



WEB SÉMANTIQUE ET ONTOLOGIES

WEB DES DONNÉES

DONNÉES LIÉES (LINKED DATA)

Philippe GENOUD – Danielle ZIEBELIN – Mohamed Wadhah Mabrouk

LIG-Soc-Sim-K

Prénom.Nom@univ-grenoble-alpes.fr



GI analyses des données, web des données, web sémantique D. Ziébelin et M.W. Mabrouk

- **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.
 - Le contrôle de connaissance se fera après le congés de toussaint.
- **Partie 2 (5 à 4 séances de 3h, projet soit 14h): 8 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
 - Linked Data

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 ...

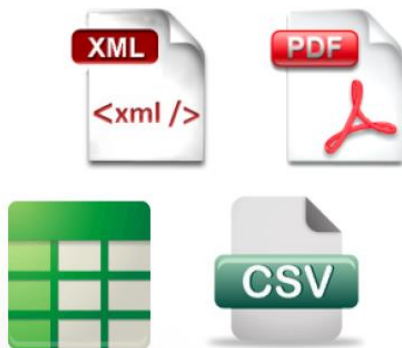
Make our datasets accessible and interoperable on the Web...

- 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>
```

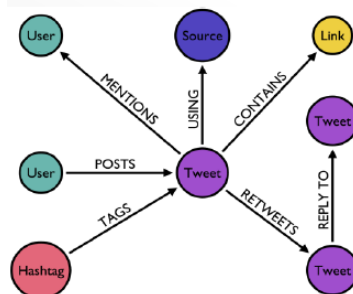
Documents



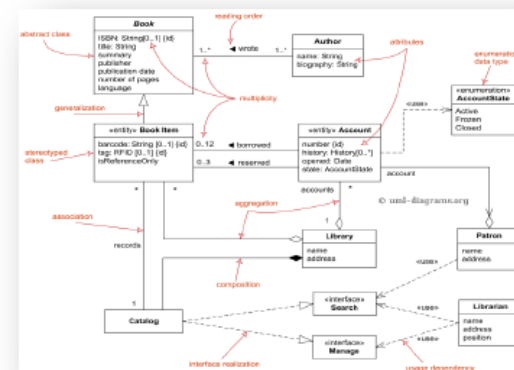
Tabular

ID	NAME	

Graph



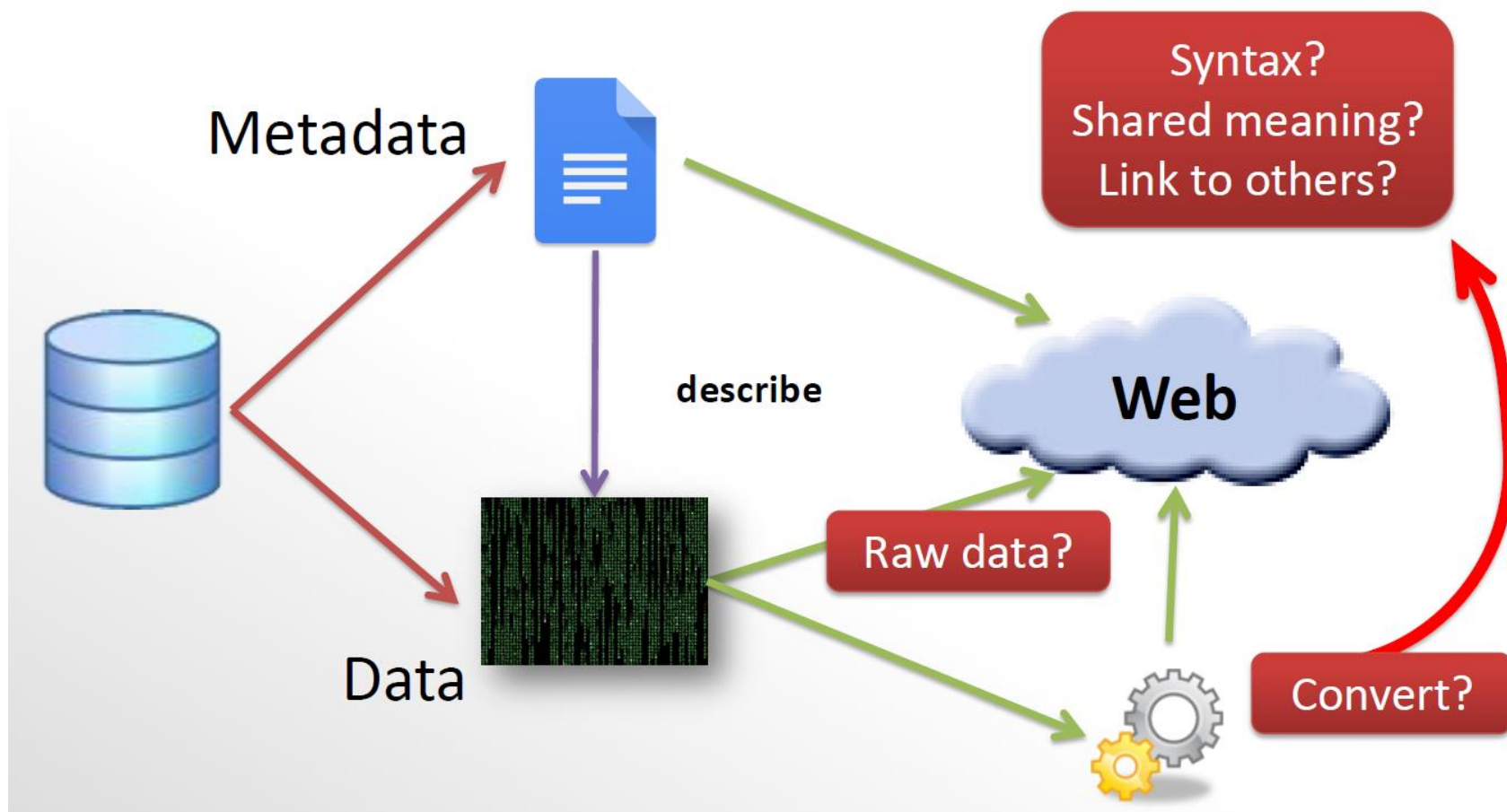
Object-Oriented



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 :

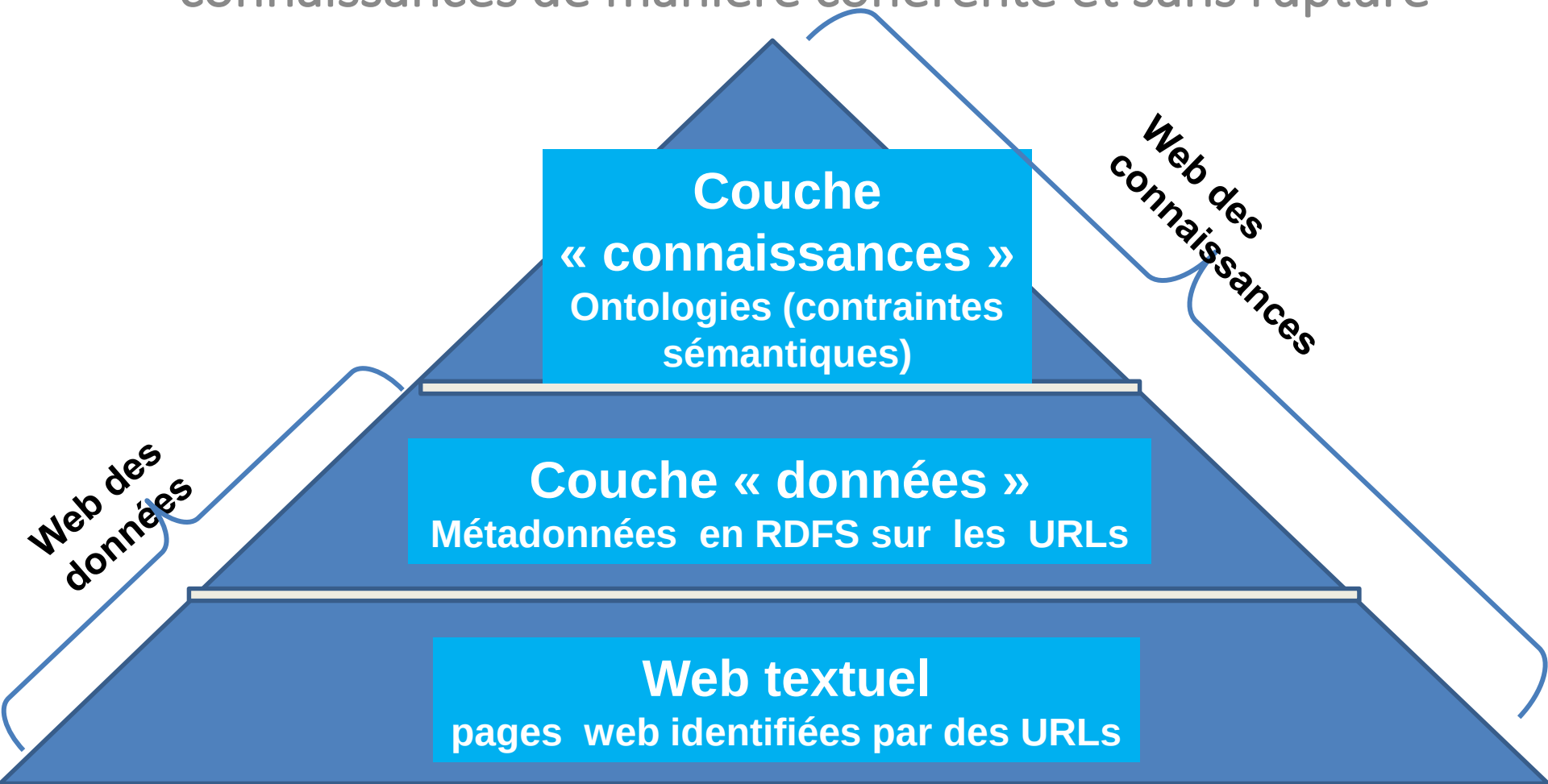
“ The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The first steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the 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

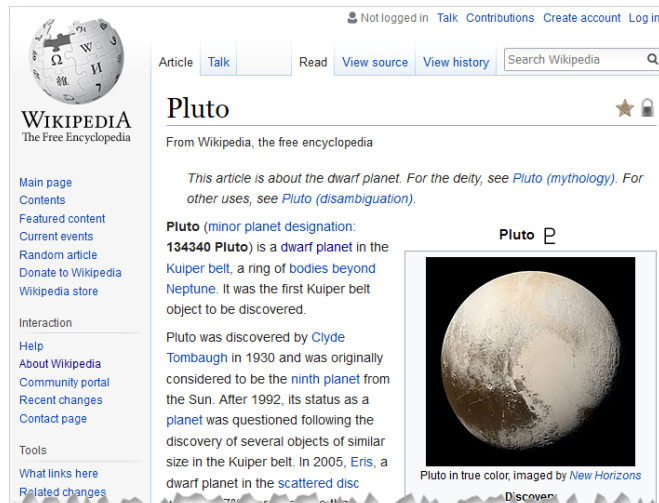
Le web est en train d'évoluer d'un web textuel à un web de connaissances de manière cohérente et sans rupture



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)

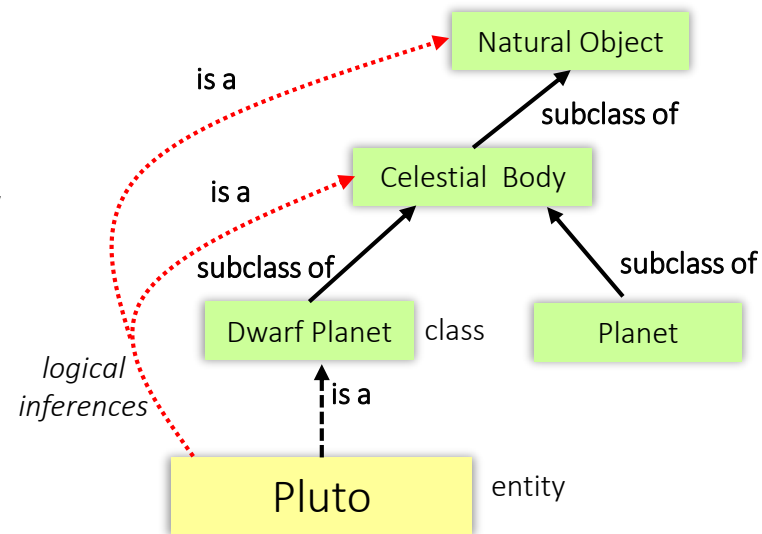


- The Semantic Web (Web of Data) is for machines

- 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

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Découvrir Inria, son Actualité, ses activités de Recherche & d'Innovation

Centre de recherche Inria Sophia Antipolis - Méditerranée

Présentation | Actualités | Recherche | Innovation

Accueil > Centre > Sophia > Agenda > MOOC sur le web sémantique

Enseignement

MOOC web sémantique et web de données

Un nouveau cours réalisé par le Mooc Lab sera bientôt disponible sur la plateforme France Université Numérique : il s'agit du MOOC Inria-uTOP « Web Sémantique et Web de données » de Fabien Gandon, Olivier Corby et Catherine Faron-Zucker, de l'équipe Wimmics (Inria, I3S, CNRS, UNS).

Plus de 2700 personnes se sont d'ores et déjà inscrites à ce cours qui débutera le 2 mars prochain!

Date : 2/03/2015 au 27/04/2015

Ce MOOC vous propose de vous former aux standards du Web de données et du Web sémantique. Il vous présentera de nombreux langages permettant d'agir sur les données du Web, tels que SPARQL, RDF, OWL, et bien d'autres.

Il s'adresse plus particulièrement aux ingénieurs en informatique et aux étudiants, notamment dans le domaine des systèmes d'information, et peut être suivi en complément d'une formation classique aux technologies de base du Web.

OPEN Channels Courses Podcast News

Knowledge Graphs - Foundations and Applications **Self-paced course**
Offered by Prof. Dr. Harald Sack

Learnings Discussions Progress Certificates Course Details Announcements Recap

Despite the fact that it affects our lives on a daily basis, most of us are unfamiliar with the concept of a knowledge graph. When we ask Alexa about tomorrow's weather or use Google to look up the latest news on climate change, knowledge graphs serve as the foundation of today's cutting-edge information systems. In addition, knowledge graphs have the potential to elucidate, assess, and substantiate information produced by Deep Learning models, such as Chat-GPT and other large language models. Knowledge graphs have a wide range of applications, including improving search results, answering questions, providing recommendations, and developing explainable AI systems. In essence, the purpose of this course is to provide a comprehensive overview of knowledge graphs, their underlying technologies, and their significance in today's digital world.

Self-paced since November 21, 2023
Language: English
English
Big Data and AI, Expert

Course information

A **knowledge graph** is a structured representation of knowledge that is used to provide a comprehensive and interconnected view of a specific domain. In this course we will discuss the **key features** and characteristics of knowledge graphs. You will learn all what is necessary to design, implement, and apply knowledge graphs. The focus of this course will be on basic semantic technologies including the underlying principles of knowledge representation and symbolic AI. This includes information encoding via RDF triples, knowledge representation via ontologies with OWL, efficiently querying knowledge graphs via SPARQL, latent representation of knowledge in vector spaces, as well as knowledge graph applications in innovative information systems, as e.g., semantic and exploratory search. Furthermore, the role of knowledge graphs in artificial intelligence and machine learning will be discussed, as well as their potential to improve **explainability** and **trustworthiness** of "black box" deep learning models such as Chat-GPT.

