



# Semi-Supervised Structuring of Complex Data

**Marian-Andrei RIZOIU**

ERIC Laboratory, University Lumière, Lyon, France

*IJCAI 2013*

*August 5<sup>th</sup>, 2013*

Beijing, China

# The context of my PhD work - complex data

**France** (structure)

From Wikipedia, the free encyclopedia

*This article is about the country. For other uses, see [France \(disambiguation\)](#).*

See also: [France portal](#) and [Outline of France](#)

**France** (English <sup>i</sup>/ˈfræns/ **FRANS** or <sup>i</sup>/ˈfrɑːns/ **FRAHNSS**; French: [fʁɑ̃s] <sup>i</sup>listen), officially the **French Republic** (French: *République française* French pronunciation: [ʁepyblik fʁɑ̃sɛz]), is a unitary semi-presidential republic located mostly in Western Europe,<sup>[note 12]</sup> with several overseas regions and territories. Metropolitan France extends from the Mediterranean Sea to the English Channel and the North Sea, and from the Rhine to the Atlantic Ocean. From its shape, it is often referred to in French as *l'Hexagone* ("The Hexagon").

France is the largest country in Western Europe and the third-largest in Europe as a whole. It possesses the second-largest **exclusive economic zone** in the world. France has been a **major power** with strong **cultural**, **economic**, **military**, and **political** influence in Europe and around the world.<sup>[6]</sup> France has its main ideals expressed in the 18th-century *Declaration of the Rights of Man and of the Citizen*. From the 17th to the early 20th century, France built the **second-largest colonial empire** of the time, ruling large portions of first **North America** and **India** and then **Northwest** and **Central Africa**; **Madagascar**; **Indochina** and southeast **China**; and many **Caribbean** and **Pacific Islands**.

France is a developed country,<sup>[7]</sup> possessing the world's **fifth-largest** and Europe's **second-largest** economy by **nominal GDP**. It is also the world's **ninth-largest** by **GDP at purchasing power parity**.<sup>[8]</sup> France is the wealthiest nation in Europe – and the fourth-wealthiest in the world – in aggregate household wealth.<sup>[9]</sup> French citizens enjoy a high **standard of living**, high **public education level**, and one of the world's longest **life expectancies**.<sup>[10]</sup> France has been listed as the world's "best overall health care" provider by the **World Health Organization**.<sup>[11]</sup> It is the most-visited country in the world, receiving 79.5 million foreign tourists annually.<sup>[12]</sup>

France has the world's **fifth-largest nominal military budget**,<sup>[13]</sup> as well as (in terms of personnel) the largest military in the **EU**,<sup>[citation needed]</sup> the third-largest deployable force in **NATO**, and the **26th-largest** military in the world. France also possesses the third **largest stockpile of nuclear weapons** in the world<sup>[14]</sup> – with around 300 active warheads as of 25 May 2010 – and the **world's second-largest diplomatic corps** (behind the **United States**).<sup>[15]</sup> France is a founding member of the **United Nations**, one of the five permanent members of the **UN Security Council**, and a member of the **Francophonie**, the **G8**, **G20**, **NATO**, **OEC**, **WTO**, and the **Latin Union**. It is also a founding and leading **member state of the European Union** and the largest EU state by area.<sup>[16]</sup> In 2013, France was listed 20th on the **Human Development Index** and, in 2010, 24th on the **Corruption Perceptions Index**.

**French Republic**  
*République française*

Flag

Motto:  
"Liberté, égalité, fraternité"  
"Liberty, Equality, Fraternity"

0:00 CC ◀ ▶ MENU

Map (image)

Location of Metropolitan France (dark green)  
– in Europe (green & dark grey)  
– in the European Union (green) — [Legend]

**Area**

- Total <sup>[note 2]</sup>	674,843 km <sup>2</sup> (41st) 260,558 sq mi
- Metropolitan France	
- IGN <sup>[note 3]</sup>	551,695 km <sup>2</sup> (47th) 213,010 sq mi
- Cadastre <sup>[note 4]</sup>	543,965 km <sup>2</sup> (47th) 210,026 sq mi

