For knowledge

Jérôme Euzenat
Not a plea, a call to brains

In this talk, I want to emphasize:

- Knowledge is important
- We lost sight of this
Table of contents

Before the web
   Our civilisation is based on knowledge evolution

The semantic web
   The semantic web has tools and potential to contribute

It’s the data, stupid
   Data-centrism is effective, but stops knowledge evolution

eScience
   Semantic eScience is a direct semantic web application

Back to evolution
   We need a more organic way to evolve web knowledge
There is something else
A good idea remains a good idea because time tells it.

Juan! Not the wheel again.

Claudio, I am sure they work better by pairs.
Knowledge (what I mean by)

Different interpretations:

- True belief (philosophy=epistemological status)
- General statements (knowledge representation=form)
Knowledge (what I mean by)

Different interpretations:

- True belief (philosophy=epistemological status)
- General statements (knowledge representation=form)

<table>
<thead>
<tr>
<th>General</th>
<th>Belief</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particular</td>
<td>capitalOf(Auckland, France)</td>
<td>capitalOf(Wellington, New Zealand)</td>
</tr>
<tr>
<td>General</td>
<td>$\forall x, y, capitalOf(x, y) \rightarrow largestCityOf(x, y)$</td>
<td>$\forall x, y, capitalOf(x, y) \rightarrow locatedIn(x, y)$</td>
</tr>
</tbody>
</table>
Knowledge (what I mean by)

Different interpretations:

- **True belief** (philosophy=epistemological status)
- **General statements** (knowledge representation=form)

<table>
<thead>
<tr>
<th>General</th>
<th>Belief</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particular</td>
<td>capitalOf(Auckland, France)</td>
<td>capitalOf(Wellington, New Zealand)</td>
</tr>
<tr>
<td>General</td>
<td>$\forall x, y, \text{capitalOf}(x, y) \rightarrow \text{largestCityOf}(x, y)$</td>
<td>$\forall x, y, \text{capitalOf}(x, y) \rightarrow \text{locatedIn}(x, y)$</td>
</tr>
</tbody>
</table>
Before the web
The story of a bunch of good ideas

From: Dougal Watt, 2019
The birth of knowledge

- For looking at the future of the semantic web
- start by looking at its past,
- and the past of its past.
The birth of knowledge

- For looking at the future of the semantic web
- start by looking at its past,
- and the past of its past.

- Vegetals and animals sense their environments,
- Some animals can memorise such experiments and learn
- Some animals can imitate others, even from other species
individual

Environment
individual

Data

perceives

Environment

Society

co-operates

acts

immitates

Knowledge

learns

Data

perceives

Environment

individual
individual

Data → Knowledge

Environment

perceives

learns
individual Data learns Knowledge

Environment

Data perceives individual

Knowledge learns Data

acts

perceives
individual

Data  learns  Knowledge

perceives  acts

Environment

Knowledge  learns  Data

acts  perceives

Society  co-operates

immitates
individual

Data \rightarrow Knowledge

Knowledge \leftarrow imitates

learns

acts

perceives

acts

perceives

Environment

individual

Data \leftarrow learns
Society and culture

- When individuals interact often they form a society.
- Their interactions help shaping a *culture*.
- Culture is a shared way to deal with things.
- It is the cement of societies and enforces cooperation.
Society and culture

- When individuals interact often they form a society
- Their interactions help shaping a culture
- Culture is a shared way to deal with things
- It is the cement of societies and enforces cooperation
  ⇒ Traditions, training, co-operating, identifying.
Articulated language ⇒ Knowledge communication

- *Articulated* language allows *explicit* knowledge expression
- and communication
- Knowledge has not to relearnt from scratch
- Both language and knowledge are part of culture
Articulated language ⇒ Knowledge communication

- *Articulated* language allows *explicit* knowledge expression
- and communication
- Knowledge has not to relearnt from scratch
- Both language and knowledge are part of culture
⇒ Teaching, universities, and conferences.
Knowledge discovery and transmission

Knowledge discovery
▶ observing,
▶ experimenting,

Knowledge transmission:
▶ imitating,
▶ training,
▶ co-operating,
▶ communicating,
▶ teaching
Knowledge discovery and transmission

Knowledge discovery
▶ observing,
▶ experimenting,

Knowledge transmission:
▶ imitating,
▶ training,
▶ co-operating,
▶ communicating,
▶ teaching
Recording