Attention: This course is currently in self-study mode, in which you do not have access to graded assignments/exams. Therefore, we can only issue you a certificate of participation.

What you'll learn

- Basic understanding of knowledge graphs
- Basic understanding of ontologies
- Basic understanding of Semantic Web Technologies
- Basic understanding of ontology design and knowledge graph construction
- Basic understanding of knowledge graph embeddings

Who this course is for

- students
- practitioners of computer science, digital humanities, and information sciences
- enthusiasts with an AI related background and an interest in symbolic knowledge representation

Featured content

0.0 Lecture Overview

LEARNERS

CURRENT Today 7,114

COURSE END Nov 21, 2023 4,694

COURSE START Oct 11, 2023 2,552

Training - Epimorphics

https://www.epimorphics.com/training/

Cours WS 1 Introduction au Web Sémantique

TRAINING MODULES

Epimorphics offers a range of core training modules. An Epimorphics training is a mixture of presentation and practical exercises.

BENEFITS OF LINKED DATA

A management-level introduction to understanding the benefits of linked data. Covers motivations, principles, practices and technologies, and the key features of linked data solutions. This course will help organisations building data strategies and developing plans to integrate and share data.

Half a day

INTRODUCTION TO LINKED DATA

A practically oriented introduction to the concepts, standards, technologies and tools needed to understand and implement linked data solutions. Understand key benefits of linked data, gain firsthand experience of exactly what it is and how it works, evaluate its relevance to your situation and get started with the technology. Can be taken as one or two days, the second day adds technical detail and provides more advanced practice.

1-2 days

SPARQL (LINKED DATA QUERY LANGUAGE)

The one day version of this course introduces core elements of using SPARQL (Linked Data Query Language) to query linked data and can be flexibly extended to cover more advanced topics including named graphs, query federation and SPARQL Update.

1-2 days

LINKED DATA MODELLING AND ONTOLOGY DEVELOPMENT

After a general introduction to the principles of linked data modelling, attendees are guided through the process of developing a model for a familiar domain. Using this example attendees learn how to apply common modelling patterns, ontology components and standard tools that they can reuse in their own work.

1 - 2 days

BESPOKE LINKED DATA TRAINING & SUPPORT

We can adapt our core training offerings to meet your specific needs and put together courses on other topics, for example 'publishing statistical data as linked open data', 'The Linked Data API and Elda'. We also offer follow up support packages to help your staff apply what they have learned.

1-3 days

REGISTRY TRAINING

This course provides a review of the features and facilities provided by the Reference Data Manager, together with training on its operation. It explains how to load new items, how to update existing registers and code lists, and how to query and retrieve data from the platform. It also covers how to add new users to the system and manage administrative functions.

1-3 days

logilab

Société | Compétences | Services | Produits | Formations | Logiciels libres

Introduction au Web Sémantique

SERVICE-INTRO • 1 jour

Objet

Le Web sémantique, aussi nommé Web 3.0 ou Web des données, est la suite logique du Web des documents aujourd'hui connu de tous. Il est fondé sur une vision du Web dans laquelle les machines échangent des informations exprimées dans des langages normalisés et peuvent agréger, trier et traiter des données issues de sources diverses, faisant du Web une sorte de base de données à l'échelle planétaire.

La formation proposée ici se déroule sur une journée et a pour but de permettre l'acquisition des bases techniques du Web sémantique et des repères que sont les standards de modélisation, de stockage, de publication et d'interrogation des données.

Public visé

Ce cours s'adresse aux décideurs désireux d'appréhender les concepts du Web sémantique et aux développeurs souhaitant acquérir des repères avant de se lancer dans une phase d'expérimentation ou de réalisation.

Pré-requis

Connaissance de l'architecture générale du Web et de la gestion des données au sein des systèmes d'information.

Programme

Introduction au Web Sémantique

Qu'est ce que le Web sémantique ?
définition, historique et origines, exemples

Des formations en entreprise

Pourquoi un cours web sémantique ?