**Population**  
(2012)

- Total <sup>[note 2]</sup>	65,350,000 <sup>[2]</sup> (19th)
- Metropolitan France	63,460,000 <sup>[1]</sup> (22nd)
- Density <sup>[note 5]</sup>	116/km <sup>2</sup> (89th) 301/sq mi

Indicators (numerical format)

**Specificities:**

- ➔ Different types of data
- ➔ Additional information
- ➔ Temporal dimension / dynamic data
- ➔ High dimensionality
- ➔ Diverse / distributed sources

Description (text)

Hyperlinks (external knowledge sources)

# My approach

**General objective:** extract knowledge from complex data, often in an unsupervised context  
add semantics in data analysis  
leverage available side-information (supervision)

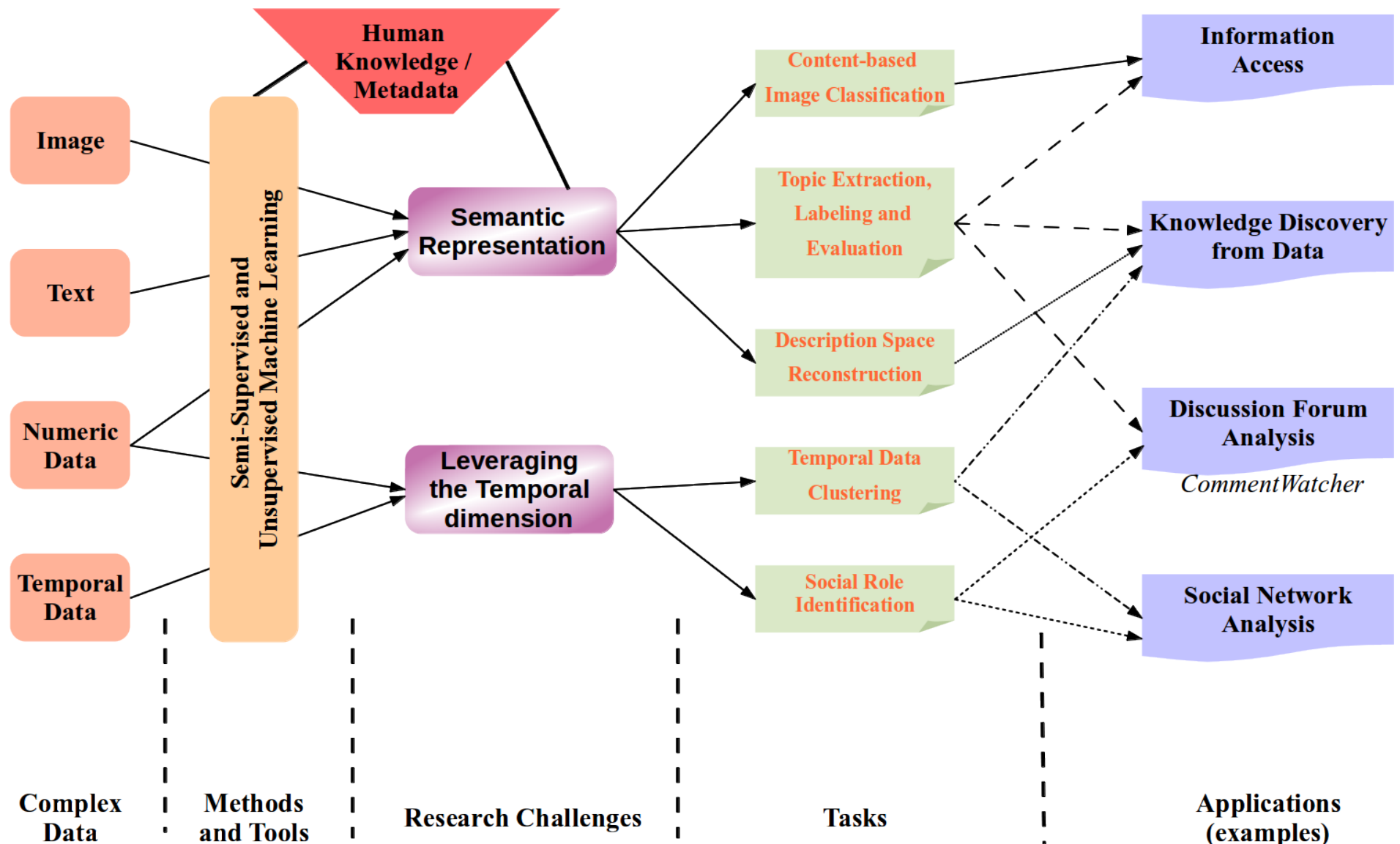
## Research challenges

Leveraging semantics when dealing with complex data

- translating data into a semantic-aware representation space
- injecting knowledge into machine learning algorithms

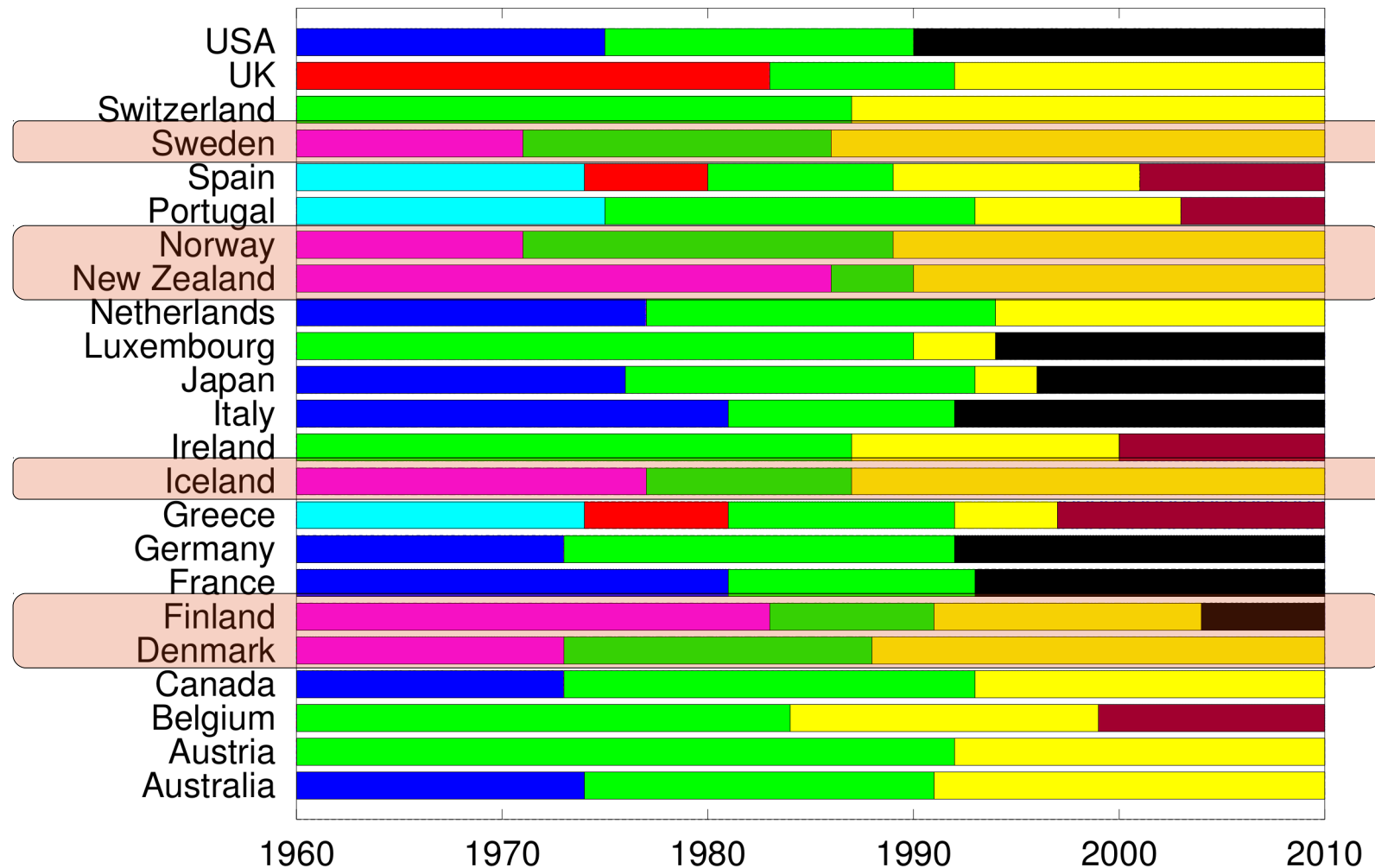
Leveraging the temporal dimension of complex data

