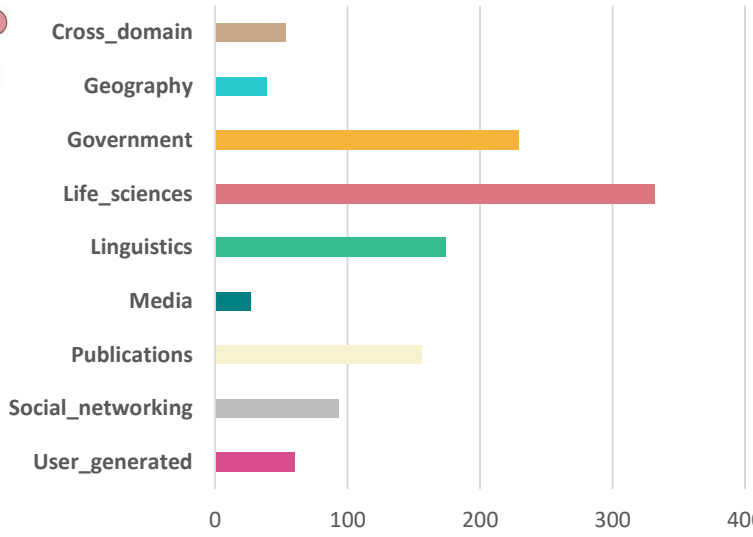
Linked Open Data Cloud



Legend
- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated

March 2019

1239 datasets

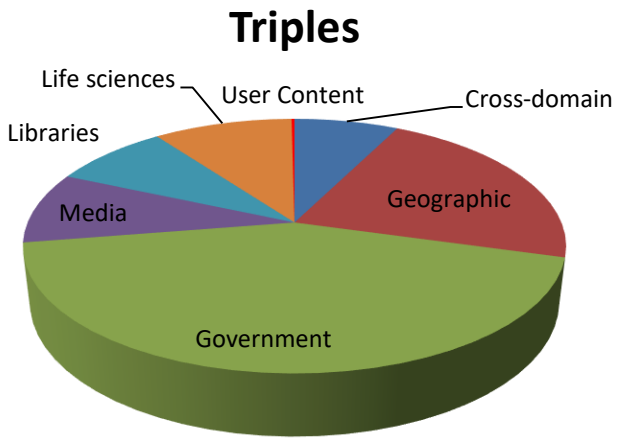


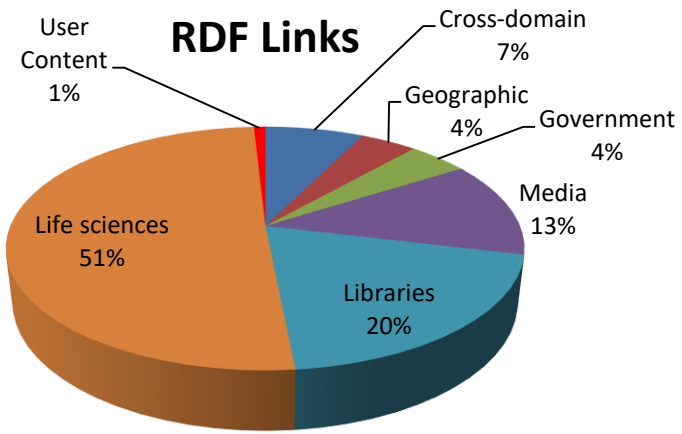
Linking Open Data cloud diagram 2019, by John P. McCrae, Andrejs Abele, Paul Buitelaar, Richard Cyganiak, Anja Jentzsch, Vladimir Andryushechkin, Jeremy Debattista
http://lod-cloud.net/

Linked Open Data Cloud

Domain	Data Sets	Triples	Percent	RDF Links	Percent
Cross-domain	20	1,999,085,950	7.42	29,105,638	7.36
Geographic	16	5,904,980,833	21.93	16,589,086	4.19
Government	25	11,613,525,437	43.12	17,658,869	4.46
Media	26	2,453,898,811	9.11	50,374,304	12.74
Libraries	67	2,237,435,732	8.31	77,951,898	19.71
Life sciences	42	2,664,119,184	9.89	200,417,873	50.67
User Content	7	57,463,756	0.21	3,402,228	0.86
(2011 September)	203	26,930,509,703		395,499,896	

http://lod-cloud.net/state





Linked Open Data Cloud

Domain	Data Sets	Triples
Cross-domain	20	1,999,085
Geographic	16	5,904,980
Government	25	11,613,525
Media	26	2,453,898
Libraries	67	2,237,435
Life sciences	42	2,664,119
User Content	7	57,463
	203	26,930,509

(2011, September)

Datasets by topical domain.

Topic	Datasets	%
Government	183	18.05%
Publications	96	9.47%
Life sciences	83	8.19%
User-generated content	48	4.73%
Cross-domain	41	4.04%
Media	22	2.17%
Geographic	21	2.07%
Social web	520	51.28%
Total	1014	

(2014, August)

State of the LOD Cloud 2014

Version 0.4, 08/30/2014

This document provides statistics about the structure and content of the crawlab
Linked Data sources implement the Linked Data best practices.