Des technologies adoptées

DATA.GOV DATA TOPICS IMPACT APPLICATIONS DEVELOPERS CONTACT

DATA CATALOG / Datasets Organizations ?

Federal datasets are subject to the U.S. Federal Government **Data Policy**. Non-federal participants (e.g., universities, organizations, and tribal, state, and local governments) maintain their own data policies. Data policies influence the usefulness of the data. **Learn more** about how to search for data and use this catalog.

geospatial

Datasets ordered by Relevance

You are searching in the list of datasets. Show results in [entire Data.gov site](#).

Filter by location Clear

Enter location...

90,628 datasets found for "geospatial"

Enterprise Geospatial Information Services
Department of Homeland Security — The Enterprise Geospatial Information Services is a web-based application and service that provides geospatial data and information to the Department of Homeland Security.

HRSA Geospatial Data Warehouse
U.S. Department of Health & Human Services Geospatial Data Warehouse (HGDW) provides health resources...

Distinct Agency Names in Geospatial Metadata
State of Oklahoma — Provides a listing of the unique agency names that contain the name from the geospatial metadata catalog on geo.data.gov.

Distinct Agency Names in Geospatial Metadata
State of Hawaii — Provides a listing of the unique agency names that contain the name from the geospatial metadata catalog on geo.data.gov.

Distinct Agency Names in Geospatial Metadata
State of Oregon — Provides a listing of the unique agency names and the number of datasets that contain the name from the geospatial metadata catalog on geo.data.gov. The list was...

Organization Types Clear All

- Federal Government (42698)
- State (34097)
- University (8229)
- State Government (5375)
- Other (102)

Show More Organization Types

Organizations Clear All

- NSGIC GIS Inventory... (34097)
- National Oceanic and Atmospheric Administration (33170)
- Earth Data Analysis Center (5535)

Geospatial display of current weather radar images (RIDGE Weather Radar)
National Weather Service, Department of Commerce — Provides GIS overlays for current weather radar results
kml/kmz

Geospatial display of current weather radar images (RIDGE Weather Radar)
State of Oregon — Provides GIS overlays for current weather radar results
CSV application/rdf+xml JSON XML

Geospatial display of current weather radar images (RIDGE Weather Radar)
State of Hawaii — Provides GIS overlays for current weather radar results
CSV application/rdf+xml JSON XML

Geospatial display of current weather radar images (RIDGE Weather Radar)
State of Oklahoma — Provides GIS overlays for current weather radar results
CSV application/rdf+xml JSON XML

application/rdf+xml

Pourquoi un cours de web sémantique : FAIR



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

Pourquoi un cours de web sémantique : Open Data

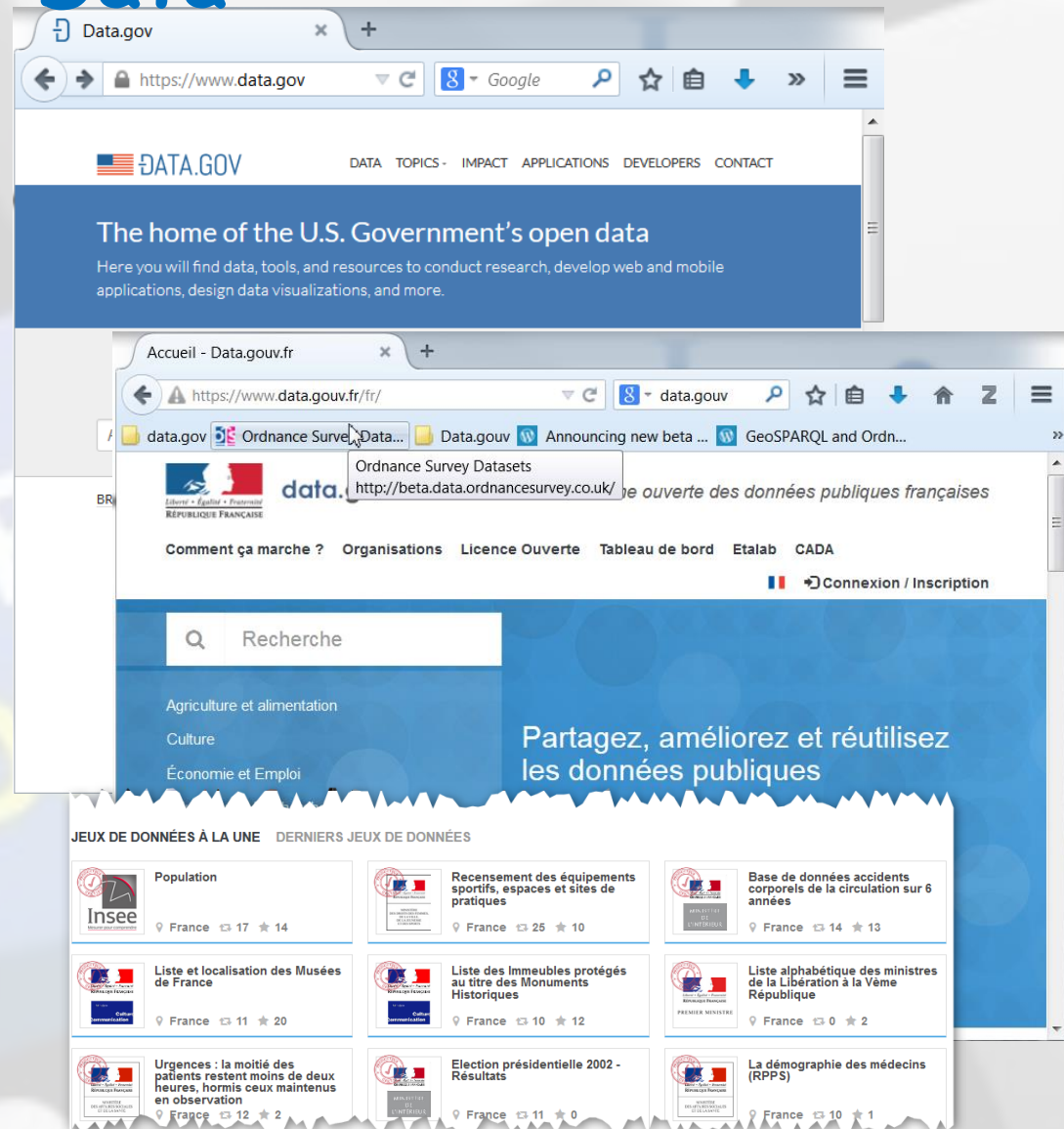
- Open Data Movement

- “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

...



What is This Thing Called Linked Data?
M. Atencia, J. David, Ph. Genoud

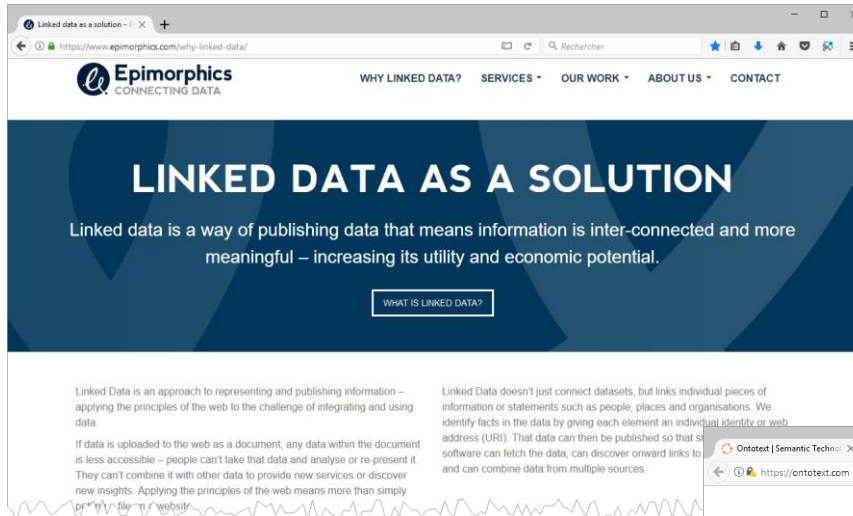
Pourquoi un cours web sémantique ?

Des technologies adoptées : 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/>
- Site public gouvernementaux : <https://ckan.org/government> pour Etats-Unis (data.gov), la France <https://www.data.gouv.fr/>, Royaume-Uni (data.gov.uk)

Pourquoi un cours web sémantique ?

- des entreprises spécialisées



Epimorphics
CONNECTING DATA

WHY LINKED DATA? SERVICES OUR WORK ABOUT US CONTACT

LINKED DATA AS A SOLUTION

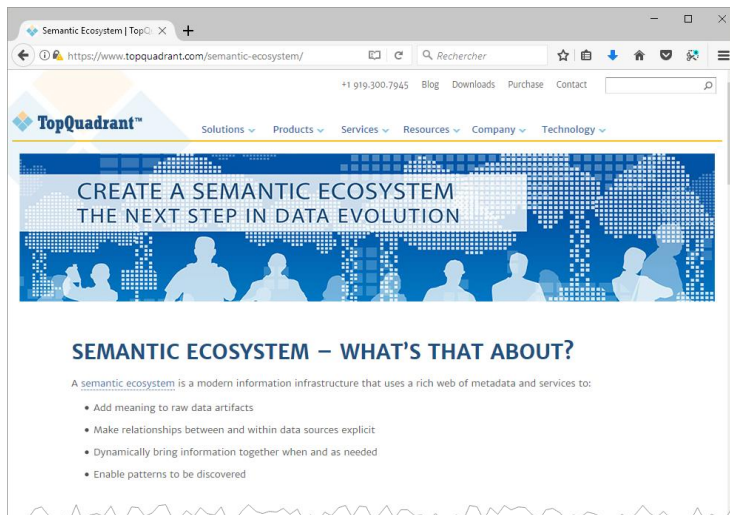
Linked data is a way of publishing data that means information is inter-connected and more meaningful – increasing its utility and economic potential.

WHAT IS LINKED DATA?

