PoliWeb project (PEPS'14)

Towards an infrastructure for collecting, storing and analyzing data flows produced by politicians

Javier Espinosa

LAFMIA, UMI CNRS Mexico

javiera.espinosa@gmail.com

Journée Cloud et Multi-Cloud, 27 Mars 2014, Université Lyon III

Geraldine Castel
CEMRA, Université Stendhal, France

Anne Marie Benoit
PACTE, CNRS, France

Genoveva Vargas-Solar
CNRS, LIG-LAFMIA, France

3 120 268

C'est le nombre total de conversations analysées sur le Web social depuis le 1er janvier 2012 à propos des élections présidentielles francaises !
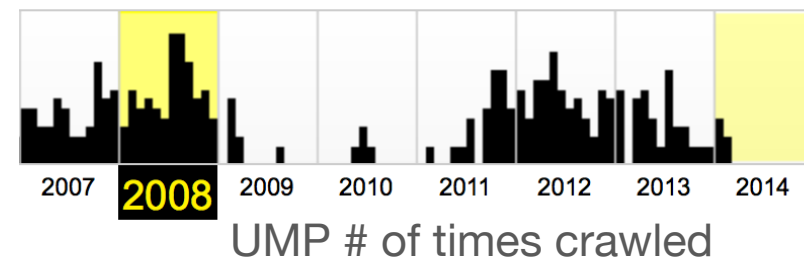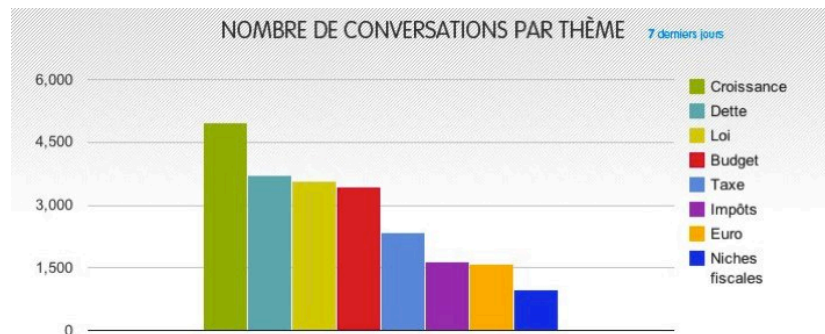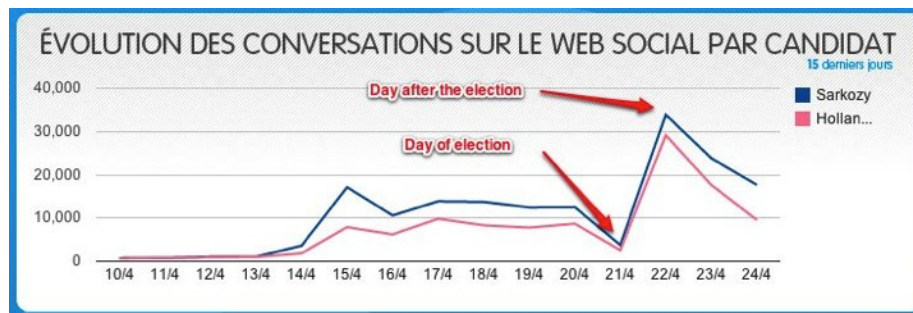
Have you been tweeting about the elections?

Doing politics in the era of big data

- 2008 was called the **"social media election"** with **1.8 million tweets** sent on election day.

- Barack Obama's appearance at the Democratic National Convention caused **4 million tweets** total during his 39 minute speech (**52,000 tweets / minute**).

- **9/10 Senators** and **Representatives** have their own Twitter accounts.

2012 in France

ÉVOLUTION DES CONVERSATIONS SUR LE WEB SOCIAL PAR CANDIDAT

LES MOTS CLÉS DE LA CAMPAGNE

NOMBRE DE CONVERSATIONS PAR THÈME

UMP # of times crawled

Sources: blogs.salesforce.com, web.archive.org

Social networks and IT in politics → Volume of data → big data → Analytics

Juridical issues

How is all of this **online political activity affecting the elections** ?

Objective

" Compare the impact of the use of technologic tools in the campaign strategies of the Europeans elections in France and UK"

- Implications
  - **Collect data** from social networks, websites and politicians' blog
  - **Curate** and **store** these data
  - Define a **continuous comparison process** that can evolve during time and as new information is integrated in the database

Challenges

- **Time and ownership**
  - Data of interest is determined by the **campaign period** (EU elections) which is short and it changes by party and country and even candidate
  - According to juridical laws in both countries access, exploitation and storage of data can be **limited** or **partially limited**

- **Data curation and storage**
  - Organization according to political and geographical organization
  - Fill-in missing information and unbalanced content retrieved about entities that must be compared

- **Provenance and pertinence**

Expected Results

- **Integrated historical and distributed database** of documents, photos, text and social networks posts
  - Data provenance, freshness
  - Structure
  - Respecting privacy and data ownership, owner anonymity

- **Analysis platform** for querying the database with respect to different criteria:
  - Geographic and temporal parameters
  - Statistics
  - Political organization and tendencies