This document updates the findings of the original State of the LOD Cloud re
publishers themselves via the datahub.io Linked Data catalog. This report is base
the ISWC2014 paper Adoption of the Linked Data Best Practices in Different To
document links the statistics to the Mannheim Linked Data catalog and enables th

Contents

http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/

Linked Open Data Cloud

- https://lod-cloud.net/# → mise à jour 2023
- Lier des jeux de données :
  - https://data.europa.eu/fr

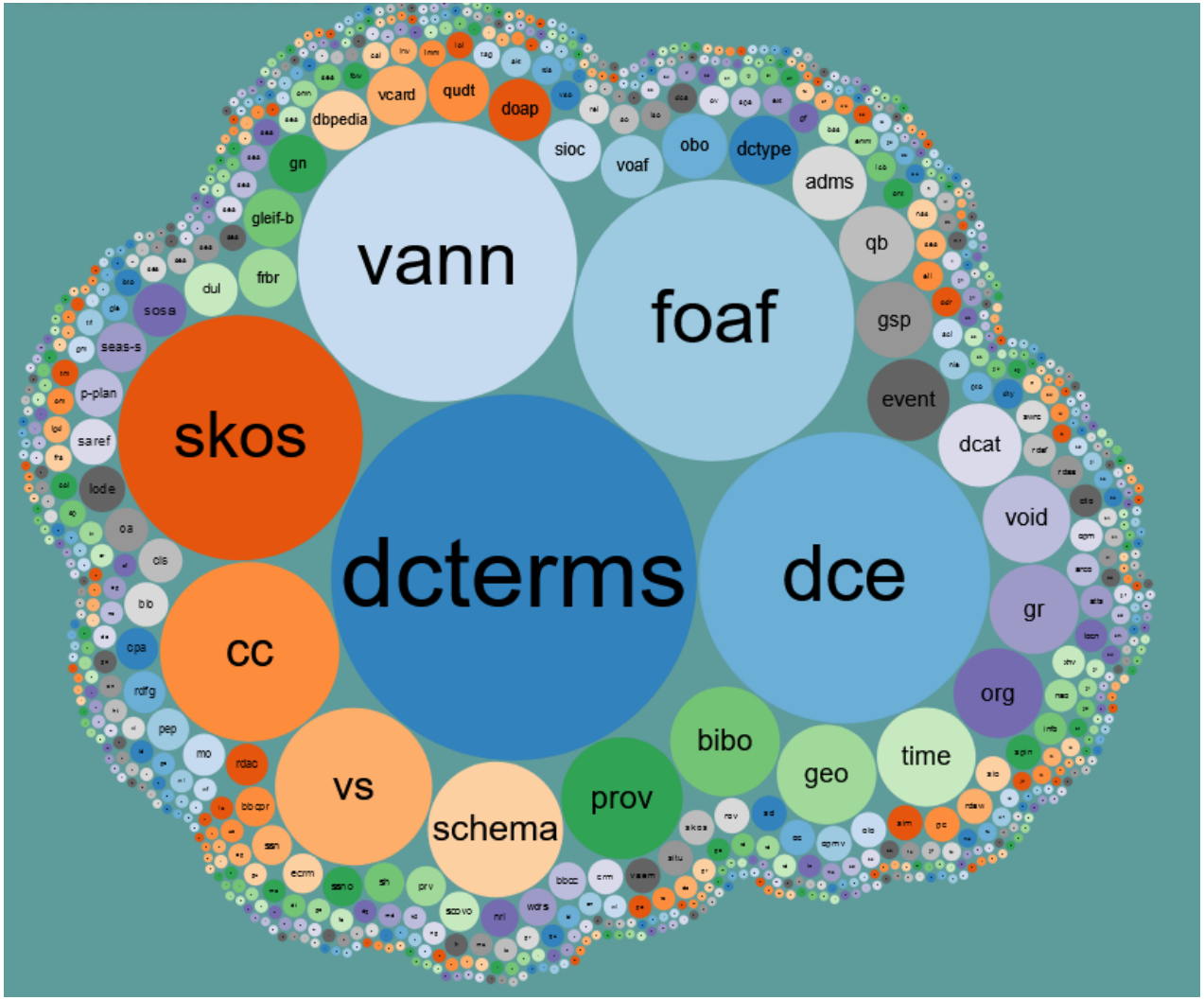
  - https://www.data.gouv.fr/fr/
  - https://www.wikidata.org/wiki/Wikidata:Main_Page

Make our datasets accessible and interoperable on the Web…

- What language can you speak ?
  - Taxonomy:
    - Practice and science of classification
    - Hierarchical categorization of controlled classes/terms
    - Nested classes under broader categories
  - Thesaurus
    - Networked collection of controlled vocabulary terms, grouped according to various types of relationship, e.g. similarity of meaning (synonyms, antonyms),
  - Ontology
    - Formal semantic description for the taxonomy terms, properties and interrelationships between categories in a domain of discourse, to facilitate conceptual search and natural language queries.
  - Folksonomy
    - Collaborative/social tagging, social classification…
    - Tag category schemes
    - No (not necessarily) hierarchical categorization
- Create, reuse and link vocabularies

Make our datasets accessible and interoperable on the Web…

https://lov.linkeddata.es/dataset/lov



A Faire : Quelques ressources à regarder

- The semantic web of data Tim Berners-Lee : https://www.youtube.com/watch?v=HeUrEh-nqtU
- Le web sémantique, web des données en 10 mn : https://www.youtube.com/watch?v=CHpZCYH4cOM
- Les vidéos de Manu Sporny:
  - IntroWeb-Sem : https://www.youtube.com/watch?v=OGg8A2zfWKg
  - Les donnéesliées: https://www.youtube.com/watch?v=4x_xzT5eF5Q
  - JSON-LD : https://www.youtube.com/watch?v=vioCbTo3C-4
- SPARQL : https://www.youtube.com/watch?v=FvGndkpa4K0
- Google knowledge graph : https://www.youtube.com/watch?v=mmQl6VGvX-c
- Global data space : http://linkeddatabook.com/editions/1.0/