Linked Data is an approach to representing and publishing information – applying the principles of the web to the challenge of integrating and using data.

If data is uploaded to the web as a document, any data within the document is less accessible – people can't take that data and analyse or re-present it. They can't combine it with other data to provide new services or discover new insights. Applying the principles of the web means more than simply

Linked Data doesn't just connect datasets, but links individual pieces of information or statements such as people, places and organisations. We identify facts in the data by giving each element an individual identity or web address (URI). That data can then be published so that software can fetch the data, can discover onward links to and can combine data from multiple sources.



Semantic Ecosystem | TopQuadrant

+1 919.300.7945 Blog Downloads Purchase Contact

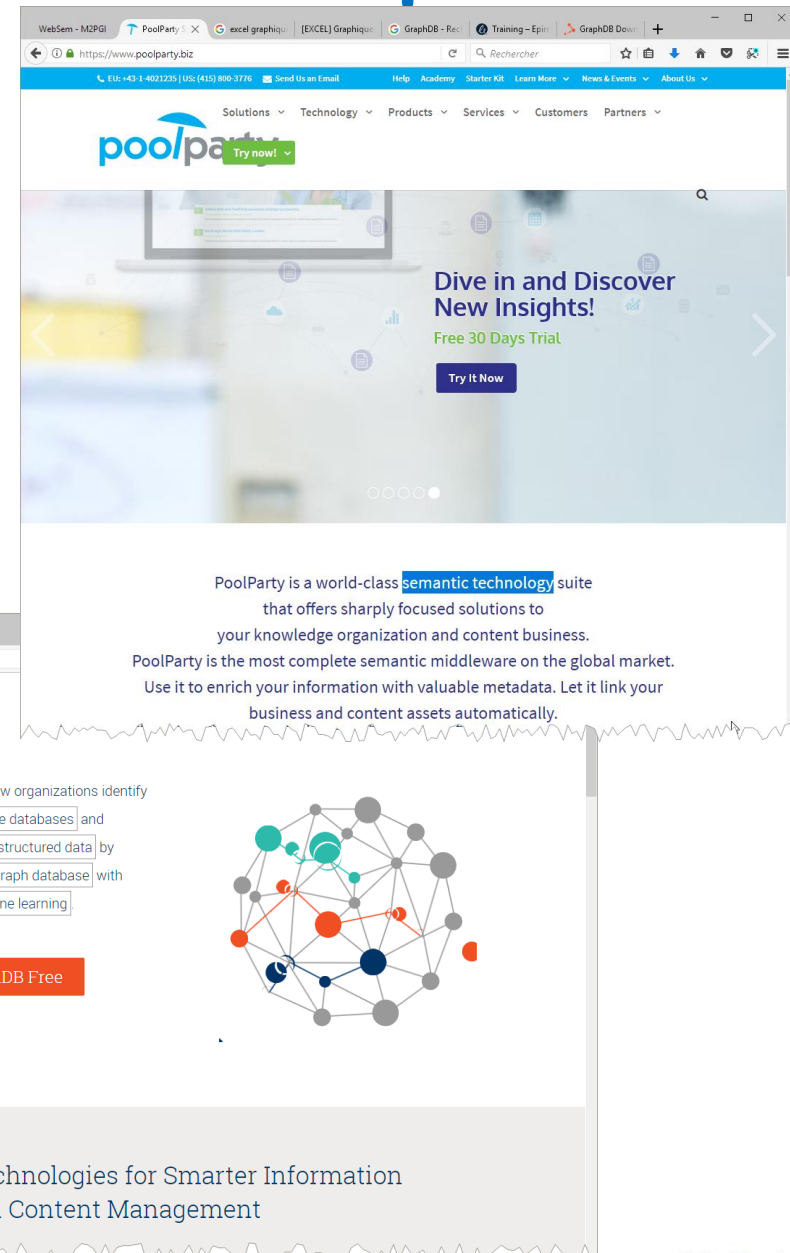
Solutions Products Services Resources Company Technology

CREATE A SEMANTIC ECOSYSTEM THE NEXT STEP IN DATA EVOLUTION

SEMANTIC ECOSYSTEM – WHAT'S THAT ABOUT?

A semantic ecosystem is a modern information infrastructure that uses a rich web of metadata and services to:

- Add meaning to raw data artifacts
- Make relationships between and within data sources explicit
- Dynamically bring information together when and as needed
- Enable patterns to be discovered



WebSem - M2PGI PoolParty excel graphique [EXCEL] Graphique GraphDB - Re Training - Epimorphics GraphDB Down

https://www.poolparty.biz

Solutions Technology Products Services Customers Partners

poolparty Try now!

Dive in and Discover New Insights!
Free 30 Days Trial
Try It Now

PoolParty is a world-class semantic technology suite that offers sharply focused solutions to your knowledge organization and content business.

PoolParty is the most complete semantic middleware on the global market. Use it to enrich your information with valuable metadata. Let it link your business and content assets automatically.

ontotext

Ontotext transforms how organizations identify meaning across diverse databases and massive amounts of unstructured data by combining a semantic graph database with text mining and machine learning.

Download GraphDB Free



Semantic Technologies for Smarter Information Retrieval and Content Management

Pourquoi un cours web sémantique ?

Des offres
d'emploi...

Jobs / Job

VIDEO

Viseo

France, Grenoble

Apply for
20 days left

ou de
stages...

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

Contexte général

Dans le cadre du projet de recherche FEDER TCVPR(1) (inventaire du patrimoine de villégiature dans les Pyrénées), nous avons pour 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 experts) à accéder aux informations relatives à ce patrimoine culturel. D'autre part, nous envisageons la publication automatique de données d'inventaire réalisées sur Wikipedia.

Objectifs

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

de Wikipedia.

De façon plus précise, les étapes à réaliser sont :
1. État de l'art sur les liens entre les données (Linked Open Data)
2. Compréhension des mécanismes de publication sur Wikidata, sur les liens entre les données (Linked Open Data)
3. Appropriation et développement d'outils de publication sur Wikidata, sur les liens entre les données (Linked Open Data)
4. Développement d'applications mobiles pour la gestion des liens entre les 3 univers (Wikipedia, Wikidata, WikiCommons)

(Linked Open Data)

Compétences requises :
- Maîtrise de la programmation (Java, Python, PHP, etc.)
- Maîtrise de la gestion des bases de données (MySQL, PostgreSQL, etc.)
- Maîtrise de la gestion des liens entre les données (Linked Open Data)
- Maîtrise de la gestion des liens entre les données (Linked Open Data)

Divers

Durée : 5 mois
Modalité : à temps plein
Lieu : LIUPPA – Equipe T2I

Comment candidater

Envoyer un CV ainsi que vos coordonnées (nom, prénom, adresse, téléphone, email) à :
marie-noelle.bessagnet@univ-pau.fr, christian.sallaberry@univ-pau.fr, philippe.rose@univ-pau.fr

19/09/2019

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 visualisation
- élaboration de Web services d'écologie
- participation à des formations de développement

Connaissances et compétences attendues

- langage Java maîtrisé et langages de programmation associés
- Web services REST appréciés
- langages du Web Sémantique (RDF, OWL, SPARQL...)

La personne recrutée sera amenée à travailler avec des acteurs externes du domaine. Un bon relationnel est donc un atout.



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 œuvre 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.

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

Leading provider of software and services in the areas of Machine Learning, Natural Language Processing, and PoolParty Software Platform is used in large enterprises and NGOs around the globe to extract meaning from unstructured data.

Intelligence and innovation and rewards strong growth opportunities. We offer great work-life balance and fun. If you are interested in this position, please send us your resume and cover letter to careers@viasema.com.
[Create alert](#)

Background More

Sort by: matches

Technologies 6d ago

Semantic Technologies 6d ago

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

Pourquoi un cours web sémantique ?

Des technologies adoptées



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.

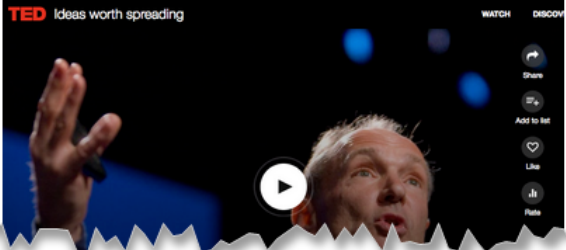
Governance

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](#)

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.



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 a framework to guide the use of Linked Data technologies is a key challenge for the Australian Government.