- Writing
- Allows to get rid of space and time
- Improves transmission accuracy
Recording

- Writing
- Allows to get rid of space and time
- Improves transmission accuracy

⇒ Books, libraries, and journals
A good idea remains a good idea
Recording $\neq$ language

- There can be culture or language without recording (oral tradition)
- There can be culture or recording without language (animals, movies)
Summary

1. Learning
2. Communicating
3. Recording

Certainly a unique feature combination of our species, without which knowledge would have to be relearned from scratch.
Summary

1. Learning
2. Communicating
3. Recording

- Certainly a unique feature combination of our species
- Without which knowledge would have to be relearnt from scratch
The web as an amplifier

- The web pushed this to a culminating point
- Knowledge more easily and readily accessible across the planet.
- Think of stackoverflow or wikipedia
- Those who have experienced the transition can only be grateful.
The semantic web
The semantic web

- A web (also) for machines
- Actually, the web was already processed by machines
The semantic web

- A web (also) for machines
- Actually, the web was already processed by machines
- It is about content:
  - metadata
  - specialised data: agenda, addresses, bibliographic records
  - anything
Data without knowledge... A map without a legend
Data without knowledge... A map without a legend

Semantic web technologies

- Expressing data and relations on the web (RDF whatever its format)
- Expressing knowledge (RDFS, various OWL species)
- Expressing queries (SPARQL and variant)
Semantic web technologies

- Expressing data and relations on the web (RDF whatever its format)
- Expressing knowledge (RDFS, various OWL species)
- Expressing queries (SPARQL and variant)
- ...and so many other friends.
Semantic web technologies

- Expressing data and relations on the web (RDF whatever its format)
- Expressing knowledge (RDFS, various OWL species)
- Expressing queries (SPARQL and variant)
- ...and so many other friends.
- Specific ontologies: work, provenance, proteins, restaurants
Knowledge ecosystem

- This could have led to a knowledge ecosystem
- In which explicit knowledge takes its part
Knowledge ecosystem

- This could have led to a knowledge ecosystem
- In which explicit knowledge takes its part

Machines may:

- Take advantage of knowledge
- Help us elaborate knowledge (Semantic washing machine)
- Contribute knowledge
Knowledge ecosystem

- This could have led to a knowledge ecosystem
- In which explicit knowledge takes its part

Machines may:

- Take advantage of knowledge
- Help us elaborate knowledge (Semantic washing machine)
- Contribute knowledge

What went wrong?
A good idea remains a good idea.
It’s the data, stupid
What happened to the semantic web?

“If people ask you ‘what happened to the semantic web?’, you [can] point at schema.org and point at the linked open data cloud.”

–Tim Berners-Lee, ACM Turing award lecture, May 2018
That’s great!

- We were overwhelmed by the lack of data.
- We rushed to publish linked data using the available technology.
That’s great!

- We were overwhelmed by the lack of data.
- We rushed to publish linked data using the available technology.

Applications:

- Statistical data aggregation
- Semantic sensor networks
- Digital assistant, maps and question answering
- Search and knowledge graph exploitation…of course
Paul-Jacques Curie was a French physicist and professor of mineralogy at the University of Montpellier. Along with his younger brother, Pierre Curie, he studied pyroelectricity in the 1880s, leading to their discovery of some of the mechanisms behind piezoelectricity. Wikipedia
Paul-Jacques Curie was a French physicist and professor of mineralogy at the University of Montpellier. Along with his younger brother, Pierre Curie, he studied pyroelectricity in the 1880s, leading to their discovery of some of the mechanisms behind piezoelectricity. Wikipedia
I was wrong

- Beginning of the 2000s, I was not optimistic about semantic search
- and was rather advising for semantic web services
I was wrong

- Beginning of the 2000s, I was not optimistic about semantic search
- and was rather advising for semantic web services

- But a good idea remains a good idea
- and these are two good ideas
The unreasonable effectiveness of data in almost any field

- Data is the fuel for the knowledge mill
- With GPUs, it unleashes machine learning power
The unreasonable effectiveness of data in almost any field

- Data is the fuel for the knowledge mill
- With GPUs, it unleashes machine learning power
- Achieving outstanding results
  - Behaviour classification and object recognition
  - Natural language translation
  - Relation discovery (drug-illness interaction)
  - Consumer profiling and recommendation
  - Search
The unreasonable effectiveness of data in almost any field

- Data is the fuel for the knowledge mill
- With GPUs, it unleashes machine learning power
- Achieving outstanding results
  - Behaviour classification and object recognition
  - Natural language translation
  - Relation discovery (drug-illness interaction)
  - Consumer profiling and recommendation
  - Search

Does data make the web more intelligible to machines? Definitely.
“…] our fellow computer scientists can both benefit from the additional semantics and structure of the data available on the Web and contribute to building and using these structures, creating a virtuous circle.”

