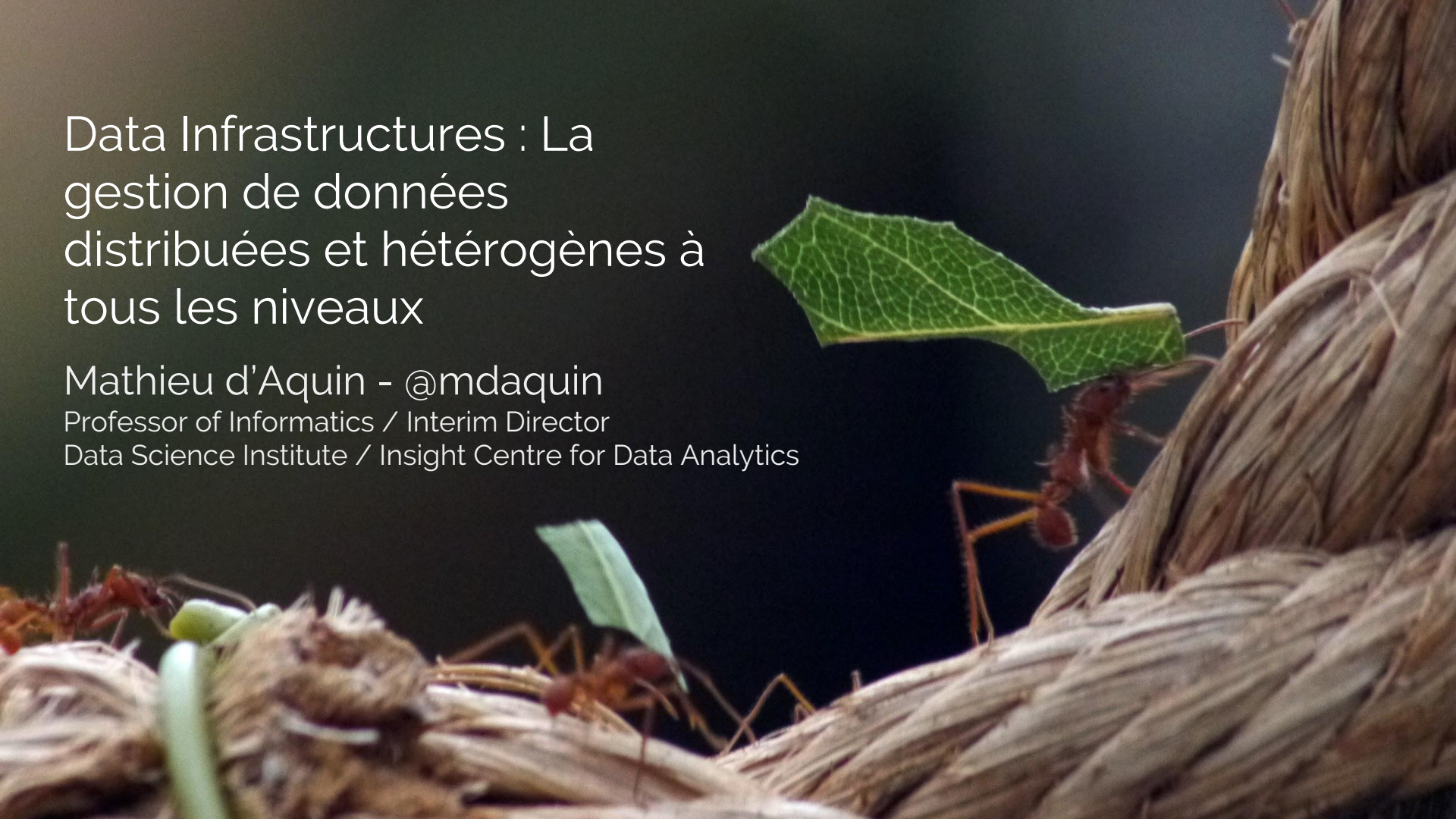
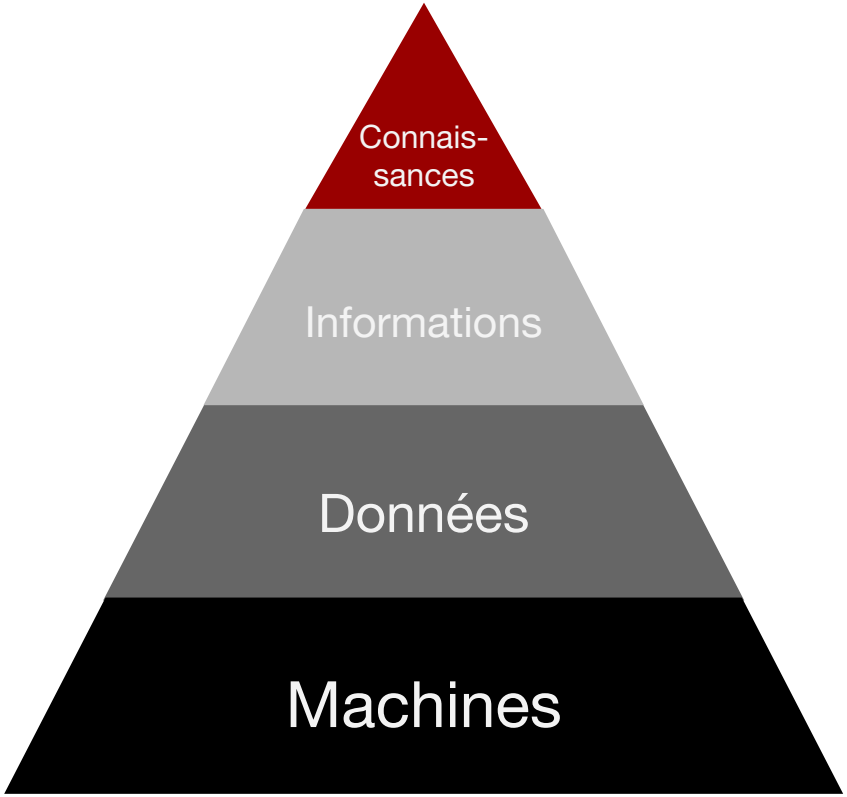


Data Infrastructures : La gestion de données distribuées et hétérogènes à tous les niveaux

Mathieu d'Aquin - @mdaquin
Professor of Informatics / Interim Director
Data Science Institute / Insight Centre for Data Analytics

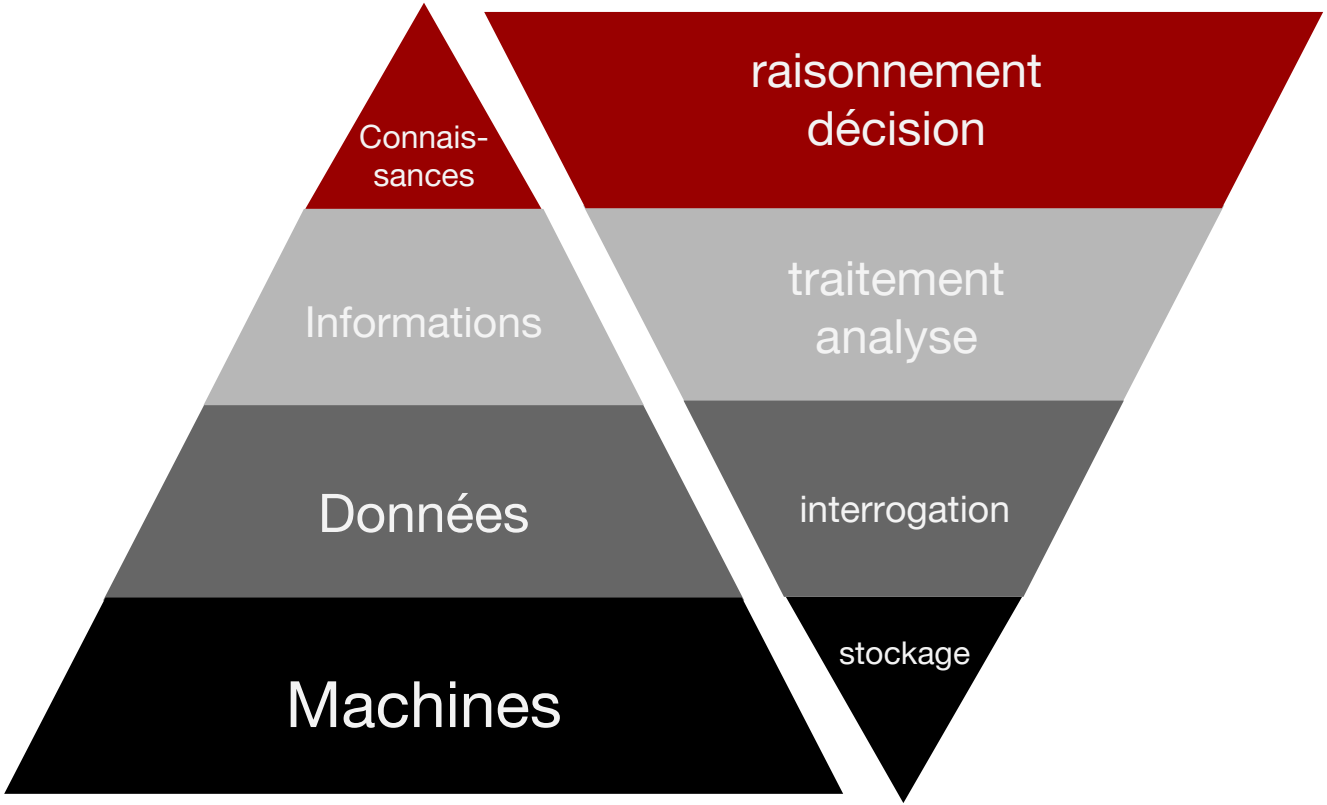
Pyramide de l'information



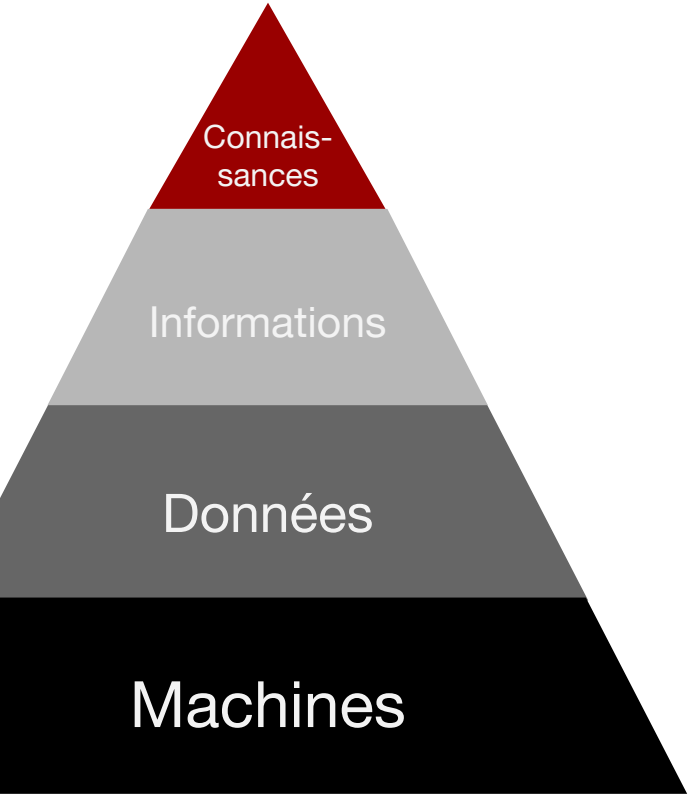
Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth. Truth is not beauty. Beauty is not love. Love is not music. Music is the best"

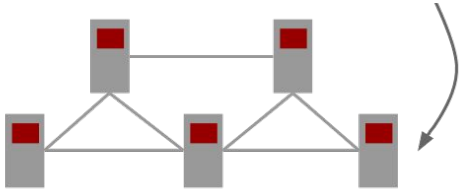
- Frank Zappa

Pyramide de l'information



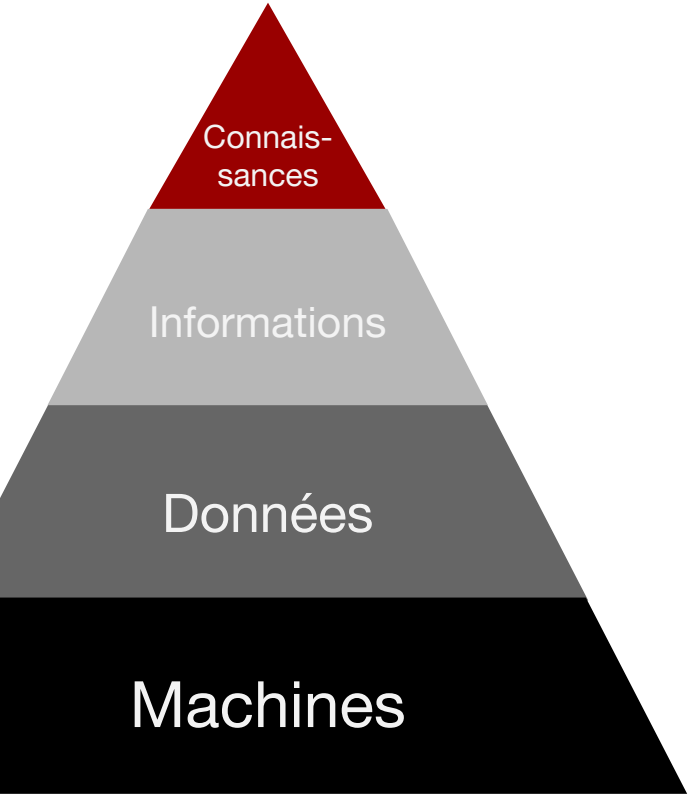
Pyramide de l'information et le Web (Sémantique)

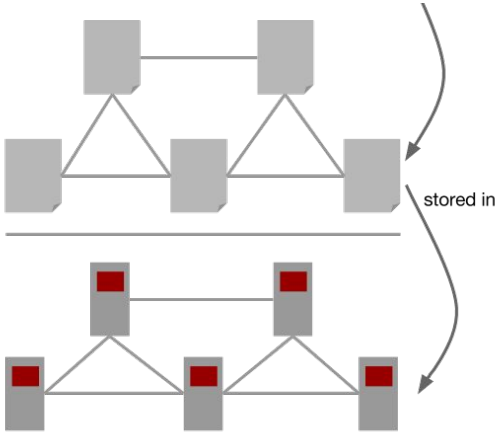




Internet
Réseau de machines

Pyramide de l'information et le Web (Sémantique)

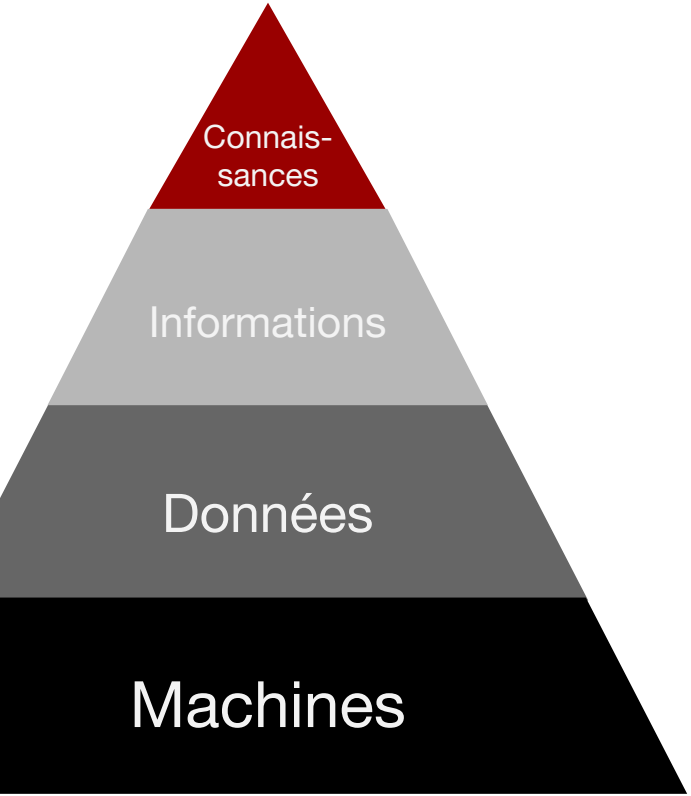


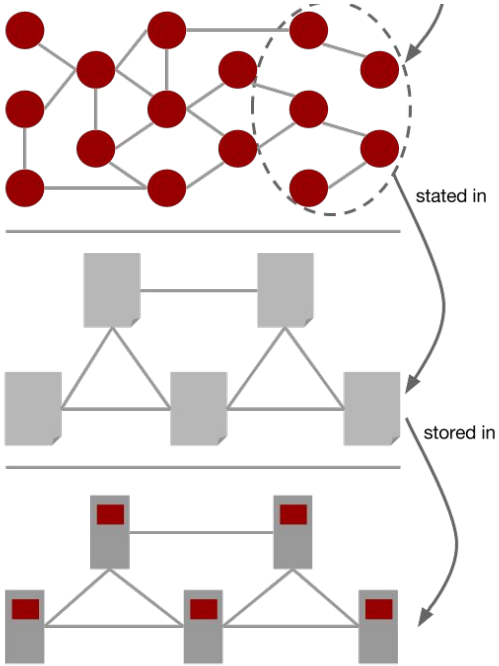


Le Web
Réseau de documents

Internet
Réseau de machines

Pyramide de l'information et le Web (Sémantique)