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

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

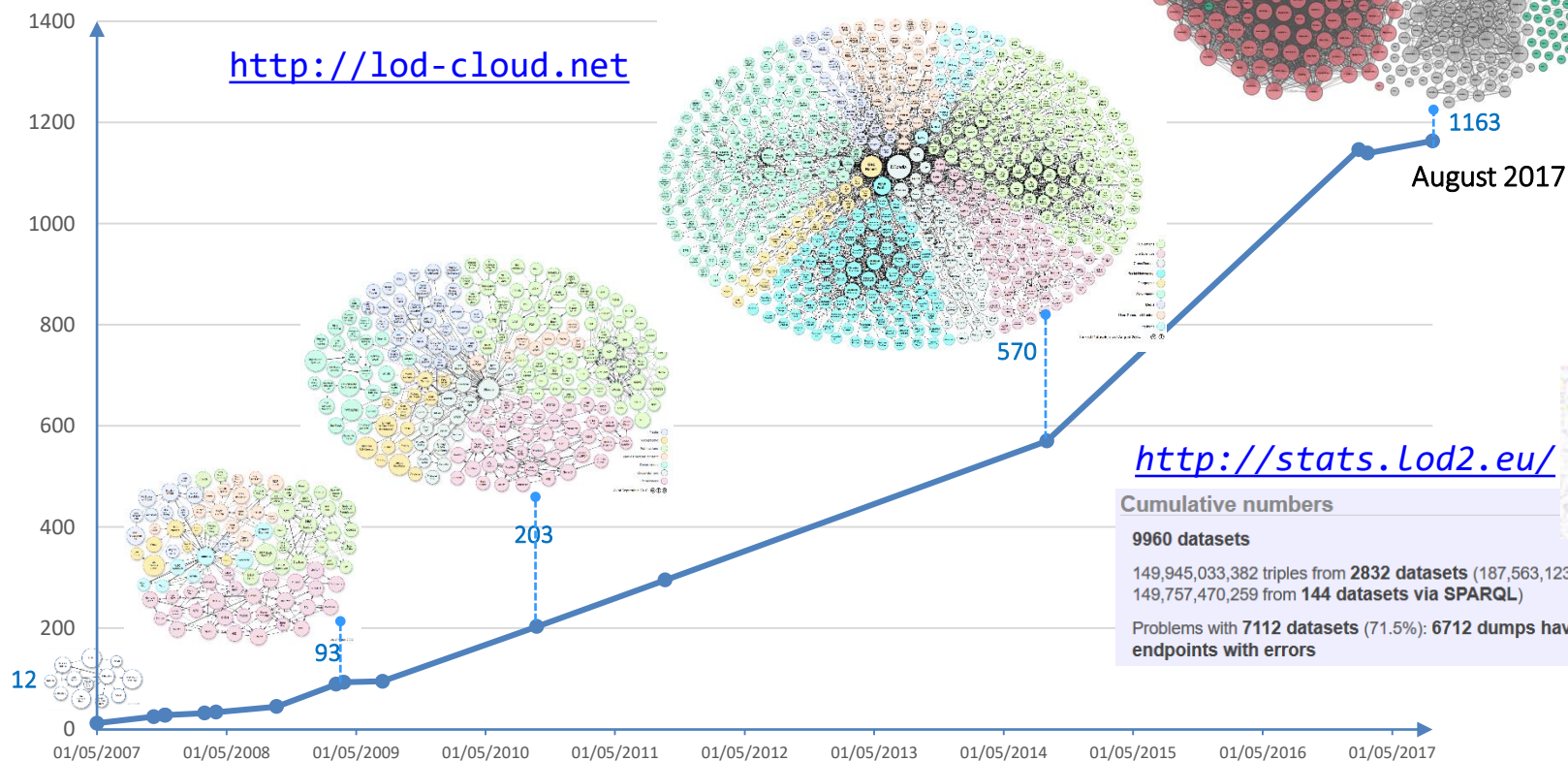
The screenshot shows the Insee website interface. At the top, there's a header with the Insee logo and navigation links: Mobile, Actualités, Agendas, Contactez-nous, Aide. Below the header is a search bar with the text 'Chercher sur le site'. A main navigation bar contains buttons for Accueil, Thèmes, Bases de données, Publications et services, Régions, Définitions et méthodes, and Accès. The left sidebar is titled 'Définitions et méthodes' and lists: Nomenclatures, Définitions, Sources et méthodes, Grilles d'analyse, Code officiel géographique, zonages d'études, Outils statistiques, and Données RDF et espace XML (which is highlighted). The main content area is titled 'Accueil > Définitions et méthodes > Données RDF et espace XML'. It features a section 'Données RDF et espace XML' with a sub-section 'Données au format RDF'. The text explains that Insee publishes data in RDF format to facilitate automatic use by applications. It lists available data: Données géographiques (COG), Codes et nomenclatures (NAF, PCS), and Données de population. It also mentions that other data will be published in the coming months and that all RDF data can be queried via SPARQL at the URL <http://rdf.insee.fr/sparql>. A link to 'Accéder à l'espace RDF de l'Insee' is provided at the bottom.

The screenshot shows the 'Eurostat - Linked Data' page. At the top, there's a header with the LATC logo and navigation links: Overview, Usage, Dataspace, Support. The main heading is 'Eurostat - Linked Data'. Below it, a paragraph states: 'This is a [Linked Data](#) version of the [Eurostat](#) data with the goal to provide 5 star Linked Open Data on the European level, in a contextually rich and up-to-date manner, useful for ETL-style business analysis or data warehousing purposes with benefits including but not limited to:'. A bulleted list follows: 'It allows for a straight-forward comparison of statistical indicators across EU countries.', 'Through providing context for statistics it facilitates the interpretation process.', and 'Enables you to re-use observations in a fine-grained way.' Below this is a section titled 'Overview'. It states: 'The following data is available under the original [Eurostat](#) license:'. A bulleted list provides links: <http://eurostat.linked-statistics.org/data/> contains the observations, <http://eurostat.linked-statistics.org/dsd/> contains the Data Structure Definition (DSD), and <http://eurostat.linked-statistics.org/dic/> contains the dictionaries. At the bottom, it says: 'Via our [SPARQL](#) endpoint you can query the entire metadata including DSDs and dictionaries.'

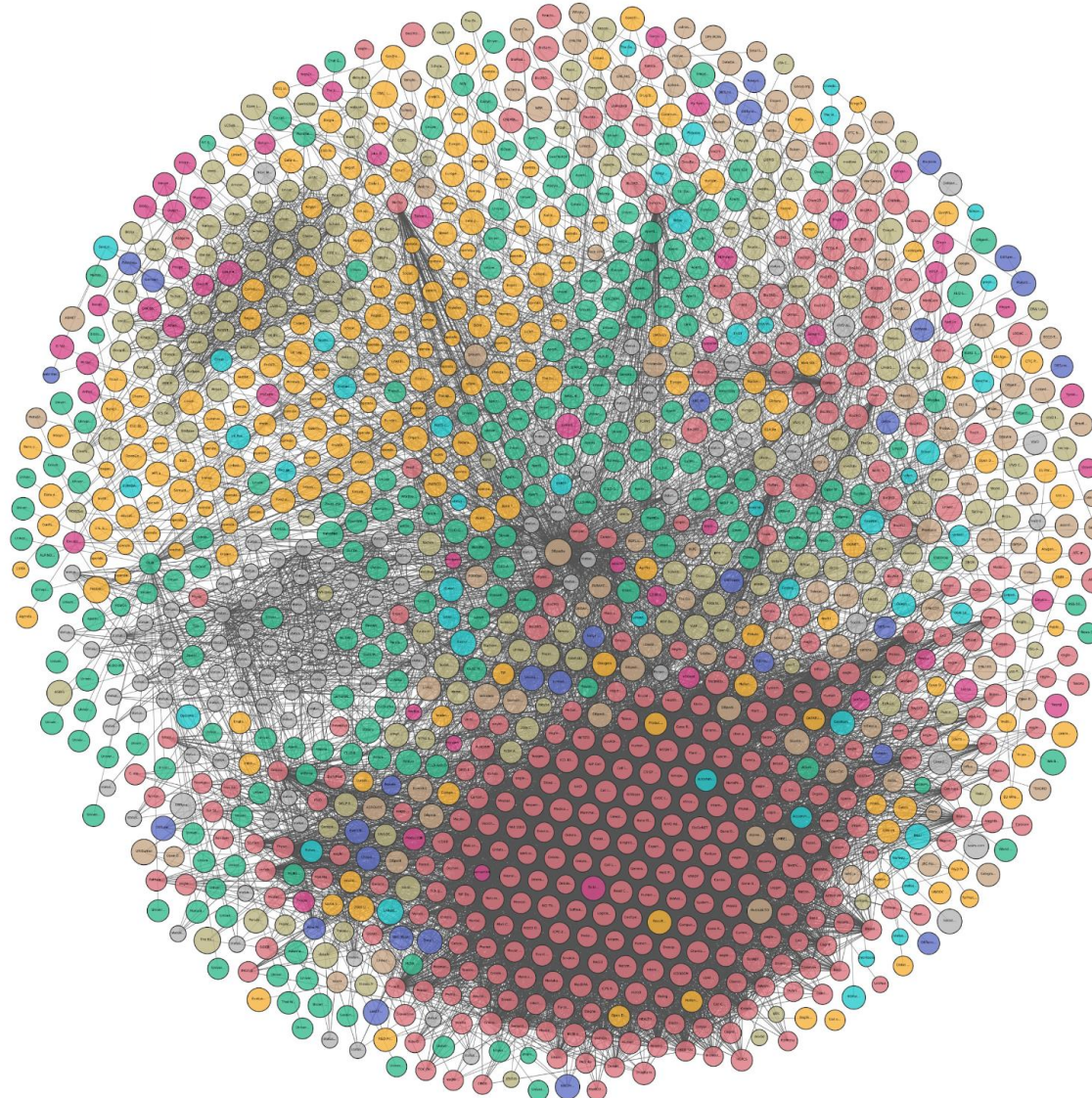
• • •

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



Linked Open Data Cloud