–Abraham Bernstein, James Hendler Natalya Noy, CACM 2016
“…] our fellow computer scientists can both benefit from the additional semantics and structure of the data available on the Web and contribute to building and using these structures, creating a virtuous circle.”

–Abraham Bernstein, James Hendler Natalya Noy, CACM 2016

There is no circle, not even vicious!
In the data realm, knowledge is misery

- Companies compete to acquire giant data meshes on which cleverly engineered engines are run to learn what is there.
- Data is eventually open, but knowledge is in the silo!
- It is not communicated: it is kept secret.
- The goal is not to share improved knowledge, but to exploit it.
- People are data providers: click, like, buy.
- Hence, knowledge does not improve.
Data revolution

▶ That’s a good name:

\[ \text{revolution} = \text{coming back to the starting point} \]

▶ No evolution

▶ Regression on the knowledge evolution ladder...

▶ Who wants to stand on the shoulders of data?
Data revolution

► That’s a good name:
  revolution = coming back to the starting point

► No evolution

► Regression on the knowledge evolution ladder...

► Who wants to stand on the shoulders of data?

► What can be done?
eScience
Knowledge in eScience

- Science is a use-case for knowledge sharing and evolution
- It is massively computer-supported
- It faces *credibility* and information *overflow* problems
- And a lot has been done in eScience
Co4: the dream of a formal collaboratory

Experiment representation

Person \(\text{writes}\) Paper

- Experiment describes research object
- Hypothesis tests
- Disk
- Object manipulates about go, obo, bio2rdf
- Software uses Data produces schema.org, dcat
- Result supports Test qualifies contradict? refute? apply?
- Subsume?
Experiment representation

Person writes Paper

Paper cite

Hypothesis tests

Manipulates disk

Uses software

Data produces schema.org, dcat

Result supports Test

Qualifies CONTRADICT?

Subsume?

Refute?

Support?

Apply?
Experiment representation

Person writes Paper

Experiment describes Paper

Paper cite

Data produces schema.org, dcat

Result supports Test

Hypothesis tests

Experiment alters research object

Object manipulates disk, go, obo, bio2rdf

Software uses

Experiment represents

Experiment representation

Person writes Paper

Experiment describes Paper

Experiment alters ResearchObject

Hypothesis tests Disk

Object manipulates about go, obo, bio2rdf

Software uses Data

Result produces schema.org, dcat

Test qualifies

contradict? subsume? refute? support? apply?
Experiment representation

Hypothesis tests Experiment
Person writes Paper

Experiment alters

Experiment describes

Paper cite

Paper

Result supports Test

Qualifies contradict?
refute?
support?
apply?
Experiment representation

Hypothesis \(\xrightarrow{\text{tests}}\) Experiment \(\xleftarrow{\text{alters}}\) Paper

Person \(\xrightarrow{\text{writes}}\) Paper

Object

- about
- manipulates

Test

- supports
- qualifies
- contradicts
- subsumes
- refutes
- supports

33
Experiment representation

Hypothesis \(\xleftarrow{\text{about}}\) Object \(\xrightarrow{\text{manipulates}}\) Experiment \(\xrightarrow{\text{tests}}\) Paper \(\xrightarrow{\text{writes}}\) Person

Software \(\xrightarrow{\text{uses}}\) Experiment \(\xleftarrow{\text{alters}}\)

Result \(\xrightarrow{\text{supports}}\) Test \(\xrightarrow{\text{qualifies}}\) Contradict? Subsume? Refute? Support? Apply?

Data \(\xrightarrow{\text{produces}}\) schema.org, dcat

So/f_tware \(\xrightarrow{\text{uses}}\) Software

Experiment describes

Research object

Method/Protocol (go, obo, bio2rdf)
Experiment representation

Hypothesis about Object

Person writes Paper

Experiment describes Software

Data produces Software

Object manipulates Software

Result supports Test

contradict? supports? refute? apply?