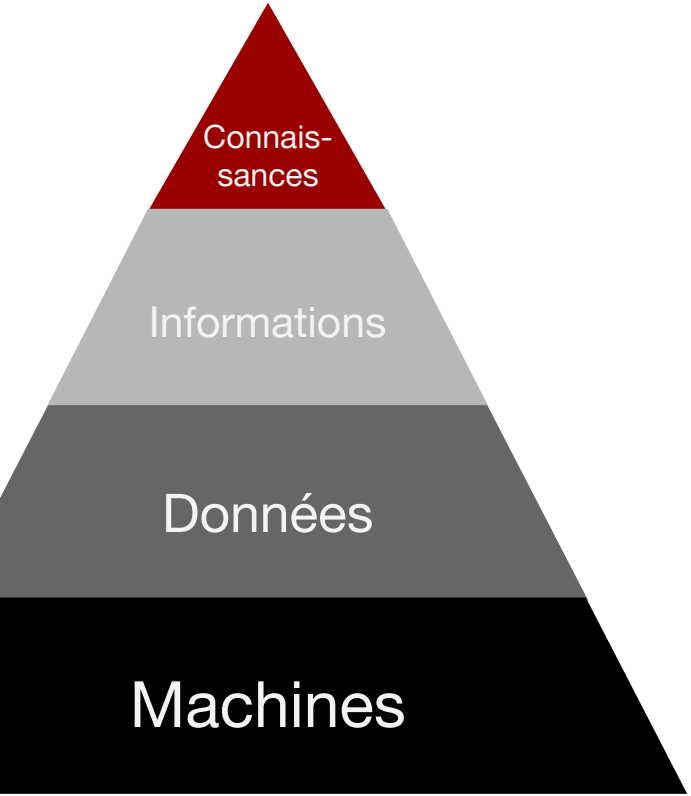


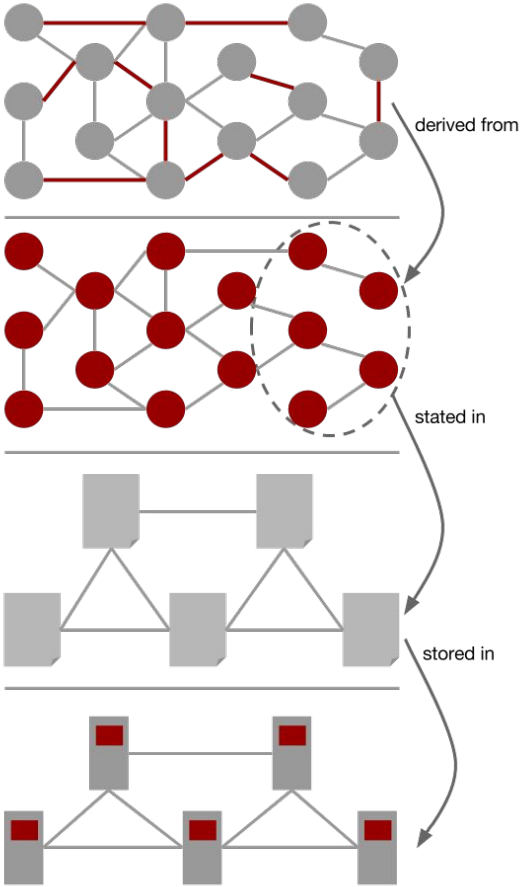
Les données liées
Réseau d'information

Le Web
Réseau de documents

Internet
Réseau de machines

Pyramide de l'information et le Web (Sémantique)



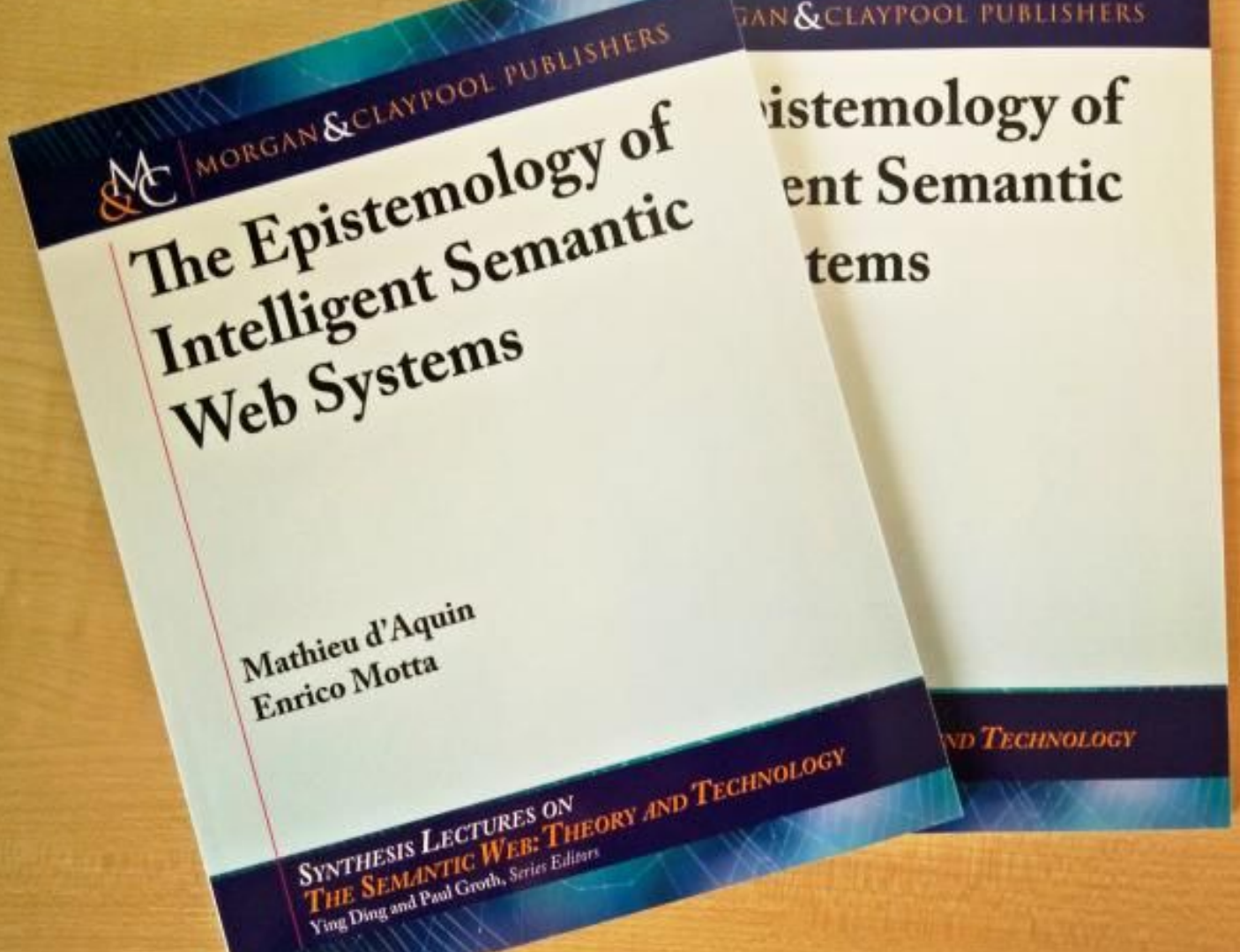


Le web sémantique
Réseau de connaissances

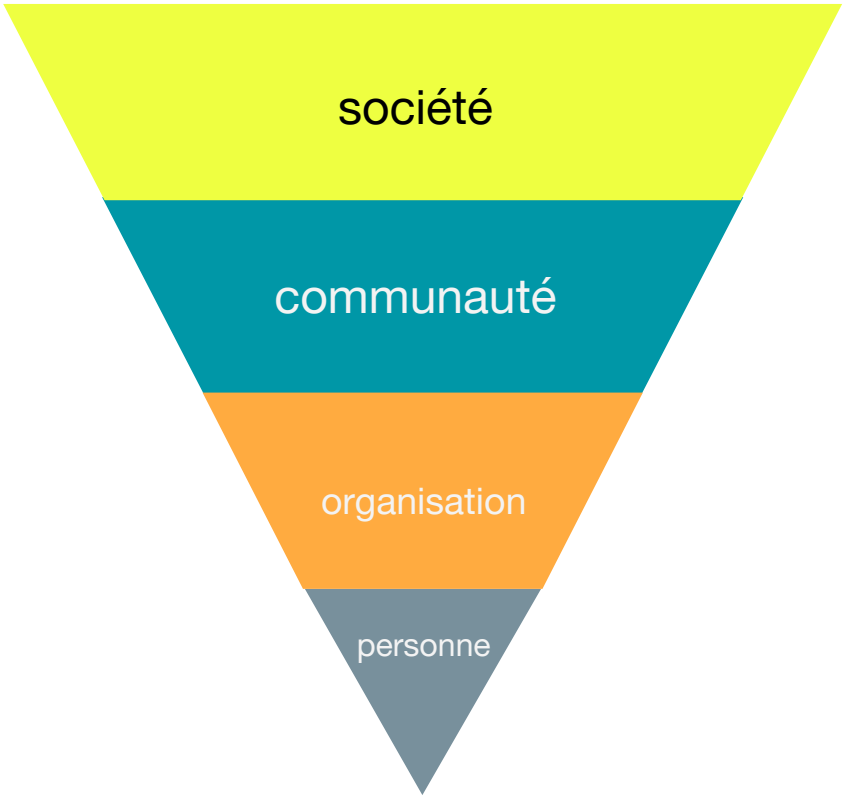
Les données liées
Réseau d'information

Le Web
Réseau de documents

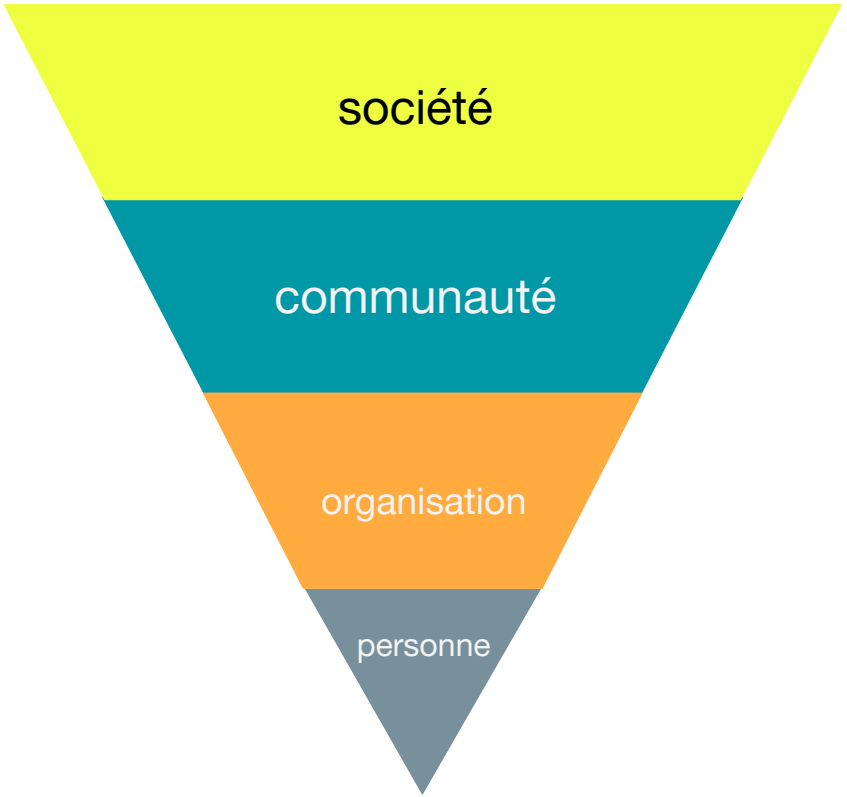
Internet
Réseau de machines


MORGAN & CLAYPOOL PUBLISHERS
The Epistemology of Intelligent Semantic Web Systems
Mathieu d'Aquin
Enrico Motta
SYNTHESIS LECTURES ON
THE SEMANTIC WEB: THEORY AND TECHNOLOGY
Ying Ding and Paul Groth, Series Editors
&CLAYPOOL PUBLISHERS
istemology of
ent Semantic
tems
ND TECHNOLOGY

Pyramide des implications du traitement de l'information

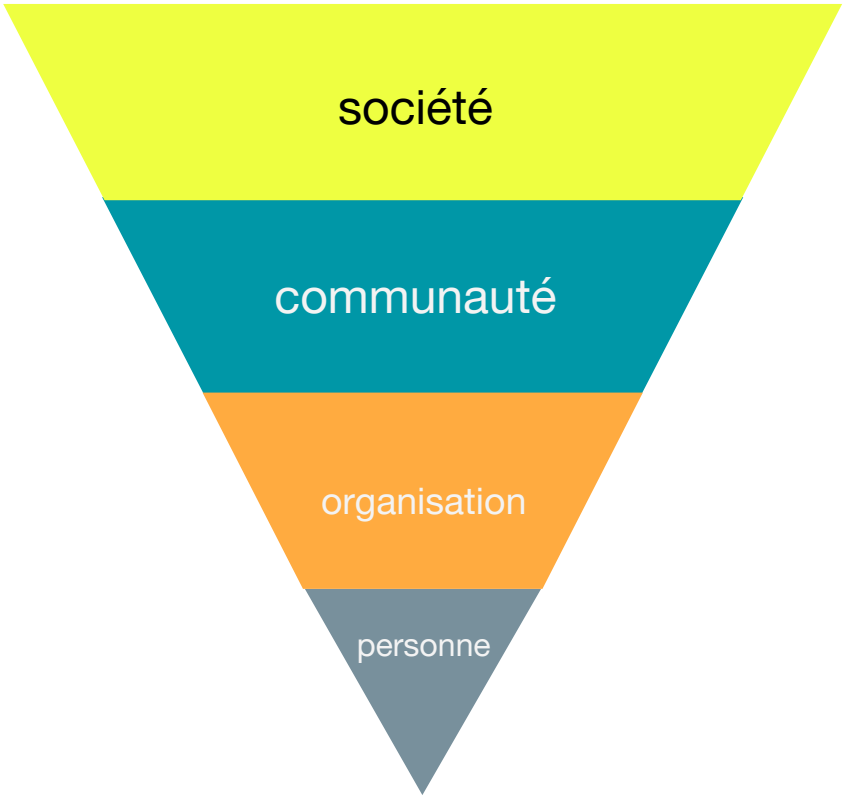


Pyramide des implications du traitement de l'information



Le traitement de l'information touche une personne, un individue.

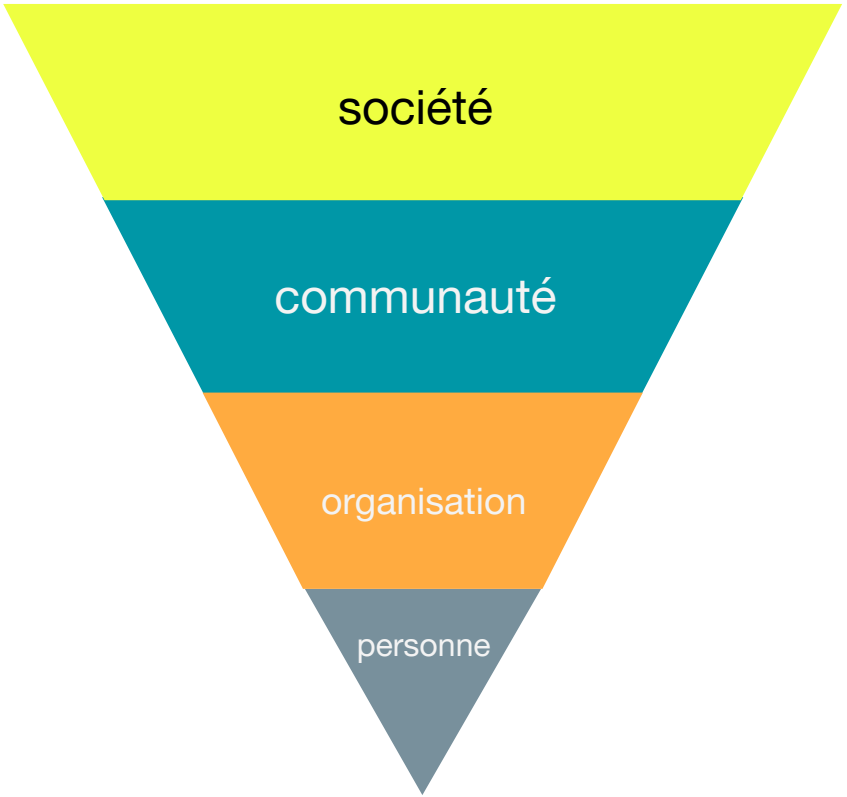
Pyramide des implications du traitement de l'information



Le traitement de l'information touche une compagnie, une entreprise, une institution, etc.

Le traitement de l'information touche une personne, un individue.