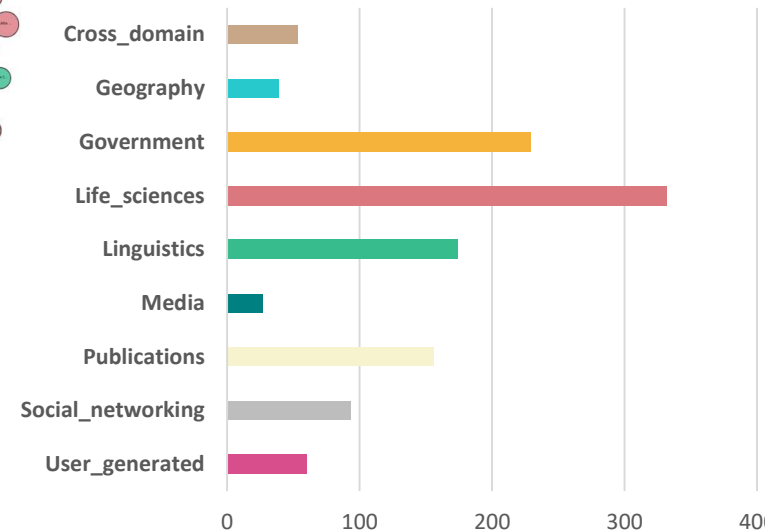


Legend



March 2019

1239 datasets



The Linked Open Data Cloud from lod-cloud.net



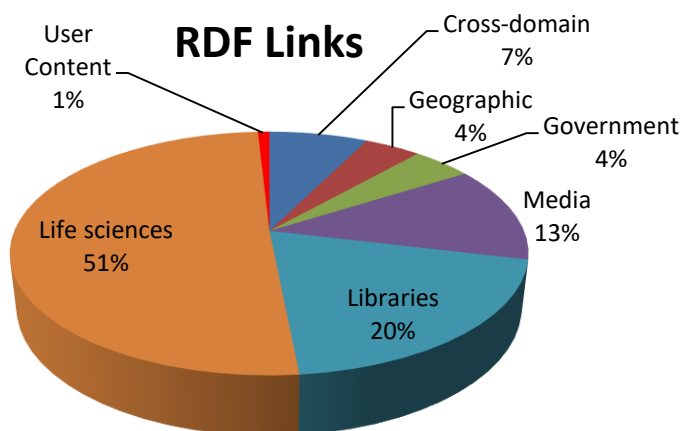
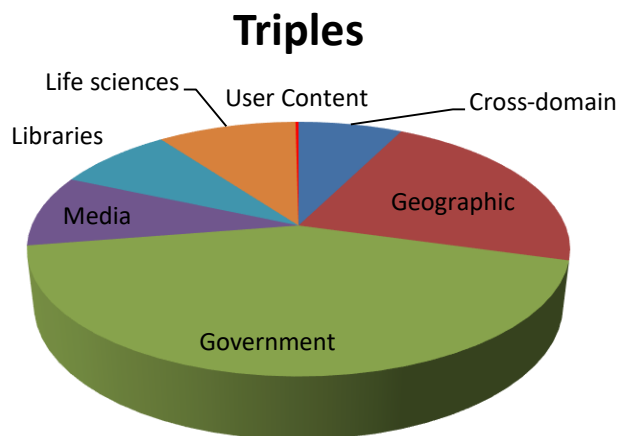
Linking Open Data cloud diagram 2019, by John P. McCrae, Andrejs Abele, Paul Buitelaar, Richard Cyganiak, Anja Jentzsch, Vladimir Andryushechkin, Jeremy Debattista

<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)

State of the LOD Cloud 2014

Version 0.4, 08/30/2014

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

This document updates the findings of the original [State of the LOD Cloud](#) report. Publishers themselves via the [datahub.io](#) Linked Data catalog. This report is based on the ISWC2014 paper [Adoption of the Linked Data Best Practices in Different Topologies](#). This document links the statistics to the [Mannheim Linked Data catalog](#) and enables the

Contents

- [1. The Linked Data Crawl](#)
- [2. Linked Data by Domain](#)
- [3. Crawlable LOD Cloud Diagram](#)
- [4. Best Practices](#)
- [4.1 Interlinking Best Practice](#)
- [4.2 Vocabulary Best Practices](#)
- [4.2.1 Usage of Proprietary Vocabularies](#)
- [4.2.2 Usage of Dereferencable Vocabularies](#)
- [4.3 Adoption of Metadata Best Practices](#)

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

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)

Linked Open Data Cloud

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

173 **36** **1 437 988**
Catalogues Pays Jeux de données

Trending datasets ?

-  Liste consolidée des personnes, groupes et entités faisant l'objet de sanctions financières de l'UE
-  Numéro d'identification du contribuable (NIF)
-  Base de données sur les ingrédients cosmétiques (Cosing) — Ingrédients et parfums
-  Le Bulletin pétrolier de la Commission européenne
-  Appels d'offres quotidiens électroniques (TED) (sous-ensembles) — avis de marchés publics

Rechercher des données

Rechercher des données 

 Agriculture, pêche, forêts et alimentation

 Economie et finances

 Education, culture et sport

 Énergie et ressources naturelles

 Environnement et climat

 Gouvernement et secteur public

 Affaires internationales

 Justice, droit et sécurité publique

 Population et société

 Régions et développement local

 Santé

 Science et technologie

 Transport

> CATALOGUES DE DONNÉES > TOUS LES JEUX DE DONNÉES > EU INSTITUTIONS DATASETS

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

Les technologies vues dans le cours

- Vocabulaires et ontologies
- Evolution des technologies du web
- Technologies des données ouvertes et liées

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)

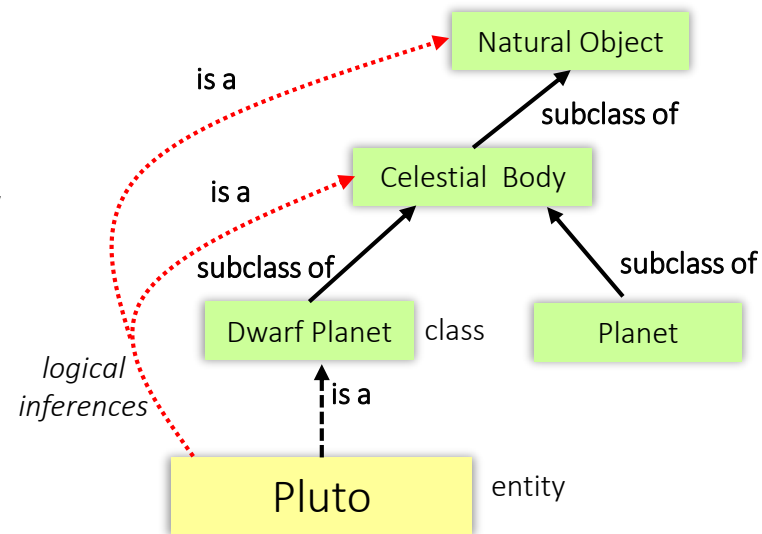


- The Semantic Web (Web of Data) is for machines

- 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

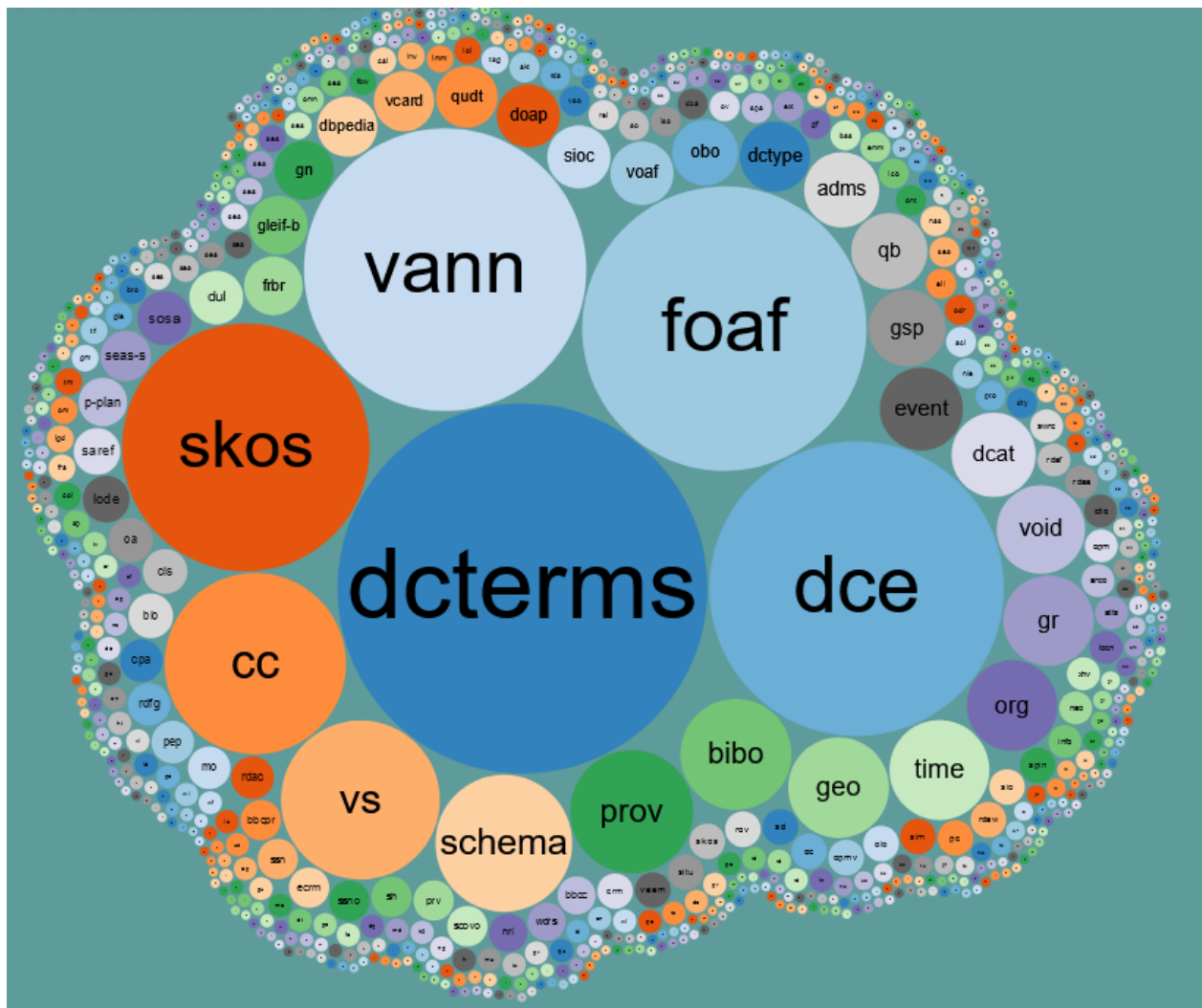


Make our datasets accessible and interoperable on the Web...

