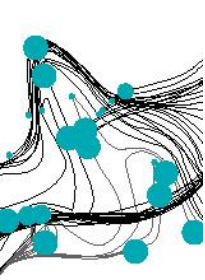# ENRICHING GEO-SOCIAL MEDIA CONTENT THROUGH GEOGRAPHIC CONTEXTUALIZATION

## EARTH OBSERVATION WITH UNCALIBRATED IN-SITU SENSORS

Frank O. Ostermann
RICH-VGI Workshop, AGILE 09.06.2015

- **Introduction: Using geo-social media APIs as sensors**

- Opportunities and challenges: Practical examples

- Outlook on future research directions

# NEW SOURCES OF GEO-INFORMATION
## GEO-SOCIAL MEDIA AS SENSORS

| | | Geography | |
|---|---|---|---|
| | | Explicit | Implicit |
| **Participation** | Explicit | **Volunteered Geographic Information (VGI)**<br><br>Open Street Map | **Volunteered Geographic Content (VGC)**<br><br>Wikipedia articles on non-geographic topics containing place names, Foursquare |
| | Implicit | **Contributed / Ambient Geographic Information (CGI/AGI)**<br><br>Public Tweets referring to the properties of an identifiable place. | **User-Generated Geographic Content (UGGC)**<br><br>Public Flickr images containing a place name or being georeferenced |

Adopted from [1]

# GEO-SOCIAL MEDIA SENSORS - WHAT'S DIFFERENT?
GEO-SOCIAL MEDIA AS SENSORS

- Often In-situ
- Rich, pre-processed information
- Uneven distribution
- Heterogeneous level of quality
- Varying but high update frequency (stream)
- Redundancy of content and channels (sharing)
- Heterogeneous structure
- Unknown source/lineage
- Unclear / changing licencing, property rights, liability (e.g. OpenStreetMap)

- Unknown/Immeasurable precision, error, completeness
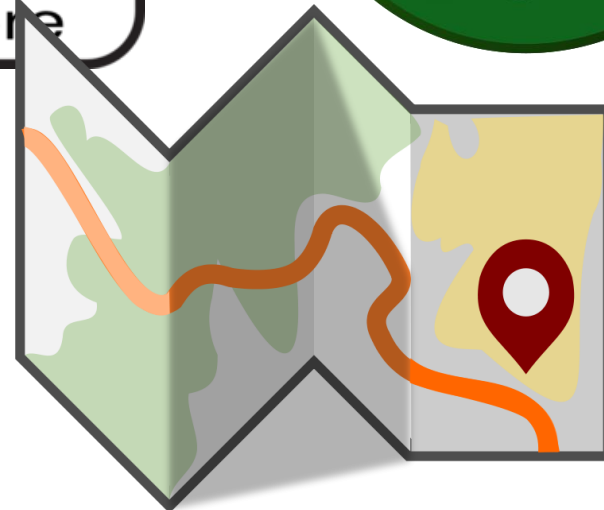- Uncertainty about the uncertainty!

- How to calibrate? (Should we?)