Pyramide des implications du traitement de l'information

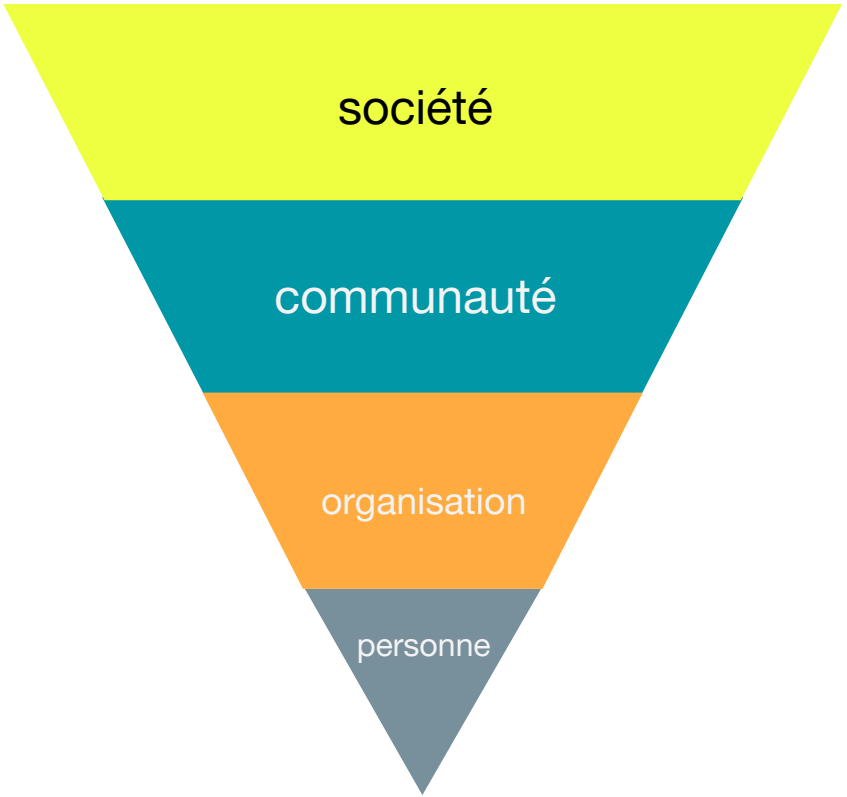


Le traitement de l'information touche des groupes sociaux, regroupements géographiques, etc.

Le traitement de l'information touche une compagnie, une entreprise, une institution, etc.

Le traitement de l'information touche une personne, un individue.

Pyramide des implications du traitement de l'information



Le traitement de l'information touche la démocratie, la liberté, l'égalité, la justice, l'éthique, etc.

Le traitement de l'information touche des groupes sociaux, regroupements géographiques, etc.

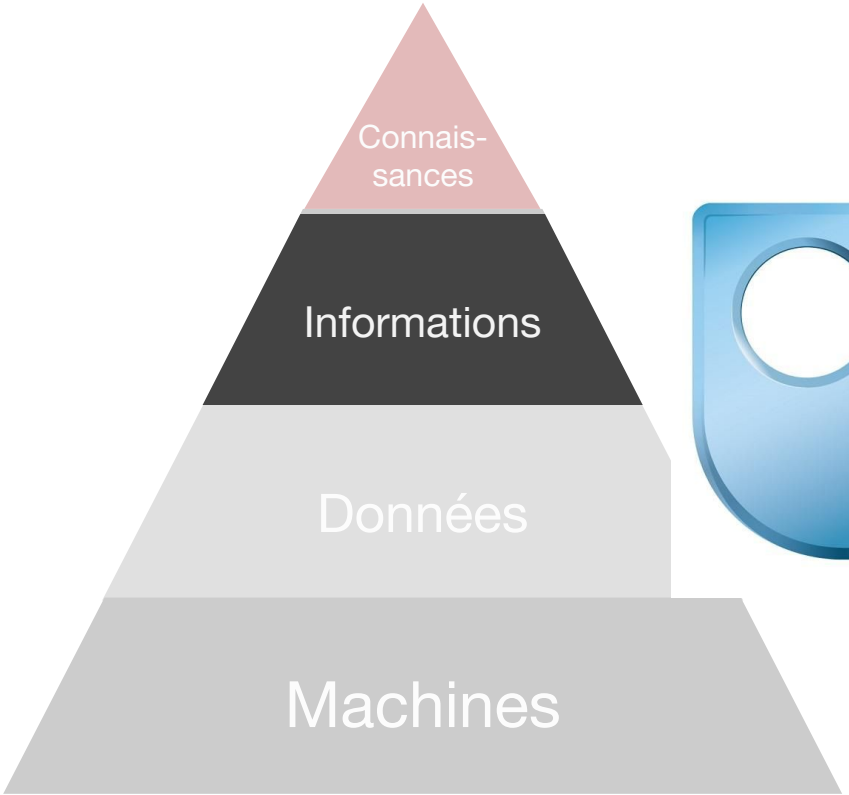
Le traitement de l'information touche une compagnie, une entreprise, une institution, etc.

Le traitement de l'information touche une personne, un individue.

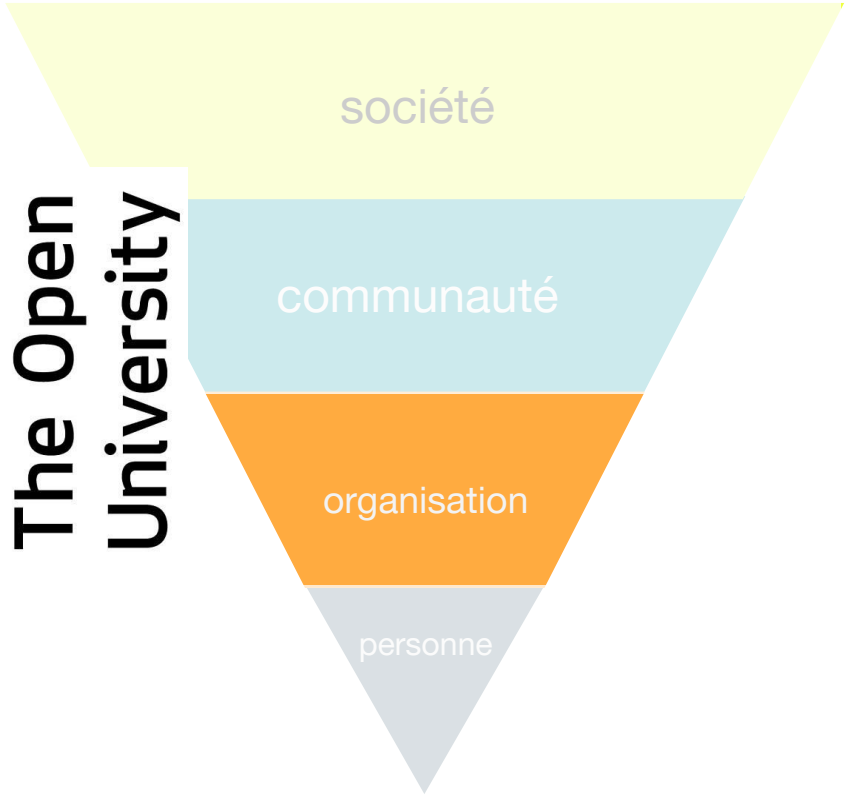
Des exemples

d'infrastructures de données à différents niveaux

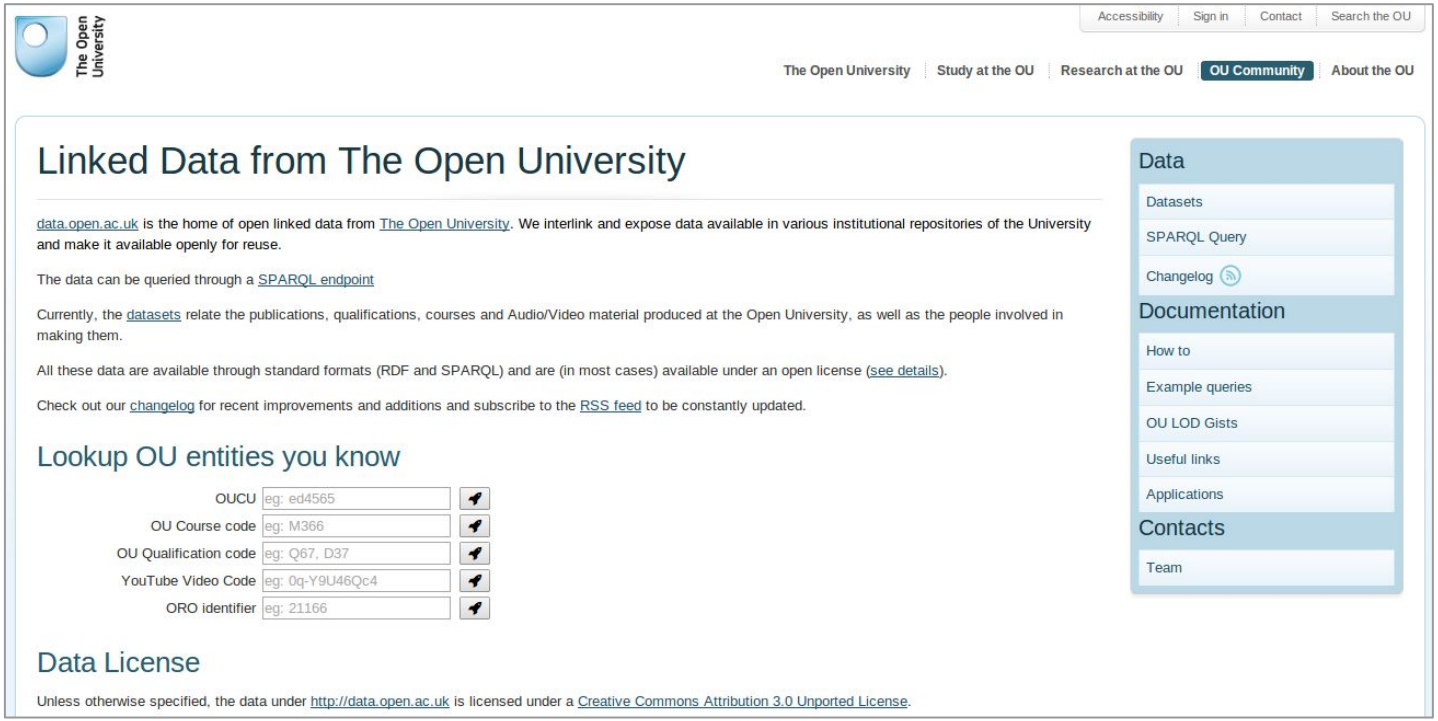
data.open.ac.uk



The Open University

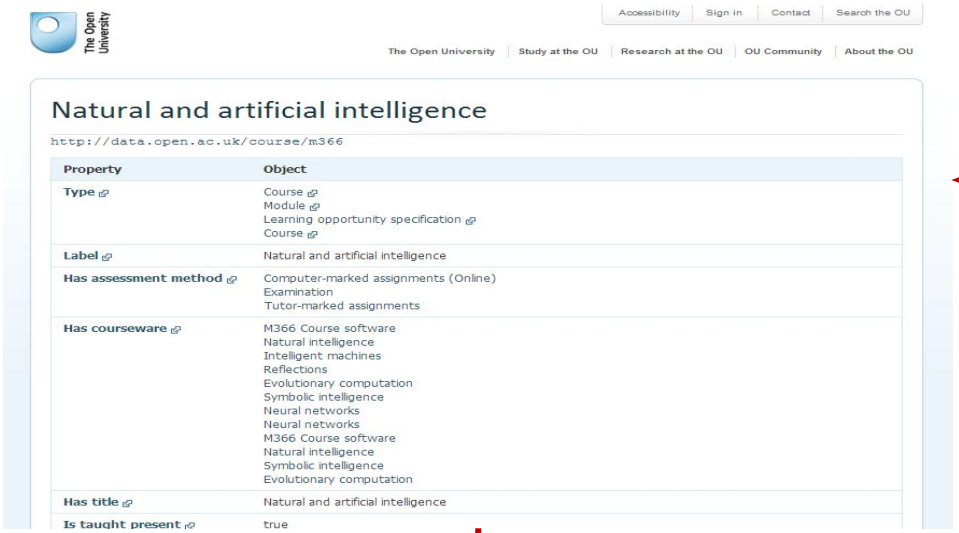


data.open.ac.uk

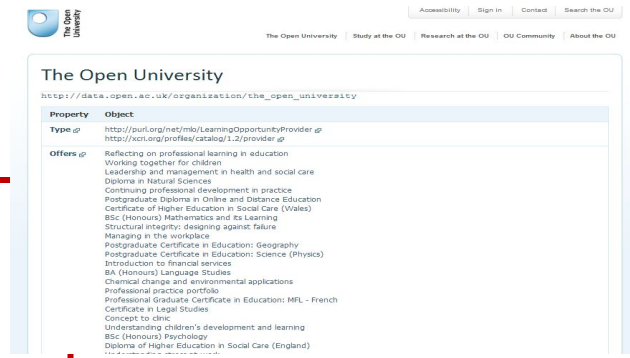


http://data.open.ac.uk/course/m366

http://data.open.ac.uk/organization/the_open_university

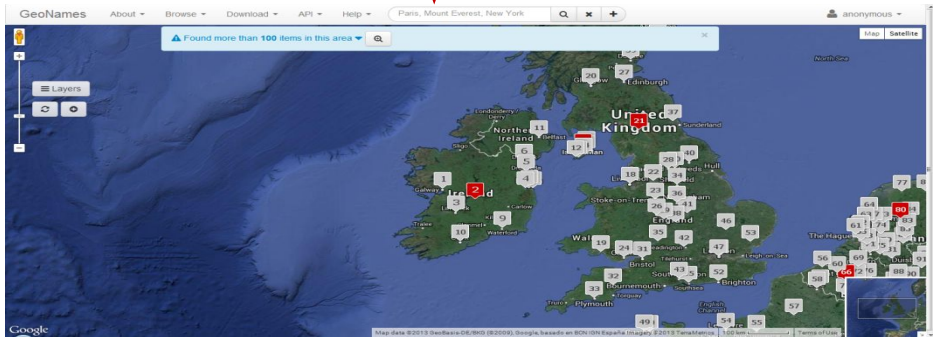


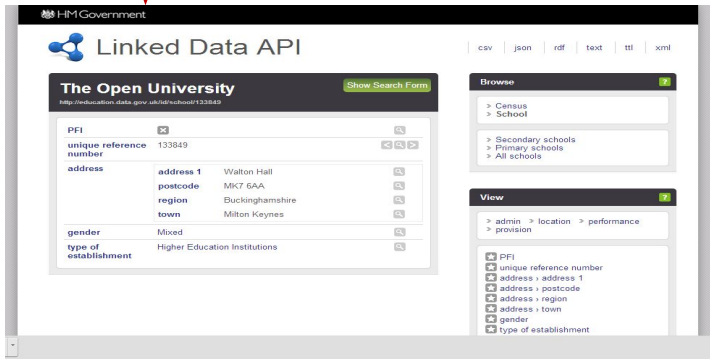
mlo:offers



mlo:location

owl:sameAs

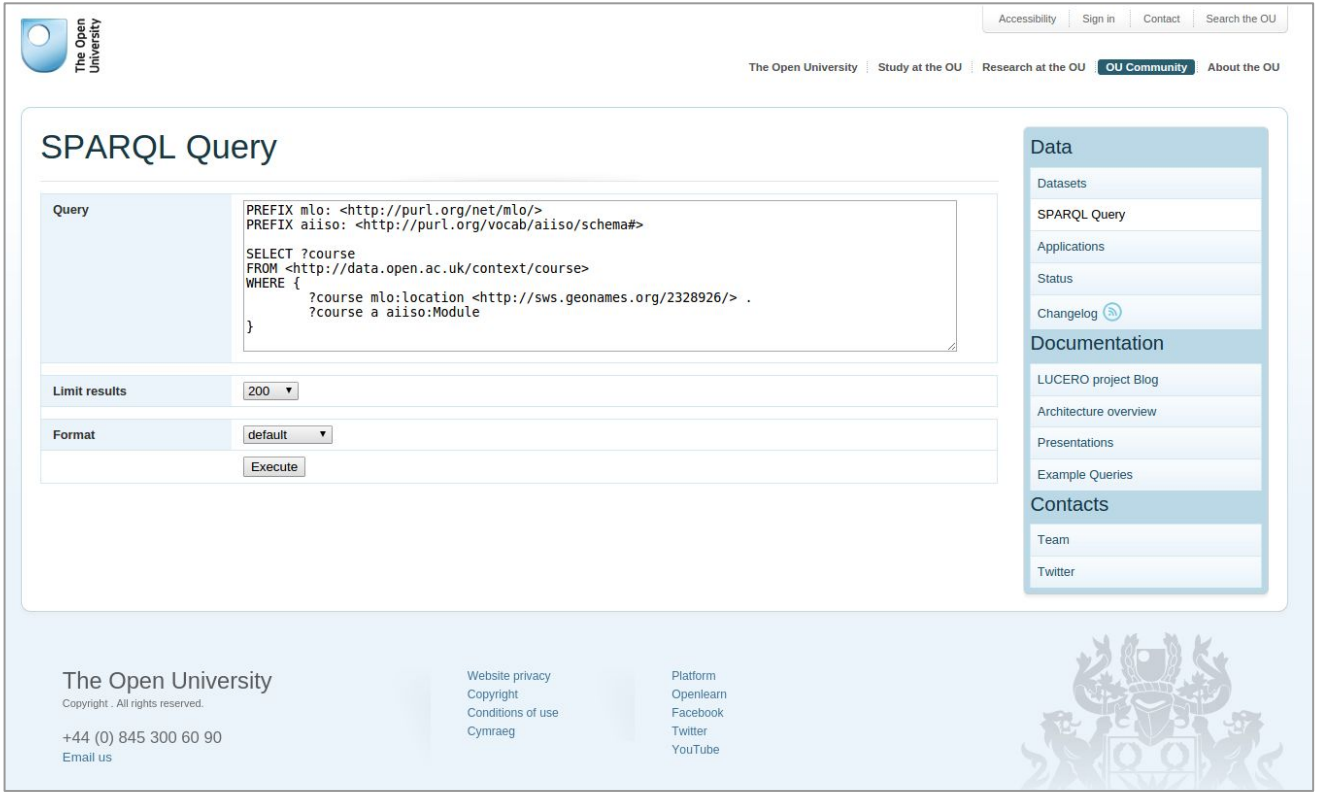




http://sws.geonames.org/2963597/ (Ireland)