- 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>



Web evolution

Web 1.0

Web 2.0

Web 3.0

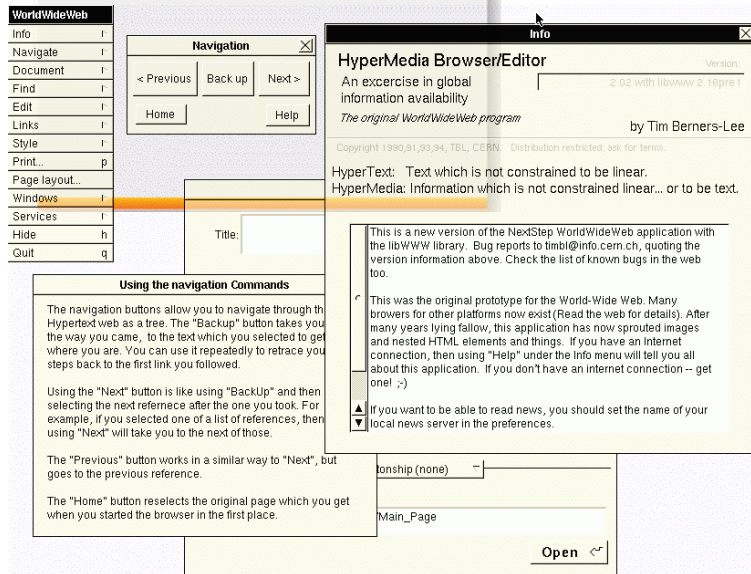
Web evolution

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

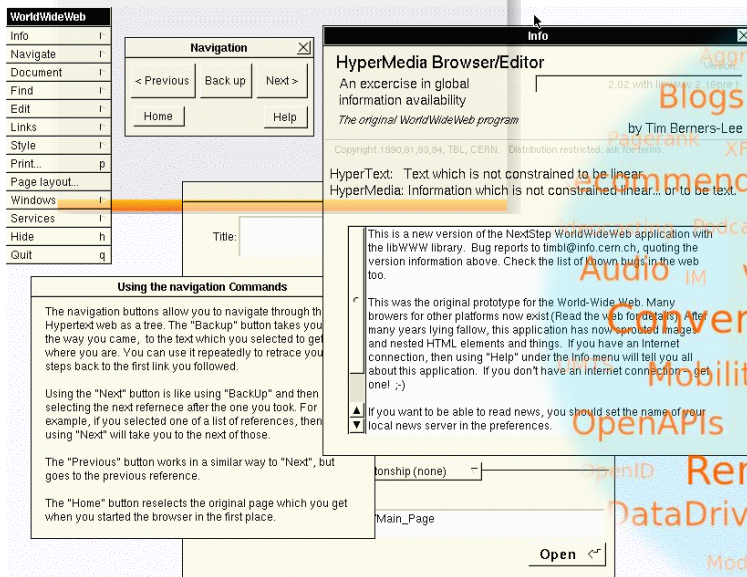
Web 1.0

- 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



Web 2.0

- Participation
- Collaboration
- Social Software
- Sharing
- Perpetual Beta
- Design
- Simplicity
- Joy of Use
- Usability
- Widgets
- Browser
- AJAX
- Pay Per Click
- CSS
- Affiliation
- Economy
- Standardization
- Microformats
- Syndication
- XML
- SOAP
- Accessibility
- REST
- Web Standards
- Semantic
- XHTML
- SVG
- Ruby on Rails
- VC
- Trust
- RSS
- Atom
- Video
- Urgence
- Casting

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

Web evolution

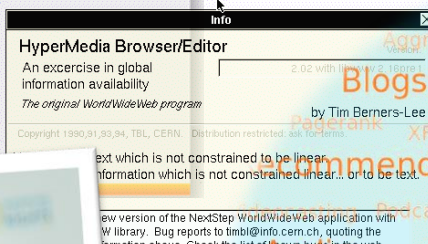
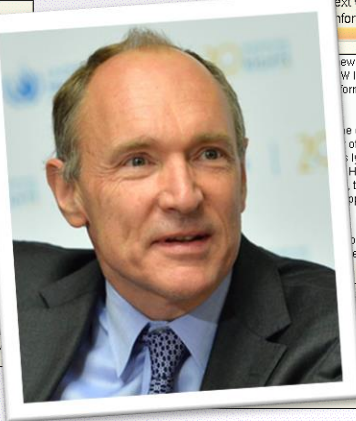
Web 1.0

- Utilisateur passif
- Essentiellement pro
- Réservé aux experts
- Guerre des navigateurs
- Contenu statique
- Formulaire
- 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



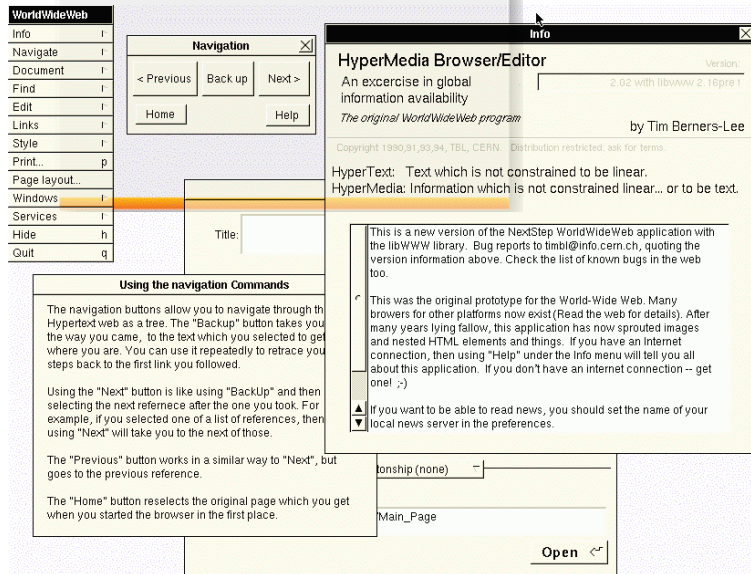
"The **web of human-readable document** is being merged with a **web of machine understandable data**. The potential of the mixture of humans and **machines** working together and communication through the web could be immense."

Tim Berners-Lee, [The World Wide Web: A very short personal history](http://www.w3.org/People/Berners-Lee/ShortHistory.html), 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
- Formulaire
- 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

- Standardization activities <https://www.w3.org/standards/>

STANDARDS

PARTICIPATE

MEMBERSHIP

ABOUT W3C

W3C » Standards

STANDARDS

W3C standards define an **Open Web Platform** for application development that has the unprecedented potential to enable developers to build rich interactive experiences, powered by vast data stores, that are available on any device. Although the boundaries of the platform continue to evolve, industry leaders speak nearly in unison about how HTML5 will be the cornerstone for this platform. But the full strength of the platform relies on many more technologies that W3C and its partners are creating, including CSS, SVG, WOFF, the Semantic Web stack, XML, and a variety of APIs.

W3C develops these technical specifications and guidelines through a process designed to maximize consensus about the content of a technical report, to ensure high technical and editorial quality, and to earn endorsement by W3C and the broader community.

If you are learning about Web technology, you may wish to start with the introduction below, and follow links for greater detail.

Web Design and Applications

Web Design and Applications involve the standards for building and Rendering Web pages, including HTML, CSS, SVG, Ajax, and other technologies for Web Applications ("WebApps"). This section also includes information on how to make pages accessible to people with disabilities (WCAG), to internationalize them, and make them work on mobile devices.

Web of Devices

W3C is focusing on technologies to enable Web access anywhere, anytime, using any device. This includes Web access from mobile phones and other mobile devices as well as use of Web technology in consumer electronics, printers, interactive television, and even automobiles.

Web Architecture

Web Architecture focuses on the foundation technologies and principles which sustain the Web, including URIs and HTTP.

Semantic Web

In addition to the classic "Web of documents" W3C is helping to build a technology stack to support a "Web of data," the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term "Semantic Web" refers to W3C's vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.

XML Technology

XML Technologies including XML, XML Namespaces, XML Schema, XSLT, Efficient XML Interchange (EXI), and other related standards.

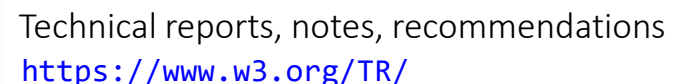
Web of Services

Web of Services refers to message-based design frequently found on the Web and in enterprise software. The Web of Services is based on technologies such as HTTP, XML, SOAP, WSDL, SPARQL, and others.

Browsers and Authoring Tools

The web's usefulness and growth depends on its universality. We should be able to publish regardless of the software we use, the computer we have, the language we speak, whether we are wired or wireless, regardless of our sensory or interaction modes. We should be able to access the web from any kind of hardware that can connect to the Internet – stationary or mobile, small or large. W3C facilitates this listening and blending via international web standards. These standards ensure that all the crazy brilliance continues to improve a web that is open to us all.

Questions About Standards?



Web standardization

- Semantic web standards

The Semantic Web Technology Stack (not a piece of cake...)