Hypothesis tests Experiment

Object about Hypothesis

Person writes Paper

Software uses Experiment

Software alters Experiment

Software describes Experiment

Software cite Experiment

Object manipulates Experiment

Software uses Experiment

Software produces Data
Experiment representation

Hypothesis ← Experiment supports Result ← Data

Object manipulates alters Software uses

tests

Person writes Paper cites

Data produces schema.org, dcat

Result supports

Test qualifies contradict? subsume? refute? support? apply?

Software alters

Research object: go, obo, bio2rdf

Data produces Software

Hypothesis tests Experiment

Result supports

Experiment manipulates Object

Hypothesis about Person writes

Hypothesis writes Person
Experiment representation

Test qualifies Result supports Data

about Object manipulates

tests Experiment

tests

Hypothesis Experiment

Person Paper

Software

uses

alters
describes

cite

produces

Data produces

Data produces

Data produces
Experiment representation

Test \(\rightarrow\) Result \(\leftarrow\) Data

Hypothesis \(\leftarrow\) Experiment \(\rightarrow\) Paper

Person \(\rightarrow\) Paper

Data produces schema.org, dcat

Software uses

Experiment describes researchobject

Result supports Data

Test qualifies Result

Object manipulates about go, obo, bio2rdf

Person writes Paper

Hypothesis tests Experiment

Disk

So/f_tware uses

Schema.org
Experiment representation

Test \rightarrow \text{Result} \leftarrow \text{supports} \rightarrow \text{Data}

\text{Hypothesis} \leftarrow \text{tests} \rightarrow \text{Experiment}

Person \rightarrow \text{Paper}

Experiment \text{describes} \rightarrow \text{researchobject}

Object \text{manipulates} \rightarrow \text{go, obo, bio2rdf}

Data \text{produces} \rightarrow \text{schema.org, dcat}

Software \text{uses} \rightarrow \text{Experiment}

\text{contradict?} \rightarrow \text{Paper}
Experiment representation

Test → Result ← supports Data

Hypothesis ← Experiment

Object → Experiment

Software → Experiment

Person writes Paper

Test qualifies Result

Data produces schema.org, dcat

Software uses Experiment

Hypothesis subsume?

Experiment refutes?

Object manipulates about go, obo, bio2rdf

Experiment alters researchobject

Result supports Test

contradict? refute? supports? qualifies? holds?

Paper cite cito

Person foaf

Data produces Data

Result describes Software

Experiment writes Paper

Person writes Paper
Experiment representation

Test \(\xrightarrow{\text{qualifies}}\) Result \(\xleftarrow{\text{supports}}\) Data

Hypothesis \(\overset{\text{support?}}{\xleftarrow{\text{about}}}\) Object \(\overset{\text{manipulates}}{\xrightarrow{\text{tests}}}\) Experiment

Person \(\xrightarrow{\text{writes}}\) Paper

Software

Data produces schema.org, dcat

Experiment alters researchobject

Result supports Test

Test qualifies Result

Object uses Software

Data

Result supports Data

Hypothesis subsume?

Hypothesis refute?

Disk

go, obo, bio2rdf

tests

Object

Person

Paper

33
Experiment representation

All these applications need knowledge

- May be used at design time, or post hoc
- May be used to probe experiment preregistration
- Far more complex than that (frbr)
Experiment representation

All these applications need knowledge

- May be used at design time, or post hoc
- May be used to probe experiment preregistration
- Far more complex than that (frbr)
- However, only applies proven technologies
Experiment representation

All these applications need knowledge

- May be used at design time, or post hoc
- May be used to probe experiment preregistration
- Far more complex than that (frbr)
- However, only applies proven technologies
- May be applied to other fields, e.g. mathematics
Experiment representation

All these applications need knowledge

- May be used at design time, or post hoc
- May be used to probe experiment preregistration
- Far more complex than that (frbr)
- However, only applies proven technologies
- May be applied to other fields, e.g. mathematics
- Here as well, machine learning could help cleaning and mining data!
What are the incentives for us to log experiments?