http://education.data.gov.uk/id/school/133849

Data -> Interrogation



Applications

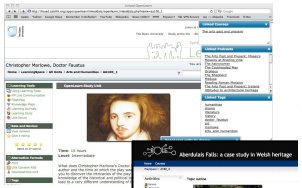
Mobile and Personal Semantics

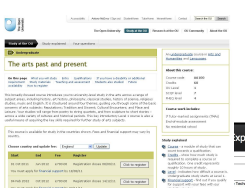
Social

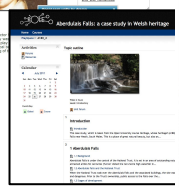
Resource Discovery

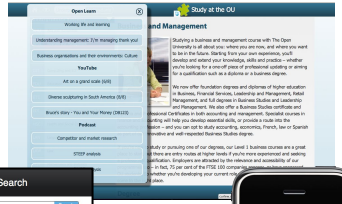
Exploration

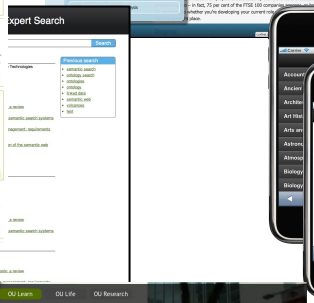
Research

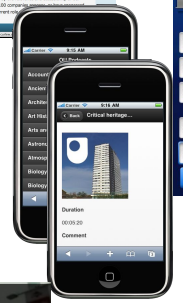


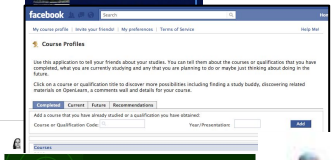


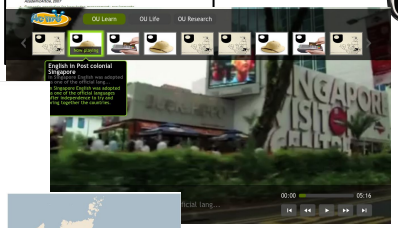


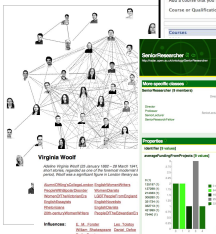


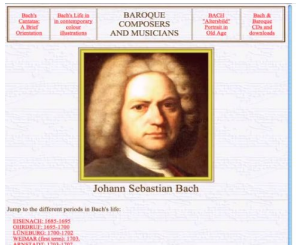


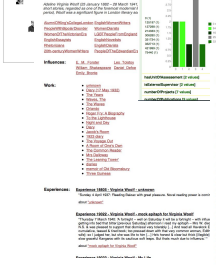


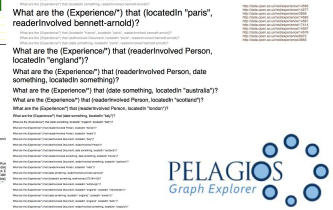












PELAGIOS Graph Explorer

Au delà d'une organisation

Linked Universities

Home About Members Vocabularies Tools Datasets and Endpoints Publications Search

Welcome to LinkedUniversities.org

Linked Universities is an alliance of european universities engaged into exposing their public data as linked data. Linked data is a set of principles to put raw data on the Web, making them Web addressable and linkable, so that they can be easily accessed, discovered, connected and reused. The idea is that data from different institutions and organisations can contribute to a common data space on the Web: the Web of Data.

There are only a few universities currently exposing their public data as linked data, using technologies such as RDF and SPARQL to give direct access to information such as their publications, courses, educational material, etc. These initiatives are currently often disconnected from each other. For every new site being developed, a lot of the efforts required to build a university linked data platform need to be re-deployed, and many of the lessons need to be re-learned. Also, we believe that the potential for linked data in education and research goes well beyond the individual benefit for each institution, as this potential can only be achieved through providing cross-university data that can be aggregated, integreated and compared. While linked data, much like the Web, relies on transparent distribution and a certain amount of self-organisation, we believe that sharing practices and collaborating in the development of university linked data platforms can significantly help towards this common goal, i.e., the creation of a Web of university data.

This portal should therefore essentially be seen as a collaborative space, where institutions and individuals involved in the exposure of university linked data can describe, share and reuse common vocabularies and practices. Our goals are therefore to:

- Identify, support and develop common linked data vocabularies, usable accross universities for common concepts such as courses, qualifications, educational material, etc.
- Describe reusable recipes, and share reusable tools, for exposing linked data in universities
- Support, through experience sharing and reuse, initiatives towards exposing university data as linked data

If you want to obtain more information about *Linked Universities*, discuss your experience with us or become a member, please send an e-mail to info@linkeduniversities.org

Linked Universities

Home About Members Vocabularies Tools Datasets and Endpoints Publications Search

Vocabularies

This page list vocabularies commonly used to describe data related to universities, together with an explaination of the main concepts involved, and links to recipes showing how to express certain situations using these vocabularies.

Course and qualification description

MLO - Metadata for Learning Opportunities

Metadata for Learning Opportunities (MLO) is a European standardized model addressing metadata sufficient for advertising a learning opportunity. The MLO standard is also designed to facilitate semantic technologies and web architectures to support several mechanisms for exchange of the information and aggregation of information by third party service suppliers.

An RDF-S version of MLO has been developed by Scott Wilson and is currently available at http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/ using the namespace http://purl.org/net/mlo/

See for example the use of MLO in course descriptions at the Open University: http://data.open.ac.uk/course/l204

XCRI-CAP - XCRI Course Advertising Profile

XCRI is UK-oriented project to establish a specification to support the eXchange of Course-Related Information. A key activity for XCRI is the development of an XML specification, the XCRI Course Advertising Profile (or XCRI-CAP for short). Learning providers can publish their courses information in the standard XCRI-CAP format, so that it can be collected easily by organisations with course search services. Opening up the offerings of learning providers creates new possibilities for value-added services and information channels for universities, colleges, and training providers.

An RDF-S version of XCRI has been developed by Scott Wilson and is currently available athttp://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/ using the namespace http://xcri.org/profiles/catalog/1.2/

See for example the use of XCRI in course descriptions at the Open University:http://data.open.ac.uk/course/l204

TEACH - Teaching Core Vocabulary

TEACH, the Teaching Core Vocabulary, is a lightweight vocabulary providing terms to enable teachers to relate things in their courses together. The Teaching Core Vocabulary is based on practical requirements set by providing seminar and course descriptions as Linked Data. The vocabulary specification is available at http://linkedscience.org/teach/ns/ and it uses this same URI as the namespace. The suggested namespace prefix is "teach".

See an example of the use of TEACH in course descriptions at the University of Muenster: http://data.uni-muenster.de/context/page/cs/course/171

University as an organisation

AIISO - Academic Institution Internal Structure Ontology

The Academic Institution Internal Structure Ontology (AIISO) provides classes and properties to describe the internal organizational structure of an academic institution.

See for example the use of AIISO in course descriptions at the Open University:http://data.open.ac.uk/course/l204

Bowlogna Ontology

The Bowlogna ontology originates from a lexicon defining terms related to the Bologna Process and aims at providing a standard schema for European universities involved in the Bologna Reform of higher-education studies.

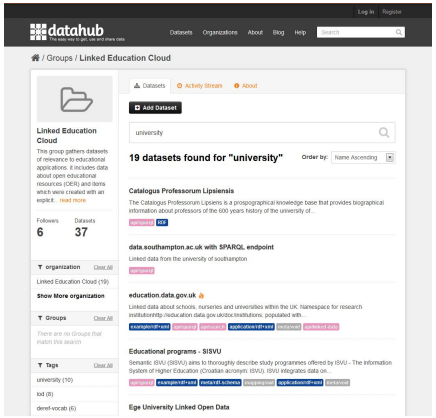
Core organization ontology

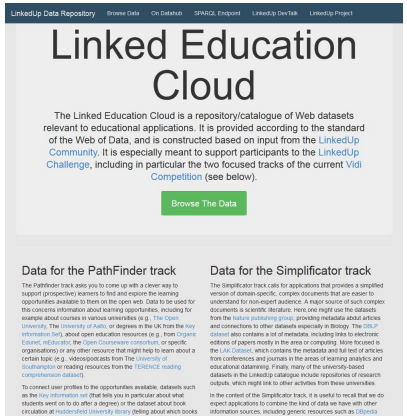
This ontology is not specific to universities, but aim at supporting linked-data publishing of organizational information across a number of domains. It is designed to allow domain-specific extensions to add classification of organisations and roles, as well as extensions to support neighbouring information such as organisational activities.

Academic Publications and Communities

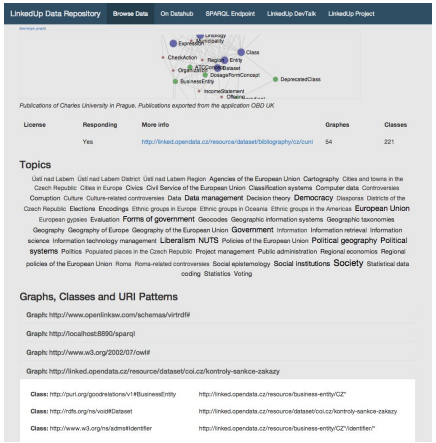
BIBO - The Bibliographic Ontology

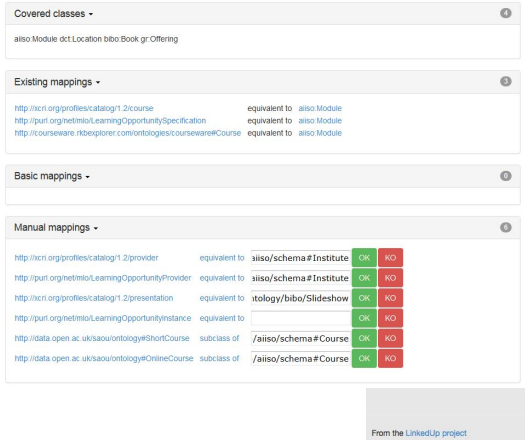
LinkedUp

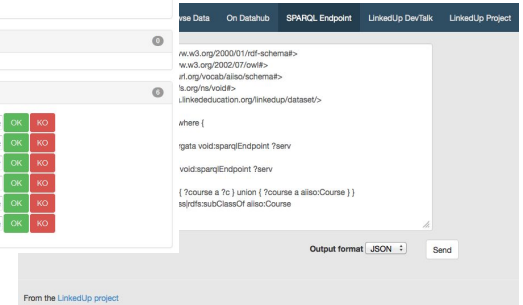

datahub
Datasets
Organizations
About
Blog
Help
/ Groups / Linked Education Cloud
Linked Education Cloud
Add Dataset
university
19 datasets found for "university"
Catalogus Professorum Lipsiensis
data.southampton.ac.uk with SPARQL endpoint
education.data.gov.uk
Educational programs - SISVU
Ege University Linked Open Data


LinkedUp Data Repository
Linked Education Cloud
The Linked Education Cloud is a repository/catalogue of Web datasets relevant to educational applications. It is provided according to the standard of the Web of Data, and is constructed based on input from the LinkedUp Community. It is especially meant to support participants to the LinkedUp Challenge, including in particular the two focused tracks of the current Vidi Competition (see below).
Browse The Data
Data for the PathFinder track
Data for the Simplificator track

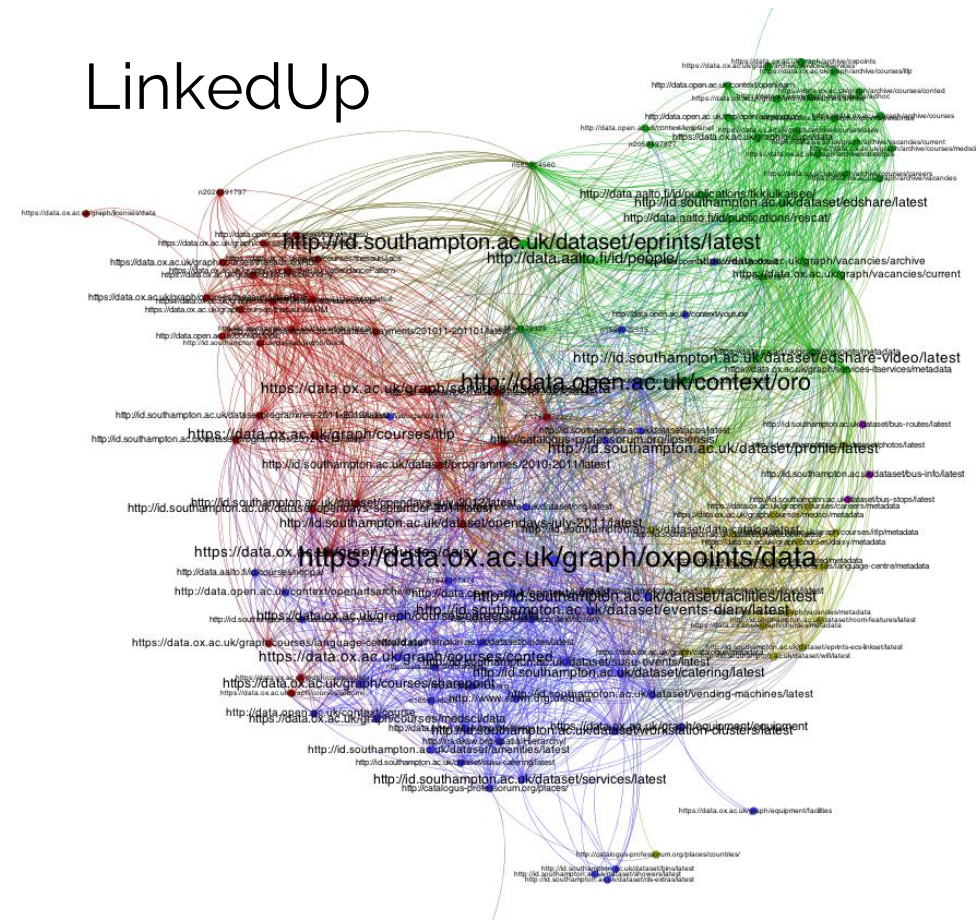

LinkedUp Data Repository
Browse Data
On Datahub
SPARQL Endpoint
LinkedUp DevTalk
LinkedUp Project

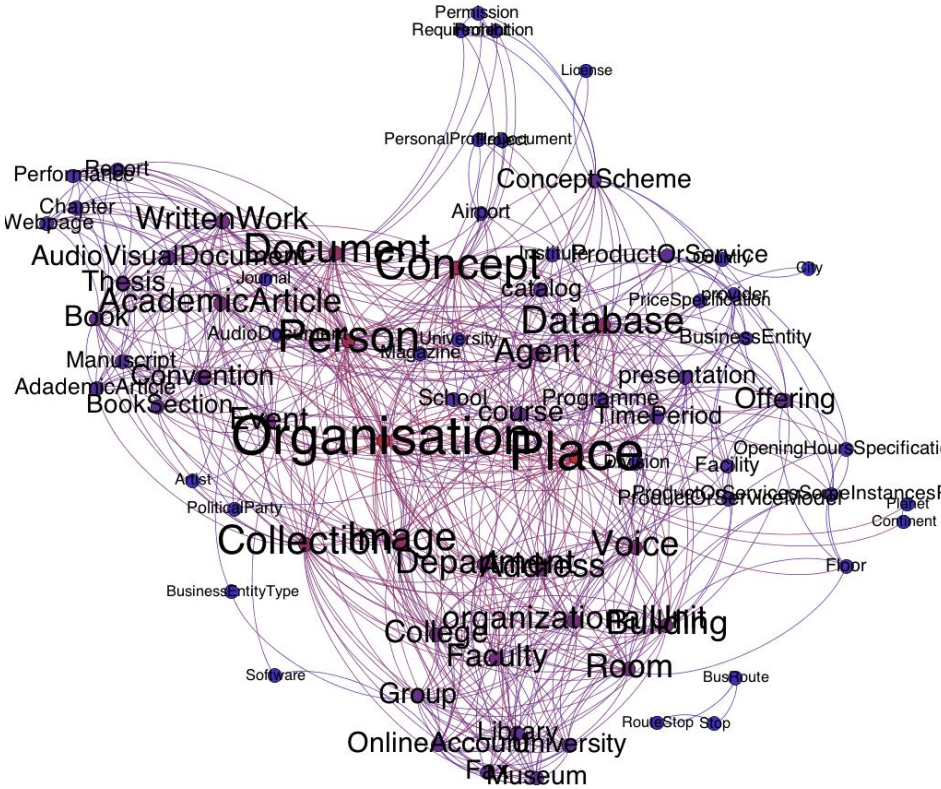

LinkedUp Data Repository
Browse Data
On Datahub
SPARQL Endpoint
LinkedUp DevTalk
LinkedUp Project
Publications of Charles University in Prague, Publications exported from the application OBD UK
Topics
Graphs, Classes and URI Patterns


Covered classes
aiiso:Module dct:Location bibo:Book gr:Offering
Existing mappings
http://xcri.org/profiles/catalog/1.2/course equivalent to aiiso:Module
http://purl.org/net/mlo/LearningOpportunitySpecification equivalent to aiiso:Module
http://courseware.rkbexplorer.com/ontologies/courseware#Course equivalent to aiiso:Module
Basic mappings
Manual mappings
http://xcri.org/profiles/catalog/1.2/provider equivalent to aiiso/schema#Institute
http://purl.org/net/mlo/LearningOpportunityProvider equivalent to aiiso/schema#Institute
http://xcri.org/profiles/catalog/1.2/presentation equivalent to tology/bibo/Slideshow
http://purl.org/net/mlo/LearningOpportunityInstance equivalent to
http://data.open.ac.uk/saou/ontology#ShortCourse subclass of /aiiso/schema#Course
http://data.open.ac.uk/saou/ontology#OnlineCourse subclass of /aiiso/schema#Course
OK
KO


On Datahub
SPARQL Endpoint
LinkedUp DevTalk
LinkedUp Project
Output format JSON
Send
From the LinkedUp project

LinkedUp





see paper at WebSci 2013

This volu
in Education Semina

In the first part of the book, two chap
the current use of linked and open data in educa
technology and the topics that are being covered.