Most apps use only a subset of the stack

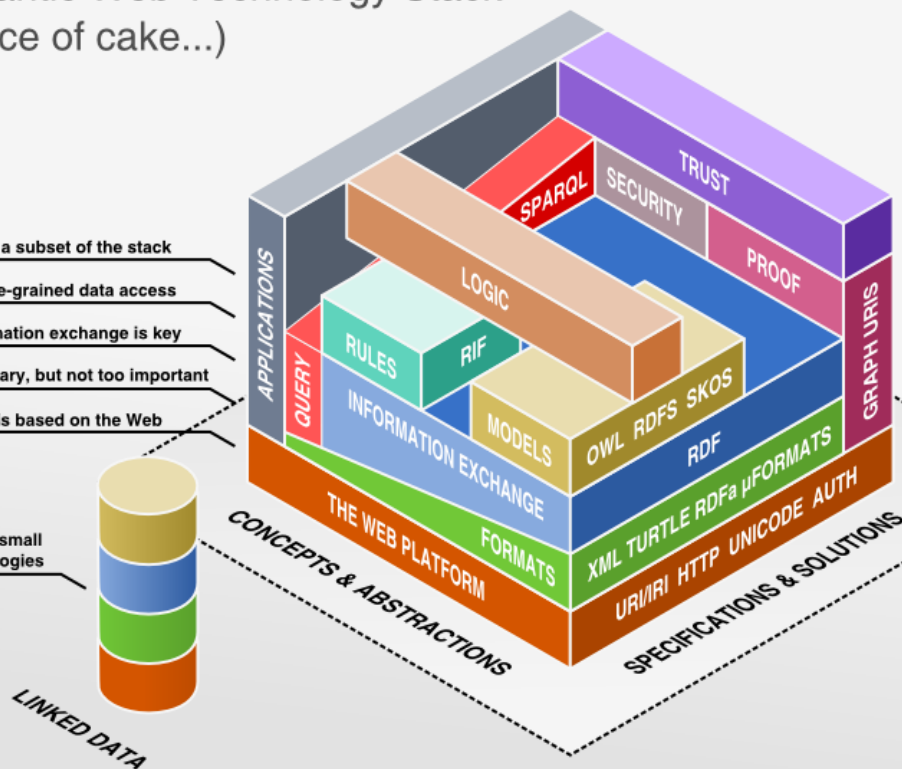
Querying allows fine-grained data access

Standardized information exchange is key

Formats are necessary, but not too important

The Semantic Web is based on the Web

Linked Data uses a small selection of technologies



W3C SEMANTIC WEB ACTIVITY

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



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/>

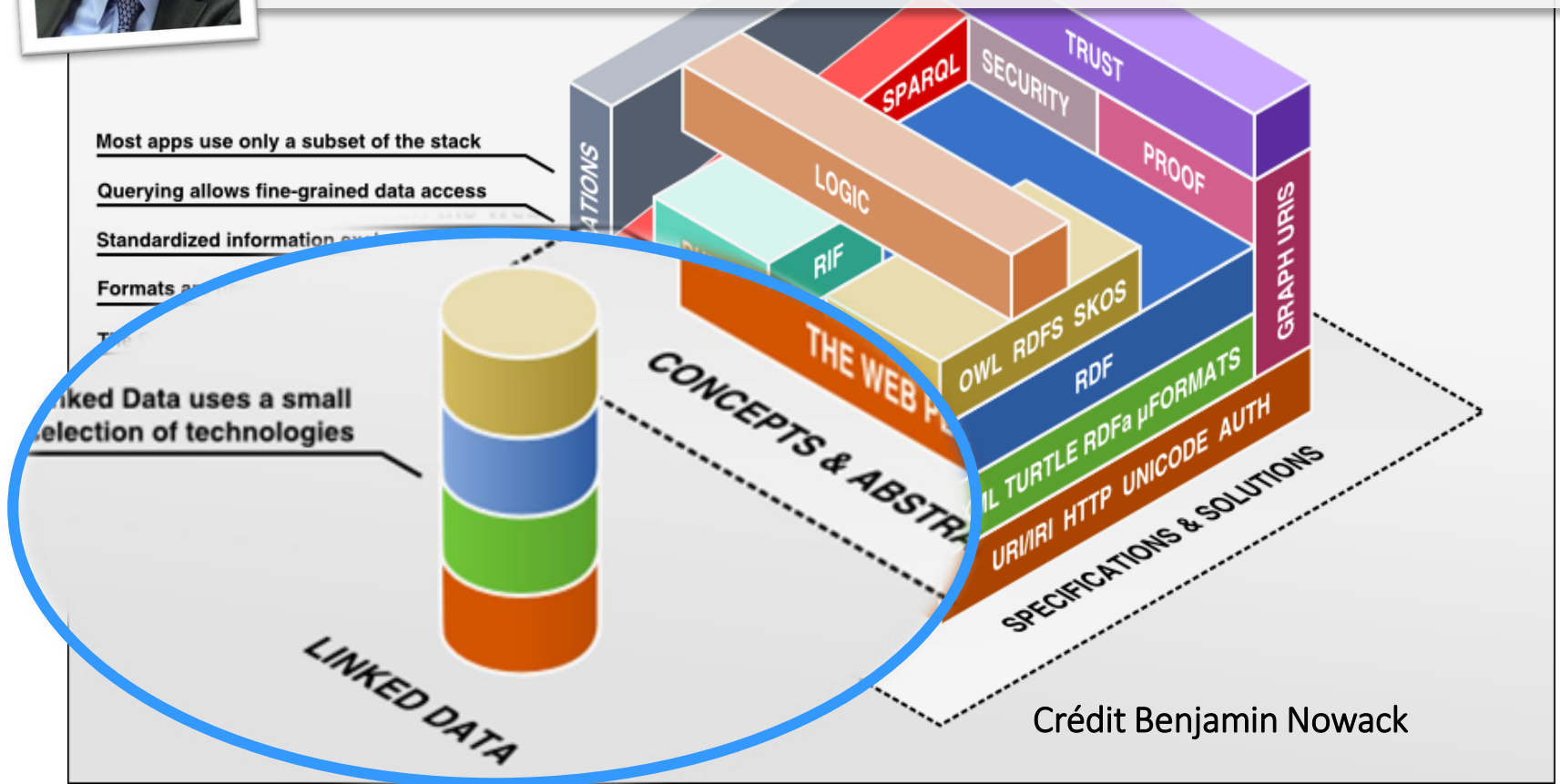
(Crédit Benjamin Nowack)

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

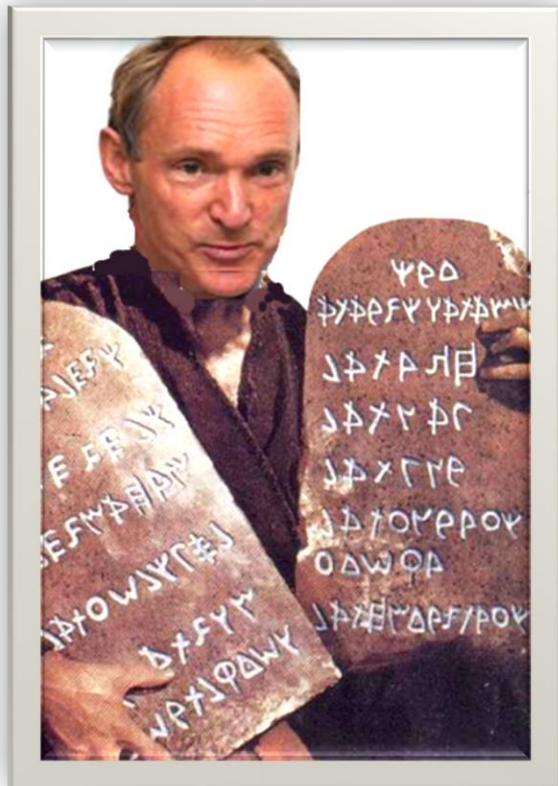
<http://www.w3.org/DesignIssues/LinkedData.html>

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>

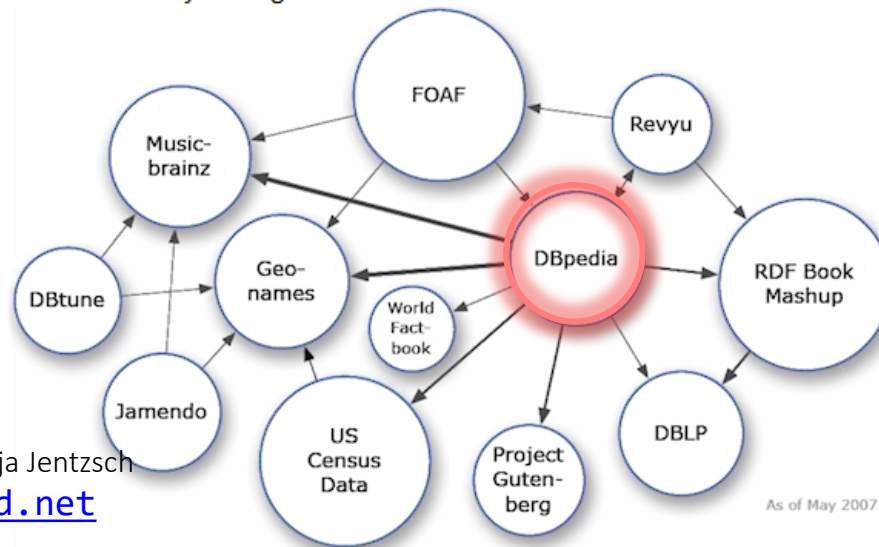


Linked Open Data (LOD)



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 SWEOW 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>

A Faire : Quelques ressources à regarder

- The semantic web of data Tim Berners-Lee :
https://www.youtube.com/watch?v=HeUrEh-nqtU&ab_channel=dilshod2006
- Le web sémantique, web des données en 10 mn :
https://www.youtube.com/watch?v=r795n3AffgA&ab_channel=ChristianAubry
- Les vidéos de Manu Sporny:
 - Intro Web-Sem :
https://www.youtube.com/watch?v=OGg8A2zfWKg&ab_channel=ManuSporny
 - Les donnéesliées:
https://www.youtube.com/watch?v=4x_xzT5eF5Q&ab_channel=ManuSporny
 - JSON-LD : https://www.youtube.com/watch?v=vioCbTo3C-4&ab_channel=ManuSporny
- Google knowledge graph : <https://www.youtube.com/watch?v=mmQl6VGvX-c>
https://www.youtube.com/watch?v=mmQl6VGvX-c&ab_channel=Google
- Global data space : <http://linkeddatabook.com/editions/1.0/>
- RDF URI SPARQL : <https://www.youtube.com/watch?v=FvGndkpa4K0>
https://www.youtube.com/watch?v=FvGndkpa4K0&ab_channel=bobdc