- Compare strategies and conditions of the elections in UK and France

Roadmap

✓ Doing politics in the era of Big Data

- Analyzing political campaign strategies in Europe
  - Data collection and curation
  - Comparing for understanding strategies: UK vs. France

- Conclusions and perspectives

Data Collection and Curation
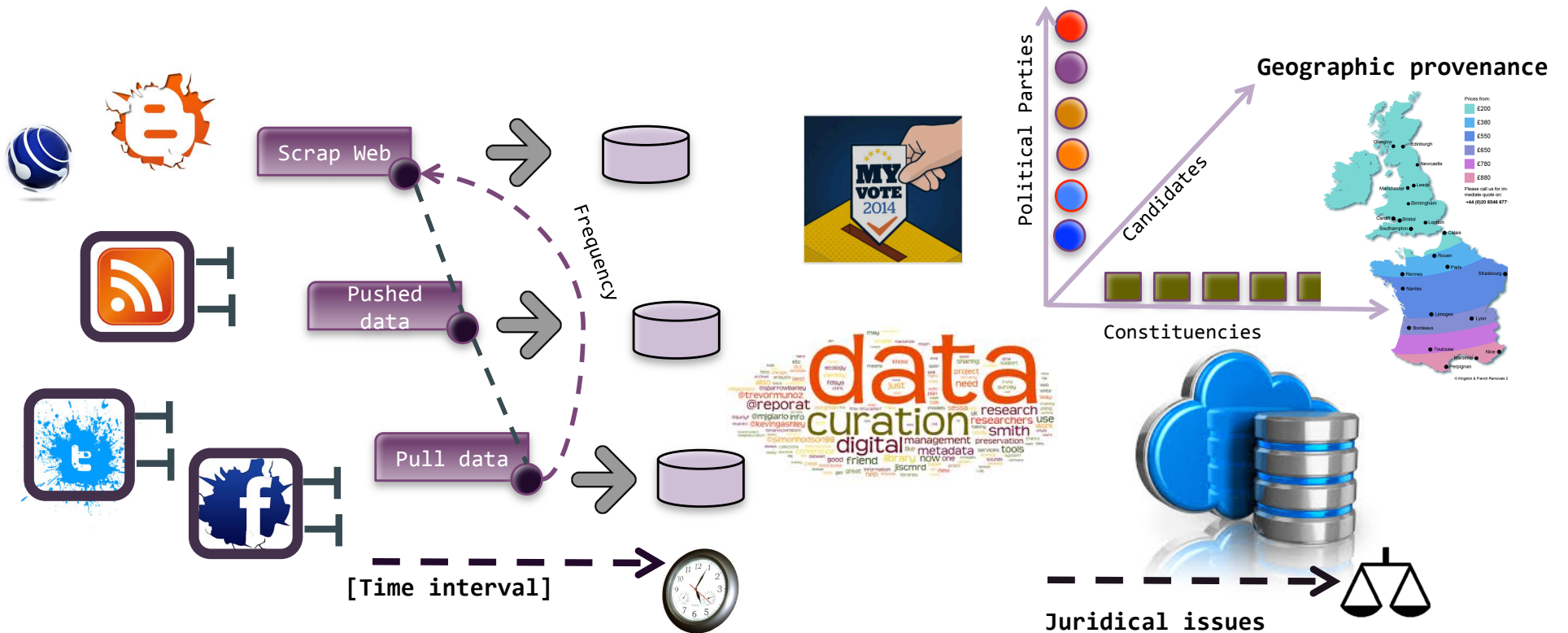
Comparison and Analysis Requirements (i)

- **Query criteria**

  - Date, candidate, party, document type, key words (frequent words/term clouds)

- **Data provenance**

  - Party, webmaster, candidate, campaign staff

- Generate an **inventory geo-localized** and **grouped by parties and militants**

  - Content types: video, text, image, document

  - Links to other content and tools: donations on line, other campaign actions, Facebook pages and support committees, agenda

Comparison and Analysis Requirements (ii)

- **Compare content** from *sites*, *personal blogs* and *pages*, *parties sites*
  - Common and different elements: content and structure (communication strategies)
  - Count and compare Facebook posts, comments, likes and shares

- **Propose visualization**
  - Comparison of tools, candidates, parties, countries
  - For example:
    → Which candidate is the most visible within the same party, among parties?
    → Compare data stemming from different sources (e.g. preferences of tools, content type) of users and parties

Roadmap

✓ Doing politics in the era of Big Data

✓ Analyzing political campaign strategies in Europe

   ✓ Data collection and curation

   ✓ Comparing for understanding strategies: UK vs. France

■ **Conclusions and perspectives**

Current Work

- Propose a juridical profile of content and tools directly and transitively used by candidates and parties

> Technology is changing the way elections are run.
> **In which extent and how?**
> We need to develop analysis tools in a multidisciplinary context to provide a comprehensive picture

- Propose a data curation process, guided by QoS aspects:
  - Juridical, temporal, provenance, reputation, geography and characteristics of the official organization of the process

?

Geraldine Castel
CEMRA, Université Stendhal, France

Anne Marie Benoit
PACTE, CNRS, France

Genoveva Vargas-Solar
CNRS, LIG-LAFMIA, France

Javier Espinosa
LAFMIA, UMI 3175 CNRS Mexico
javiera.espinosa@gmail.com