# The schema structuring my work



# Applications:

**Detecting temporal evolution patterns** in a population of entities recorded over a period of time



# Applications:

**Construct interpretable features**, which capture better the semantics of the dataset and reduce feature correlation

$\{ \text{groups}, \text{sky}, \text{tree}, \text{building}, \text{street} \}$



$\{ \text{sky} \wedge \text{building} \wedge \text{tree} \wedge \overline{\text{forest}}, \text{sky} \wedge \text{groups} \wedge \text{street} \}$

$\{ \text{groups}, \text{water}, \text{cascade}, \text{sky}, \text{tree}, \text{lawn}, \text{forest} \}$

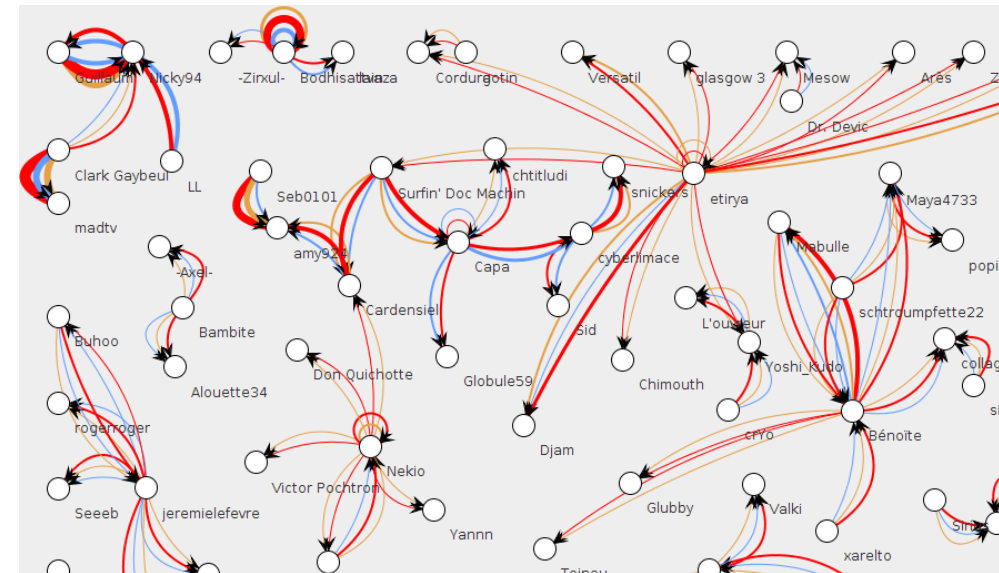


$\{ \text{groups} \wedge \overline{\text{street}} \wedge \overline{\text{interior}}, \text{water} \wedge \text{cascade} \wedge \text{tree} \wedge \text{forest}, \text{sky} \wedge \overline{\text{building}} \wedge \overline{\text{panorama}} \}$



# Applications:

**CommentWatcher**, an open-source web-based platform for analyzing discussions on web forums



- ➔ Expression cloud for each topic
- ➔ Temporal evolution of topics
- ➔ Evolution of the popularity of a topic

- ➔ Social Network modeled as a multi-graph
- ➔ **Vertexes:** the users ; **Arcs:** the comments associated with topics
- ➔ Based on the structural citation relation

**Thank you !**