The second part of the book focuses on the specific, pra
that are being put in place to exploit open and linked
today.

The goal of this book is to provide a snapshot of cu
share and disseminate the growing collective expe
data in education. This volume brings together re
practical endeavors from initiatives spread acro
the world. These initiatives are laying the founda
data in the education movement and leading t
applications.

In parallel to the printed boo
in LNCS Online.

Detailed information on L
www.springer.com/lncs

Proposals for publicatio
LNCS Editorial, Tierga
E-mail: lncs@springe

ISSN 0302-9743

ISBN 978-3-31

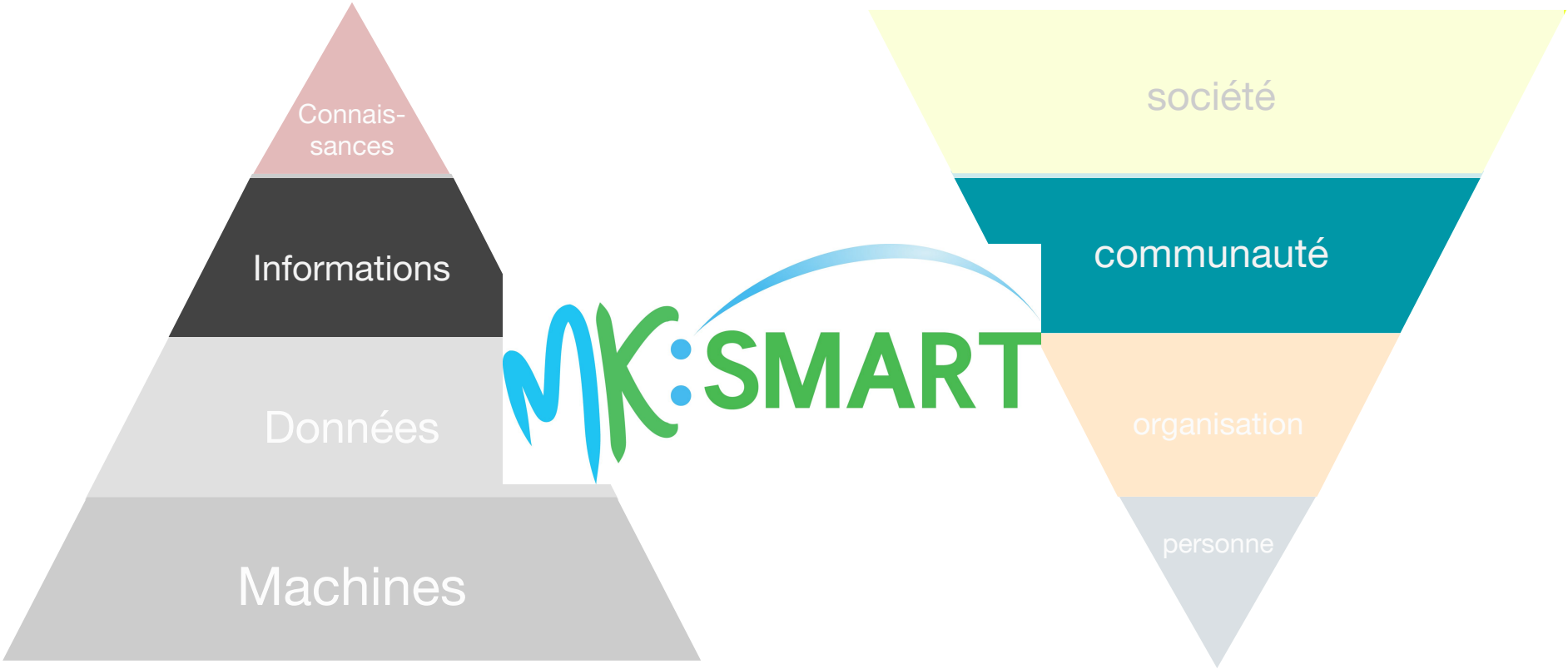
State-o
Survey

LNCS 9500

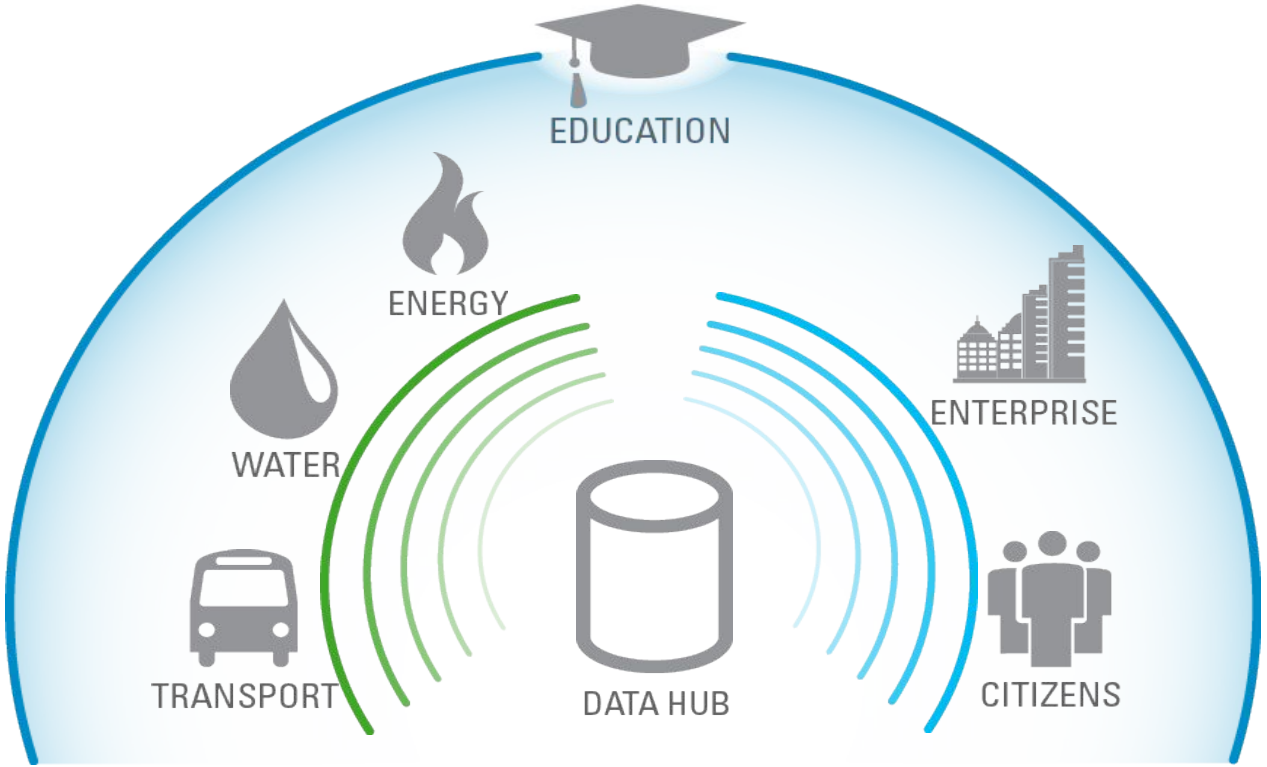
Open Data for Education

Linked, Shared, and Reusable Data for Teaching and Learning

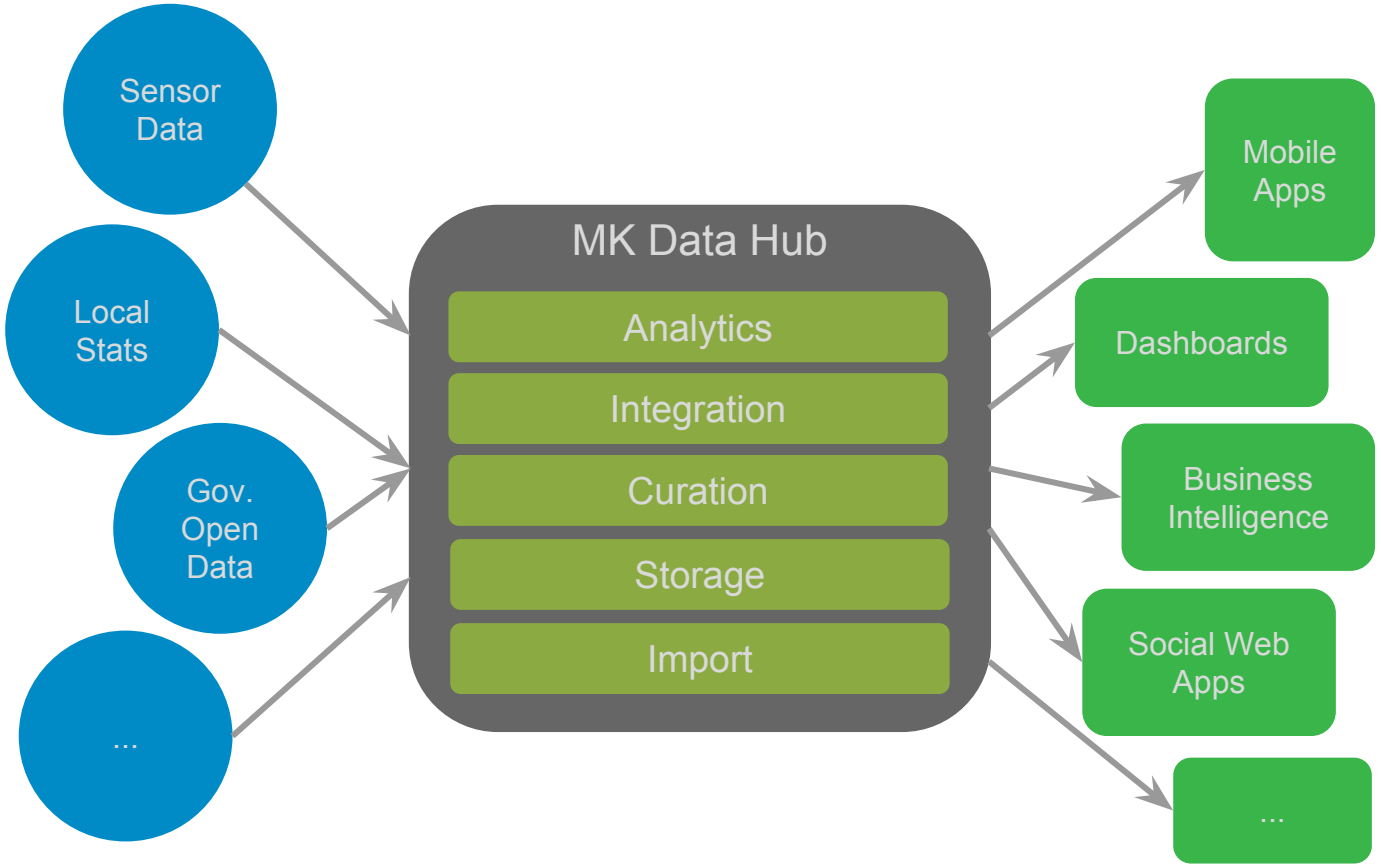
MK:Smart


Connais-
sances
Informations
Données
Machines
MK:SMART
société
communauté
organisation
personne

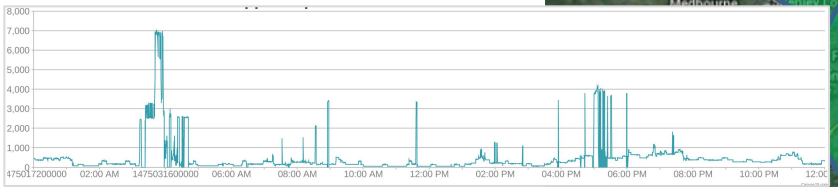
Les données au centre de la ville intelligent



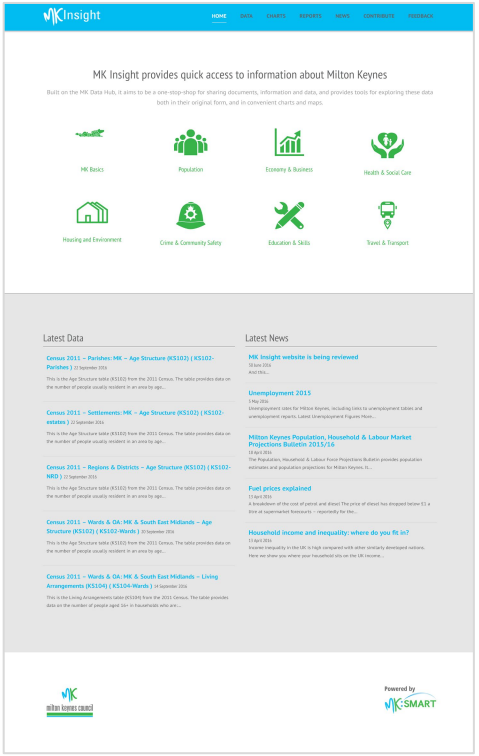
MK Data Hub

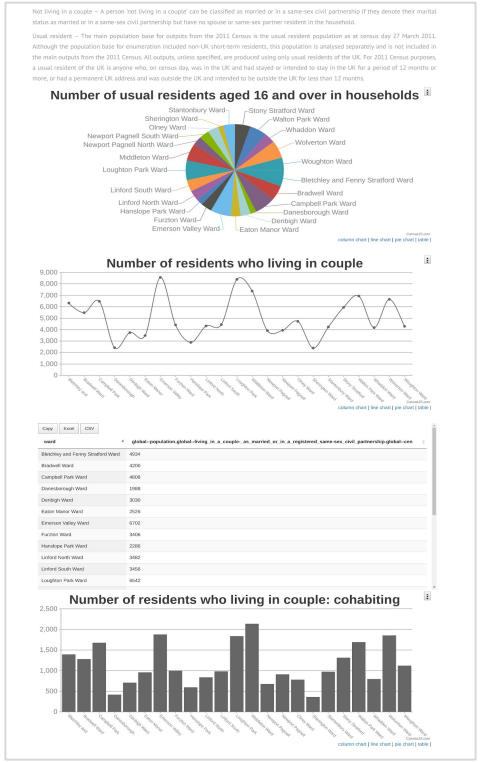

Sensor Data
Local Stats
Gov. Open Data
...
MK Data Hub
Analytics
Integration
Curation
Storage
Import
Mobile Apps
Dashboards
Business Intelligence
Social Web Apps
...

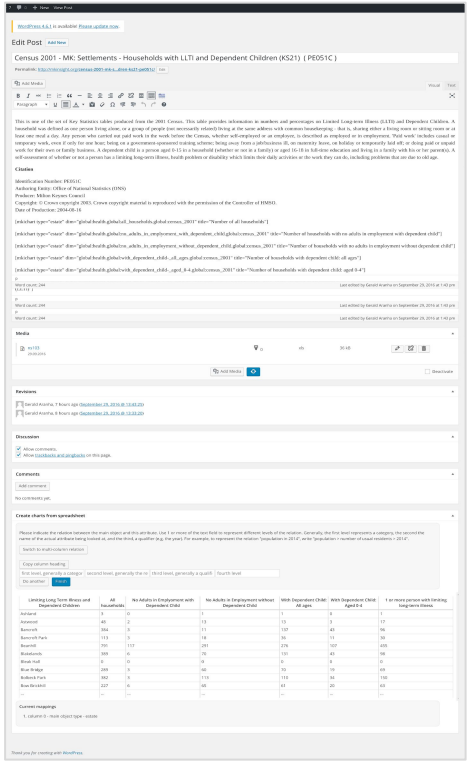
Integre les données temps réel



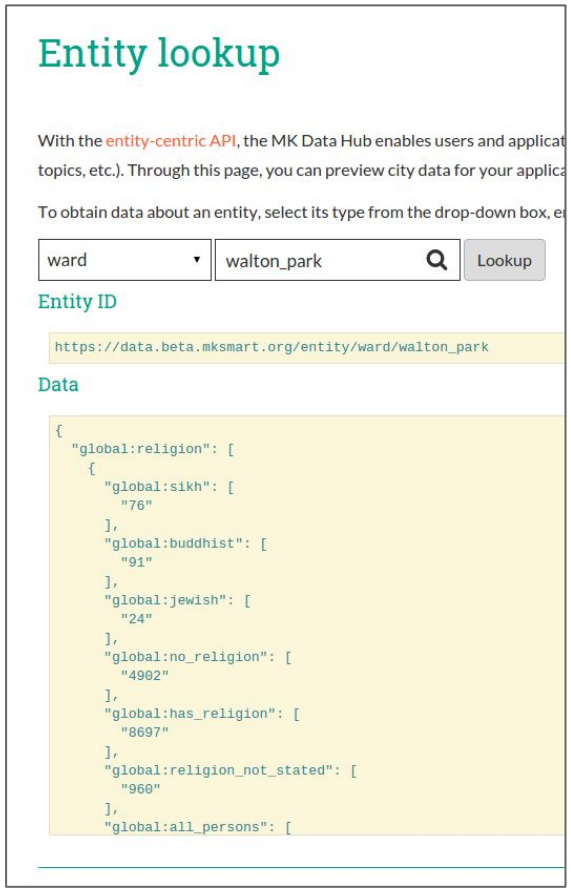
Autant que les données statique







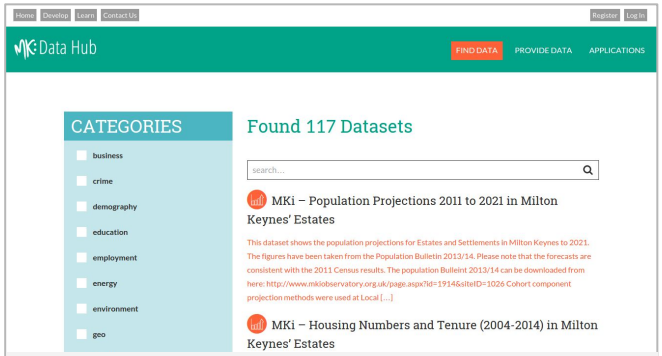
Au travers d'API intelligentes



Une API centrée entités, qui intègre les données de plusieurs sources autour d'un objet/d'une entité, sur le principe du Linked Data.