We know how to do them already.
Back to evolution
Knowledge has to evolve

- Knowledge *has* to evolve
- Not only scientific knowledge
- Not as scientific knowledge
Knowledge has to evolve

- Knowledge *has* to evolve
- Not only scientific knowledge
- Not as scientific knowledge

- It has to be done seamlessly
- while being used
Cultural evolution

- Natural evolution applied to culture
- e.g. boat design, language, social structure
- Evolutionary epistemology (applied to science)

“a fast lane of cultural evolution bypassing the traffic jams of genetic evolution”
– Yuval Noah Harari, Sapiens: a brief history of humankind, 2011
Cultural evolution

- Natural evolution applied to culture
- e.g. boat design, language, social structure
- Evolutionary epistemology (applied to science)

“a fast lane of cultural evolution bypassing the traffic jams of genetic evolution”

–Yuval Noah Harari, Sapiens: a brief history of humankind, 2011
Natural selection

An abstract control mechanism based on:

- Variation
- Selection
- Transmission
Natural selection

An abstract control mechanism based on:

- Variation
- Selection
- Transmission

Computer implementation:

- Genetic programming
- Experimental cultural evolution
Experimental cultural knowledge evolution

- Populations of agents bearing knowledge
- Using it for interacting with each others and their environment (games)
- Adapting it at the issue of each interaction
Experimental cultural knowledge evolution

- Populations of agents bearing knowledge
- Using it for interacting with each others and their environment (games)
- Adapting it at the issue of each interaction
Experimental cultural knowledge evolution

- Populations of agents bearing knowledge
- Using it for interacting with each other and their environment (games)
- Adapting it at the issue of each interaction

Used for:
- Abstract culture
- Natural language evolution
- Ontology alignment evolution
Application: semantic search engine

- A search engine uses knowledge to answer questions
- It takes feedback as usual from users
- Uses knowledge adaptation operators at the issue of each interaction
- May be applied to a (interacting) population of such engines
What has to be done

- Understanding general principles of knowledge evolution
- Articulation with the semantic web
What has to be done

- Understanding general principles of knowledge evolution
  - Theoretically
  - Experimentally
  - In relation with social sciences
- Articulation with the semantic web
What has to be done

- Understanding general principles of knowledge evolution
  - Theoretically
  - Experimentally
  - In relation with social sciences
- Articulation with the semantic web
  - Adaptation (selection)
  - Read-write web (transmission)
  - Non-breaking interoperability
  - We still do not have a semantic 404
Conclusion
Where are we?

TRANSMISSION
DISCOVERY

SW IS HERE
The semantic web and its technologies may:

- Serve filling silos to the benefits of commercial interests
- Help contributing to the knowledge of the humanity

Both may be pursued.
A good idea remains a good idea

What is the business model for these... books?

Selling glasses obviously
Knowledge (what I mean by)

**Different interpretations:**
- True belief (philosophy=epistemological status)
- General statements (knowledge representation=form)

<table>
<thead>
<tr>
<th>Data</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>capitalOf(Auckland, France)</td>
<td>capitalOf(Wellington, New Zealand)</td>
</tr>
</tbody>
</table>

∀ x, y, capitalOf(x, y) → largestCityOf(x, y)

∀ x, y, capitalOf(x, y) → locatedIn(x, y)

The unreasonable effectiveness of data in almost any field
- Data is the fuel for the knowledge mill
- With GPUs, it unleashes machine learning power
- Achieving outstanding results
- Natural language translation
- Predicting outcomes (long-term interaction)
- Consumer profiling and recommendation
- Search

Does data make the web more intelligible to machines? Definitely.

A good idea remains a good idea

A scientific revolution!

The unreasonable effectiveness of data in almost any field

<table>
<thead>
<tr>
<th>Data</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>Result</td>
</tr>
</tbody>
</table>

Experiment representation

<table>
<thead>
<tr>
<th>Experiment representation</th>
<th>Test</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis</td>
<td>Accept</td>
<td></td>
</tr>
</tbody>
</table>

What has to be done
- Understanding general principles of knowledge evolution
  - Theoretically
  - Experimentally
- In relation with social sciences
- Articulation with the semantic web
  - Adaptation (not text)
  - Read/write web (transmission)
  - Non-breaking interoperability
- We still do not have a semantic 404

Not everyone agrees

“The illiterate of the 21st century will not be those who cannot read and write, but those who cannot learn, unlearn, and relearn”

Alvin Toffler
Knowledge graphs

- I have been a computer scientist for too long now.
- But I have always worked on graphs and knowledge.
- This is so 60’s