Compile les résultats "à la demande" (i.e. au moment de la requête) selon un schema virtuel (une ontologie) et évolutif.

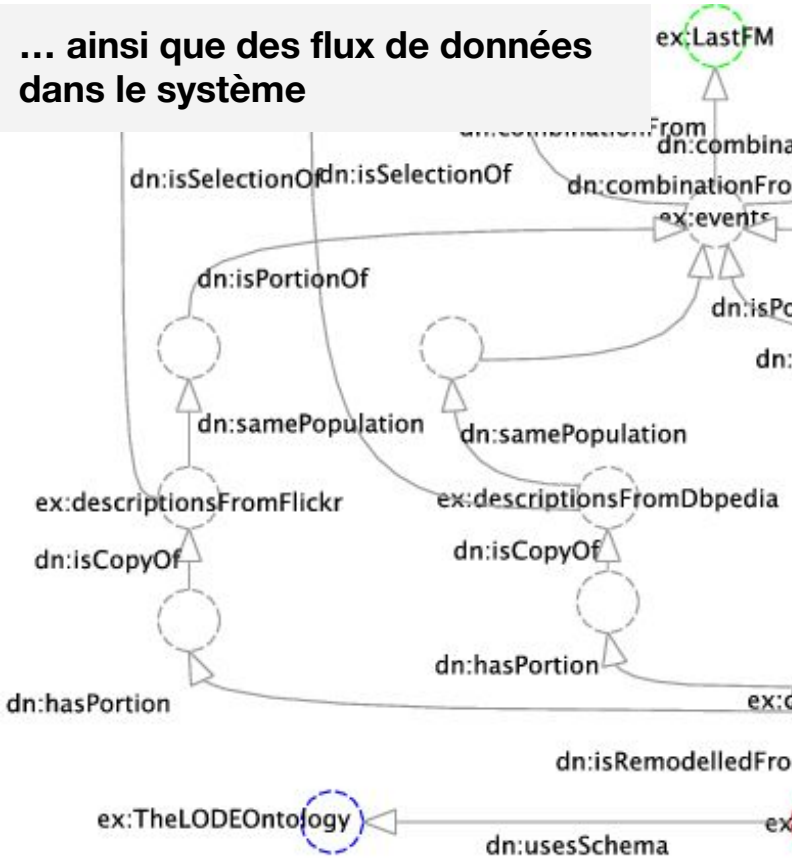
Et d'un catalogue de données semantique



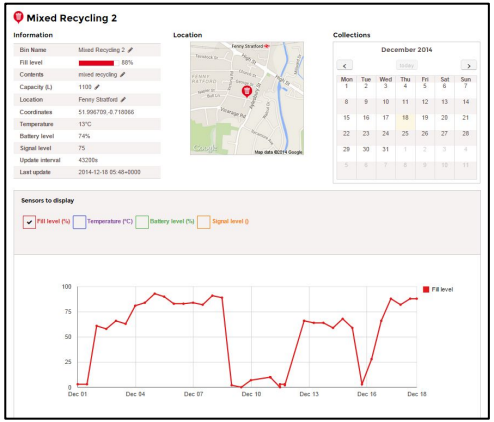
Représentation explicite et raisonnement sur les licence de données…

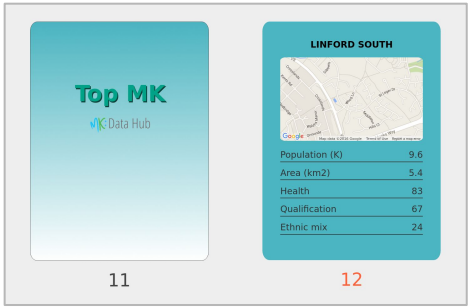
```
lic:cc-by-nc4.0 a odrl:Policy ;
 rdfs:label "CC-BY-NC" ;
 odrl:permission [
   odrl:action cc:Distribution,
ldr:extraction , ldr:reutilization ,
cc:DerivativeWorks , cc:Reproduction ;
 odrl:duty  [ odrl:action
cc:Attribution , cc:Notice ] ] ;
 odrl:prohibition [ odrl:action
cc:CommercialUse ]
```

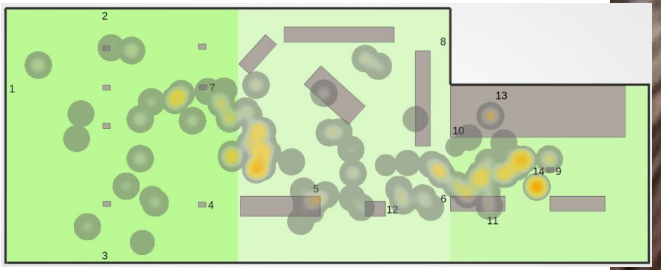
… ainsi que des flux de données dans le système

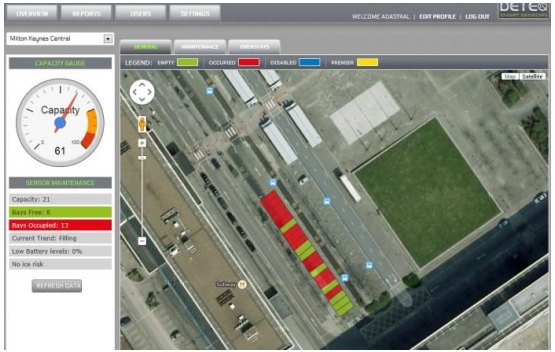


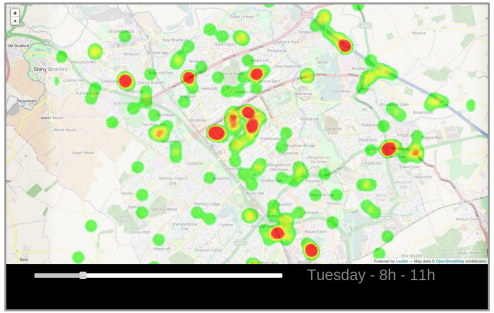
Applications

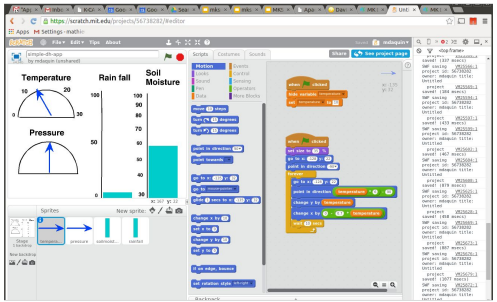




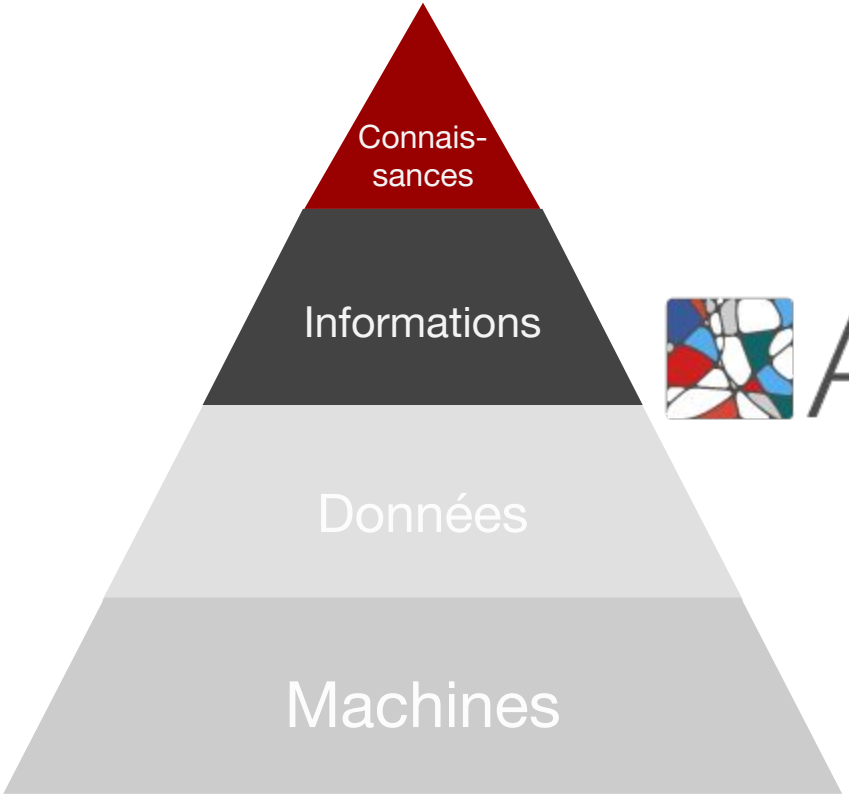




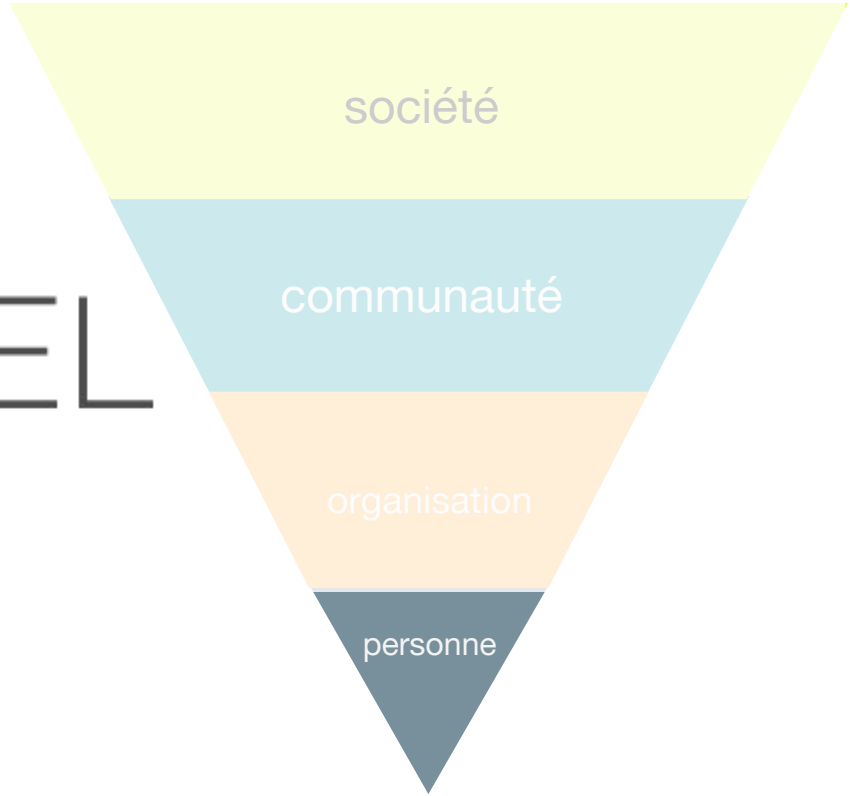




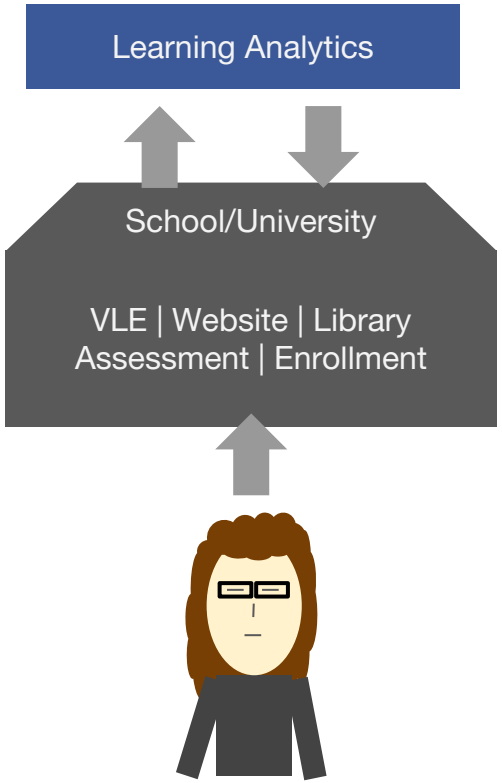
AFEL


Connais-
sances
Informations
Données
Machines

AFEL


société
communauté
organisation
personne

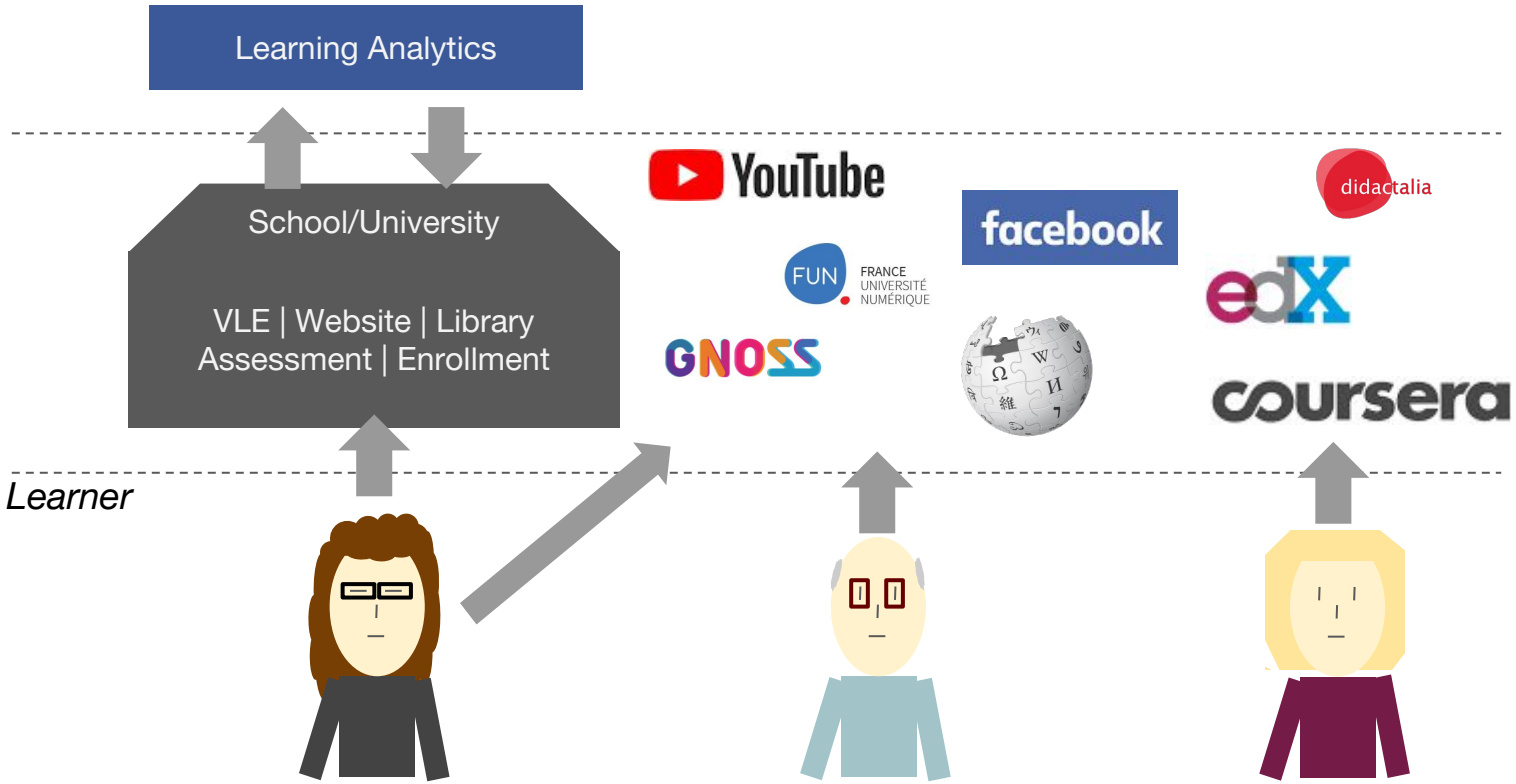
Le learning analytics (L'analyse de données d'apprentissage)



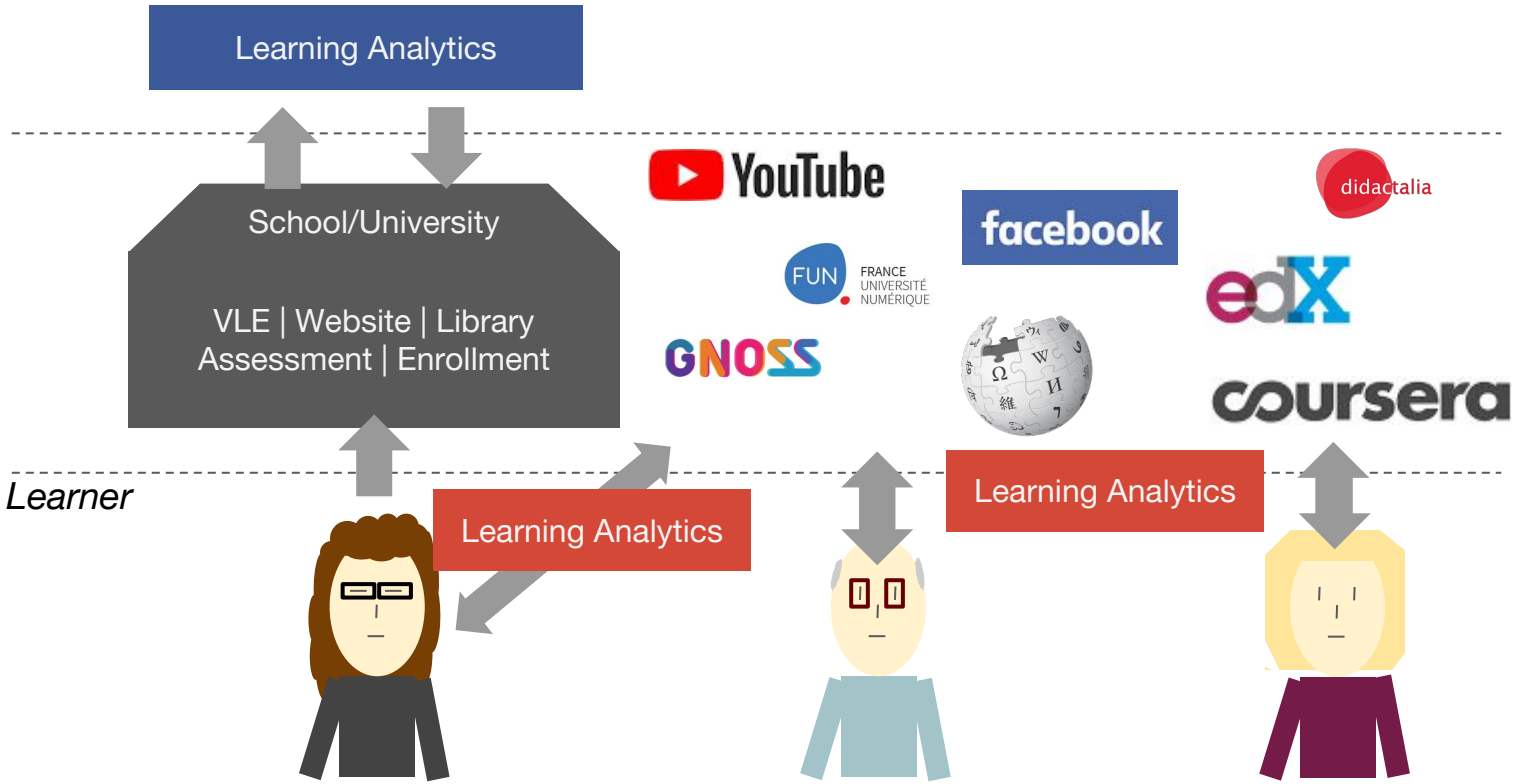
Typiquement, une université utilisent les données sur les étudiants et sur leurs activitées au travers de leur système informatique dans le but de prédire la réussite des étudiants afin de pouvoir l'améliorer....

Et si on a de la chance, aussi d'améliorer les enseignements...

Mais : Everyday learning

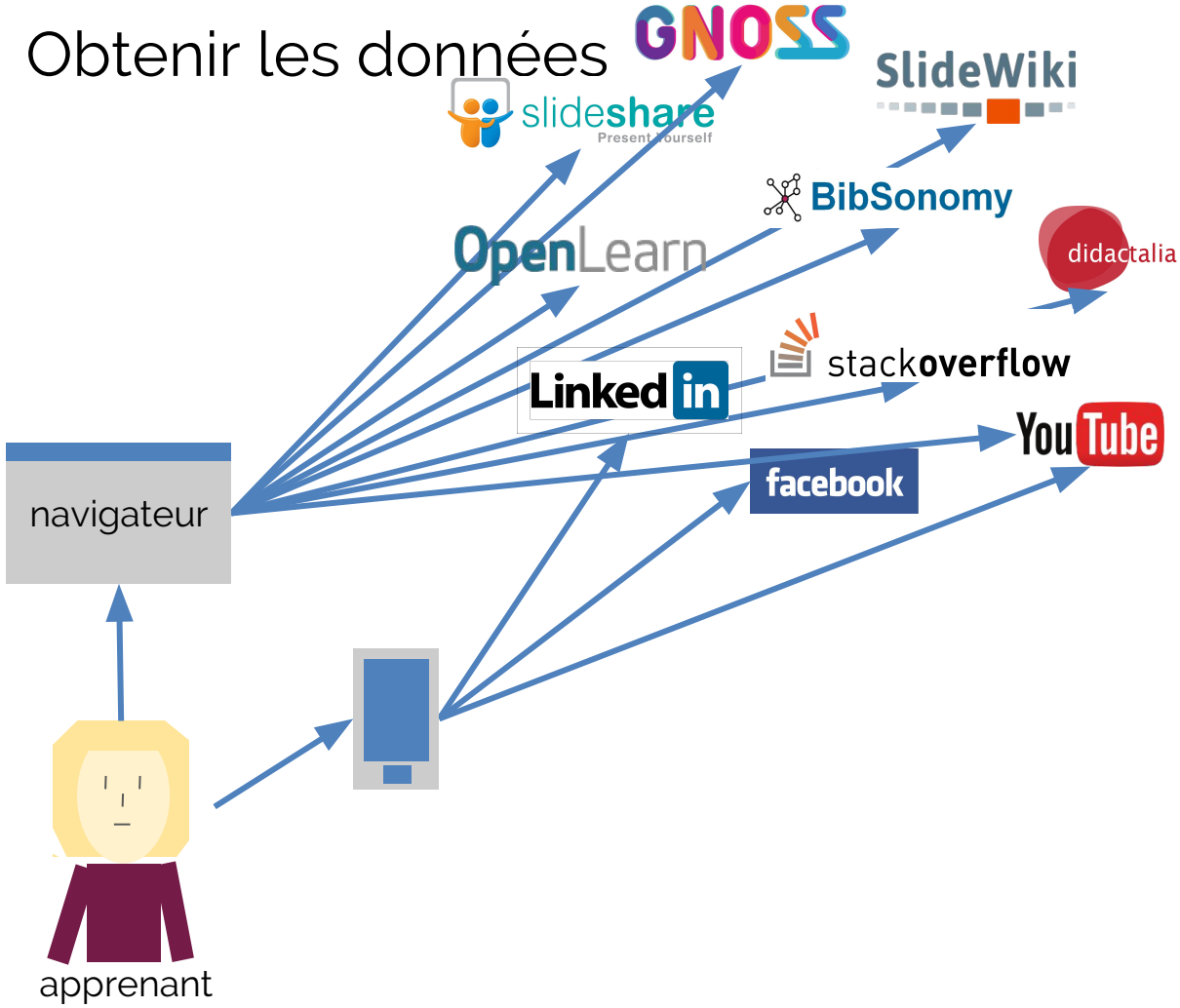


Mais : Everyday learning

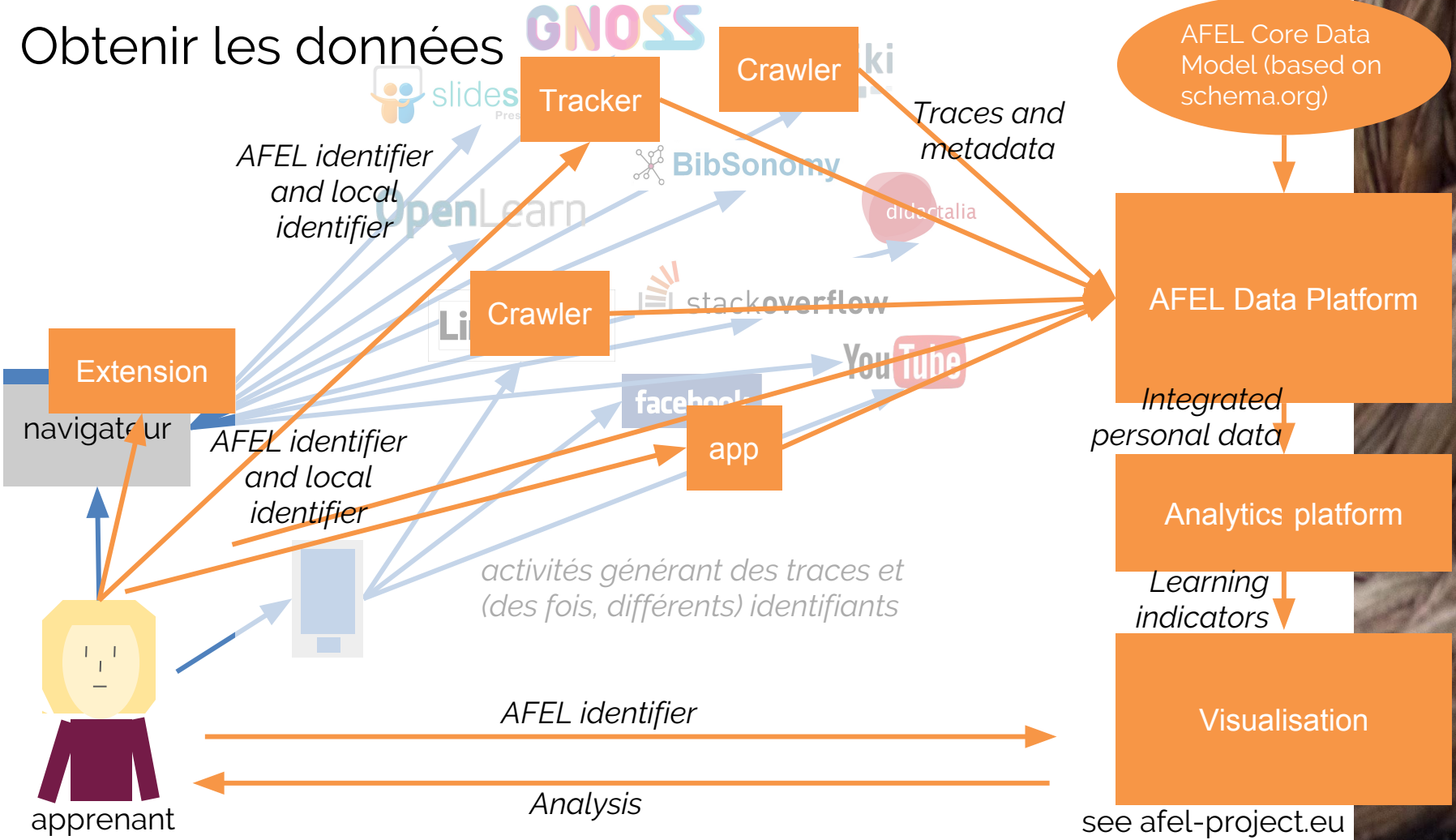


see afel-project.eu

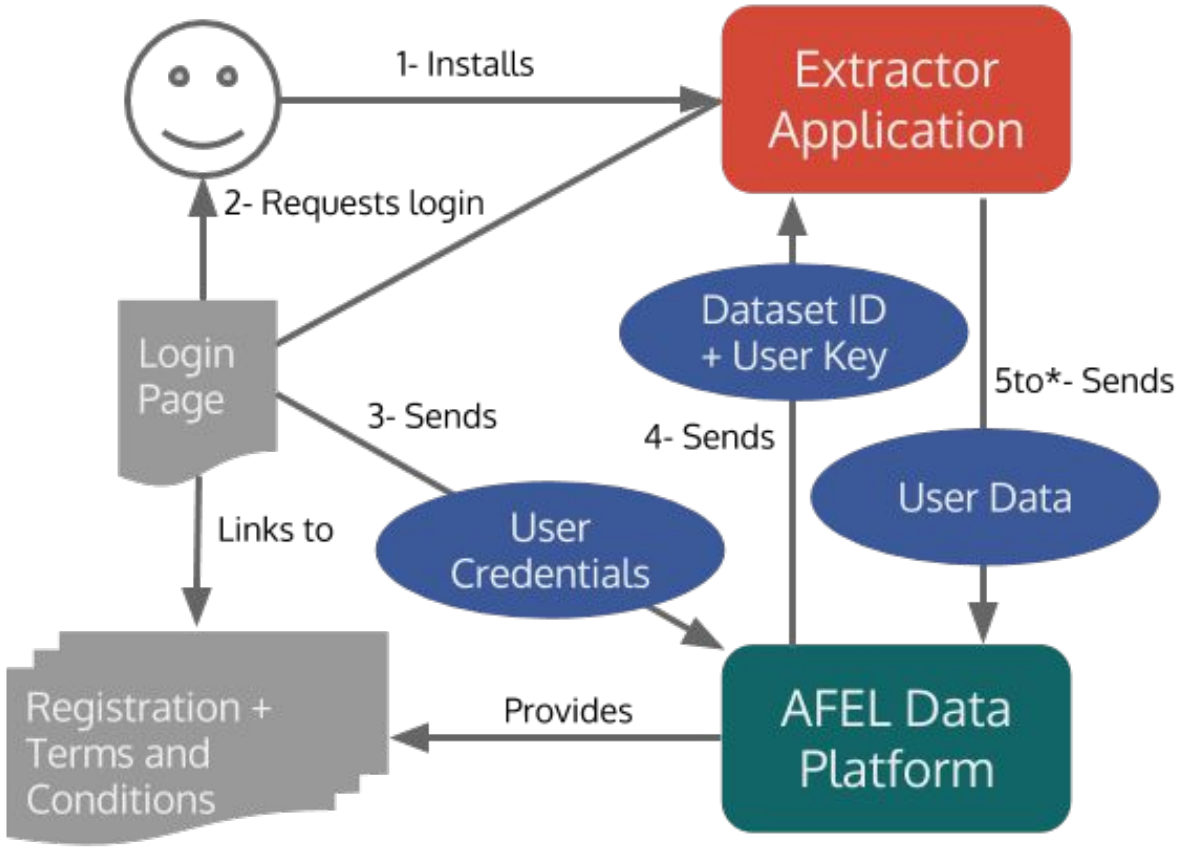
Obtenir les données



Obtenir les données

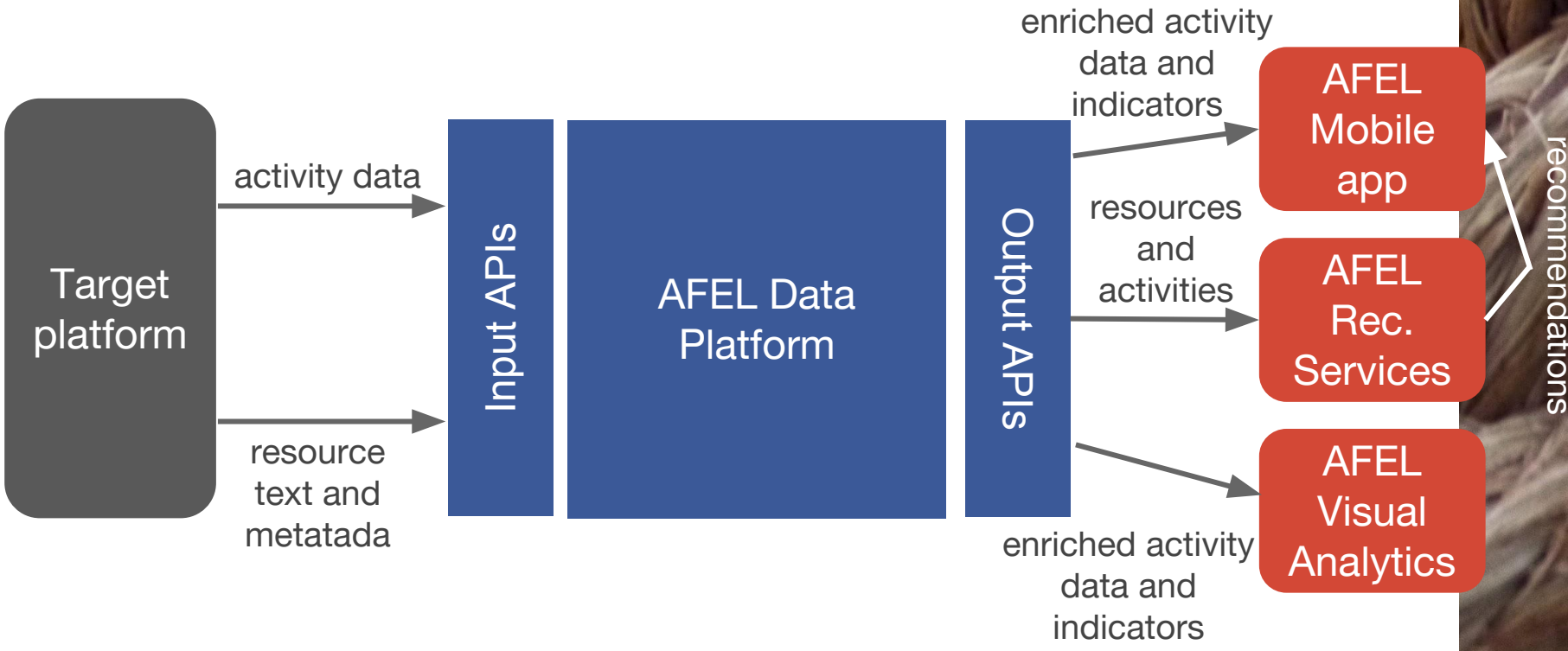


Exemple: Plugin pour navigateur



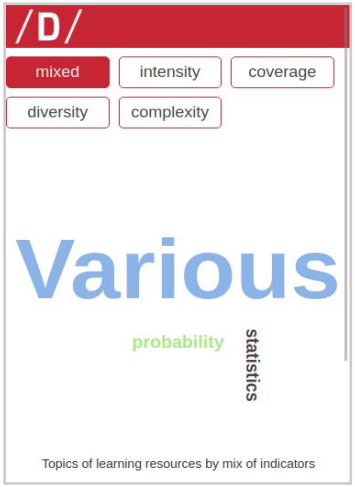
see afel-project.eu

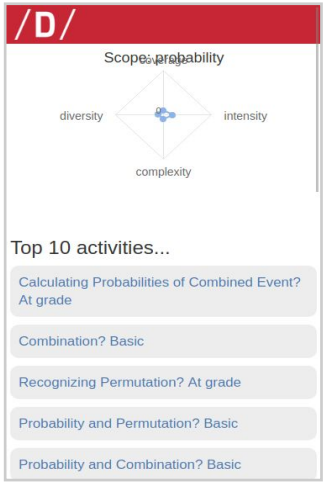
Architecture pour l'analyse d'activitées personelles

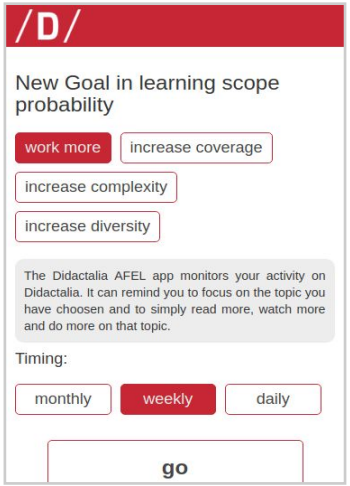


see afel-project.eu

Application







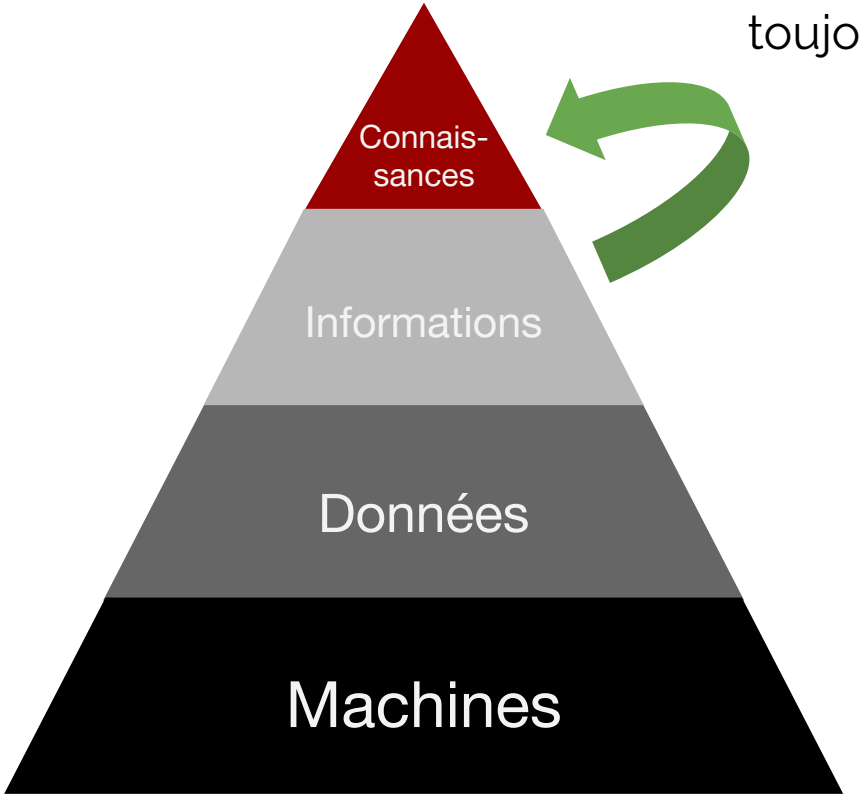
see afel-project.eu

Qu'a t'on appris ?

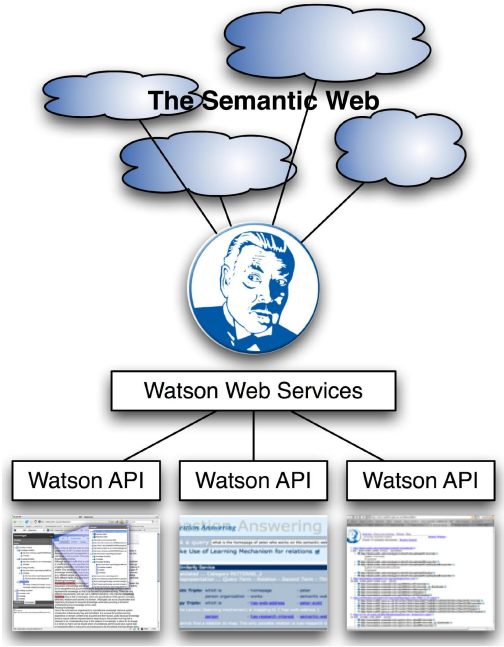
La taille des données et l'hétérogénéité sont importantes, mais ça on le savait déjà.

La diversité des données et de leurs usages est toujours un problème.

Qu'a t'on appris ?

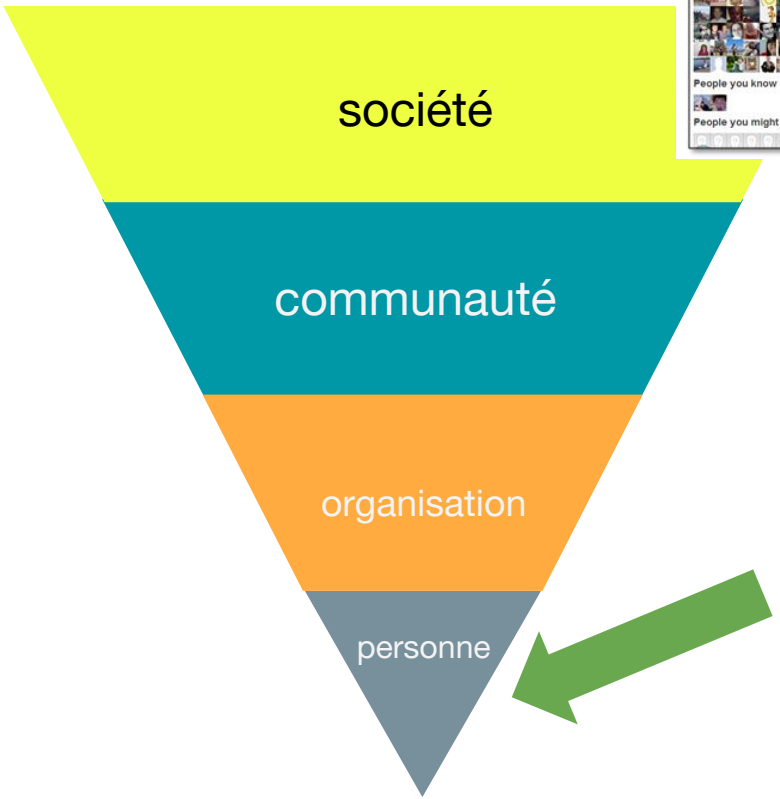


Passer de la à la est toujours difficile...



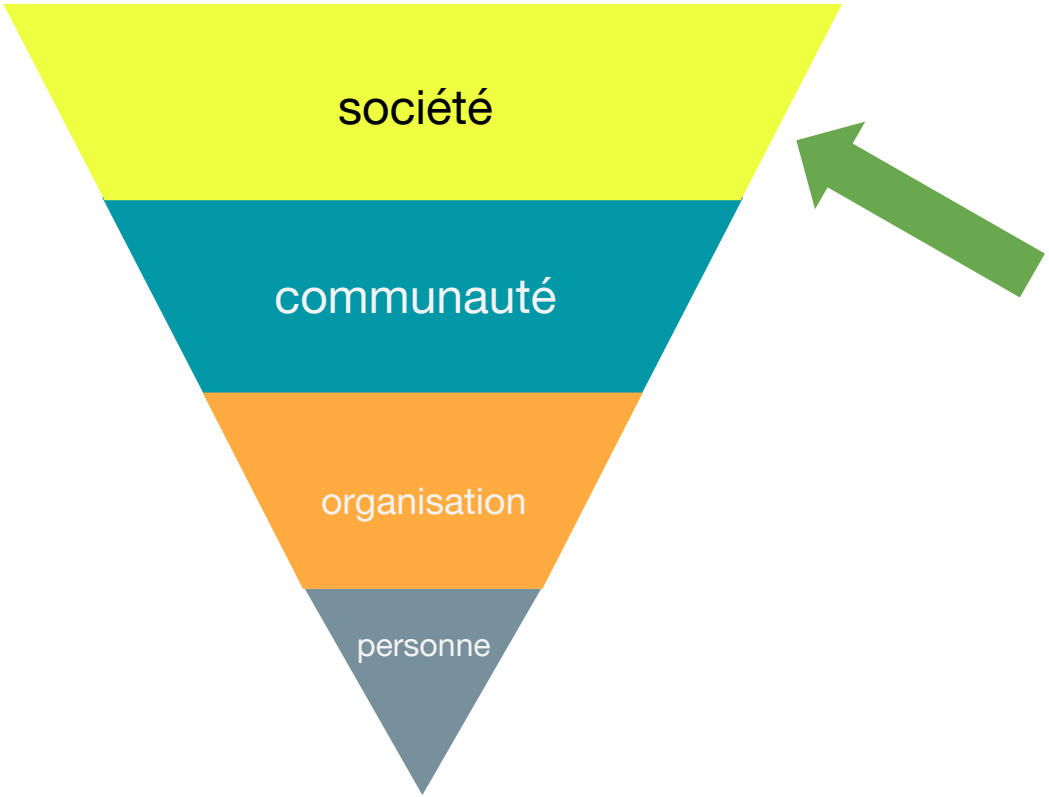
see papers from 2006 to 2009

Qu'a t'on appris ?



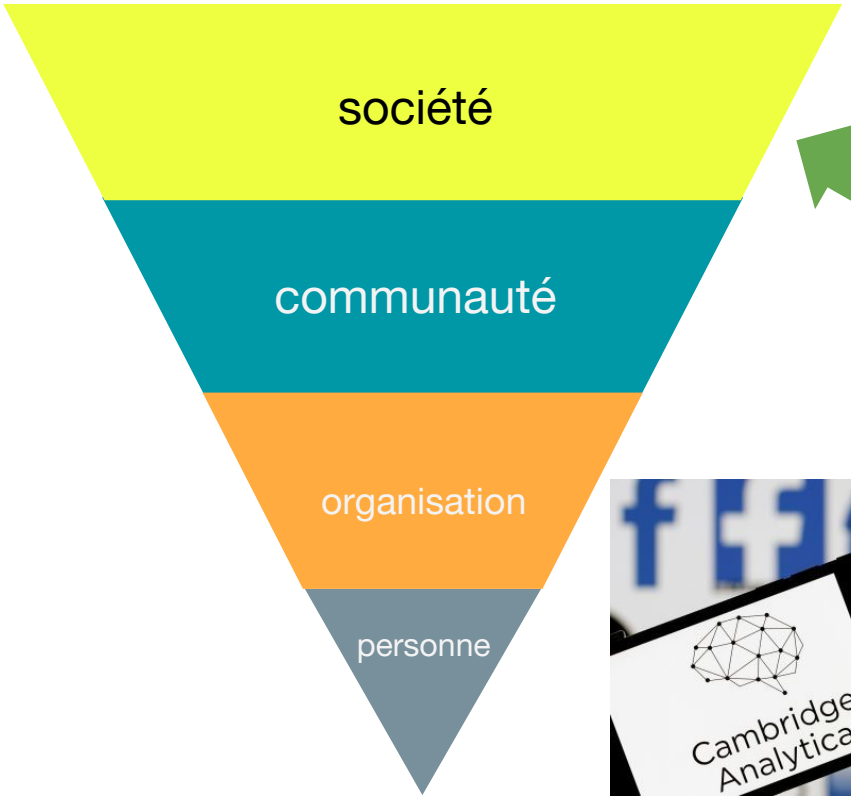
Ce niveau la est plus que jamais un problème !

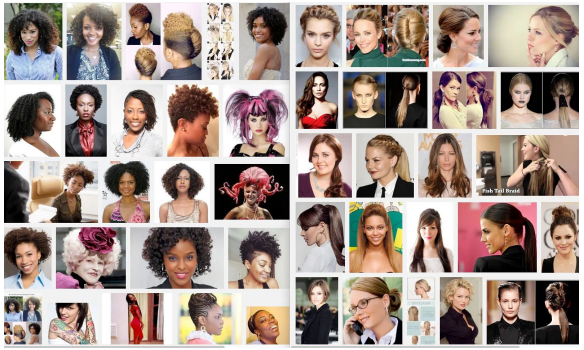
Qu'a t'on appris ?



??

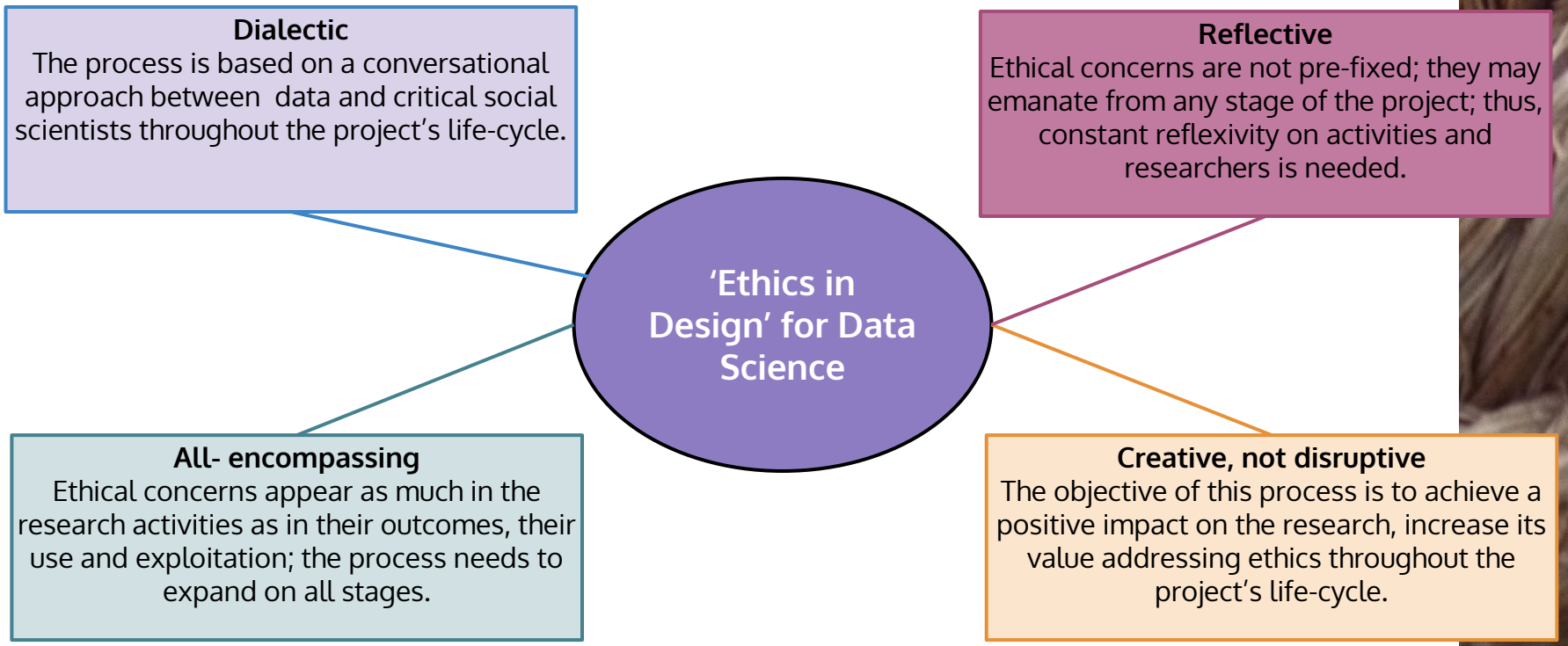
Qu'a t'on appris ?





??

Data Ethics - Comprendre les implications du traitement des données.



see paper at AIES 2018

Data Ethics - Comprendre les implications du traitement des données.

Design Fiction - utilisation de méthodes de narration pour extraire and comprendre les implications possibles des technologies, et pouvoir intégrer des méthodes préventives dans leur conception. Voir :

Workshop “Re-coding Black Mirror” à The Web Conference (WWW 2018).

‘Ethics in Design’ for Data Science

Reflective
Ethical concerns are not pre-fixed; they may emanate from any stage of the project; thus, constant reflexivity on activities and researchers is needed.

Creative, not disruptive
The objective of this process is to achieve a positive impact on the research, increase its value addressing ethics throughout the project’s life-cycle.

see paper at AIES 2018

Conclusion

On est toujours de plus en plus à la recherche des éléments d'infrastructure qui permettent de distribuer et de rendre exploitables les données.

Dans les années 1900, Nicolas Tesla investissait dans la recherche dédiée à la transmission d'électricité sans fils. Une des raisons est qu'il ne semblait pas faisable de construire l'infrastructure globale nécessaire pour amener l'électricité dans les foyers par des fils.

Concernant les infrastructures de données, peut être somme nous, de façon similaire, à la fois trop et trop peu ambitieux...

Merci

mathieu.daquin@insight-centre.org
mdaquin.net
@mdaquin