# WHO IS THE CROWD?
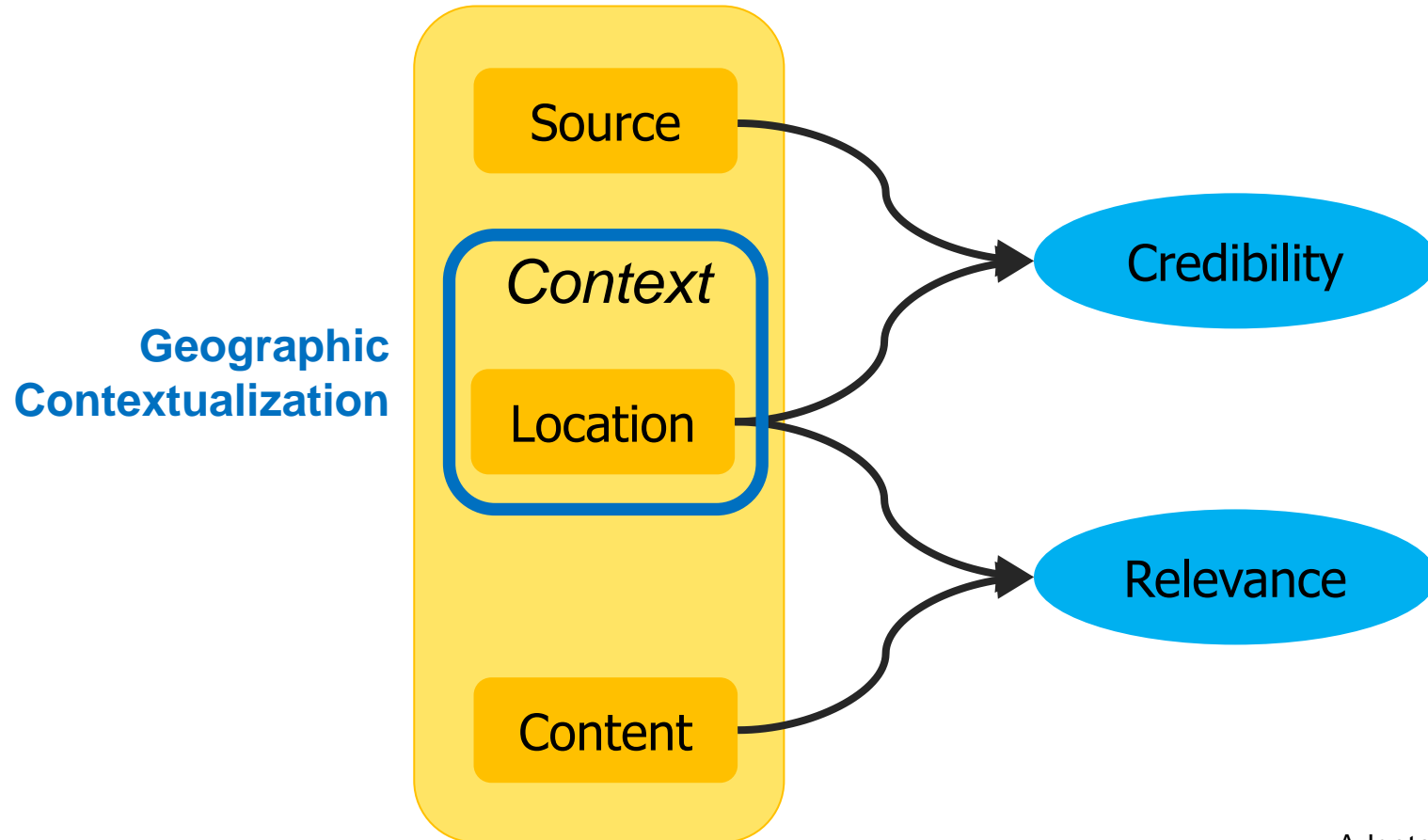## GEO-SOCIAL MEDIA AS SENSORS

# QUALITY ASSESSMENT AND CALIBRATION OF GEO-SOCIAL MEDIA
## GEO-SOCIAL MEDIA AS SENSORS
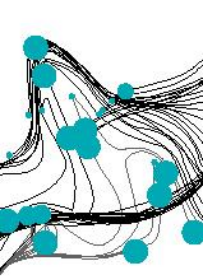


**Geographic Contextualization**

Source

*Context*

Location

Content

Credibility

Relevance

Adopted from [2, 3]

- Introduction: Using geo-social media APIs as sensors

- **Opportunities and challenges: Practical examples**

- Outlook on future research directions

# GEO-SOCIAL MEDIA AND CRISIS MANAGEMENT
## EXAMPLES

| Social media offers… | Crisis management needs… |
|---|---|
| rich up-to-date information ✔ | up-to-date information |
| new paths of communication ✔ | redundant paths of communication |
| noise, uncertain lineage and accuracy ✘ | high-quality and reliable information |

Crowd-sourced data curation faces limits of

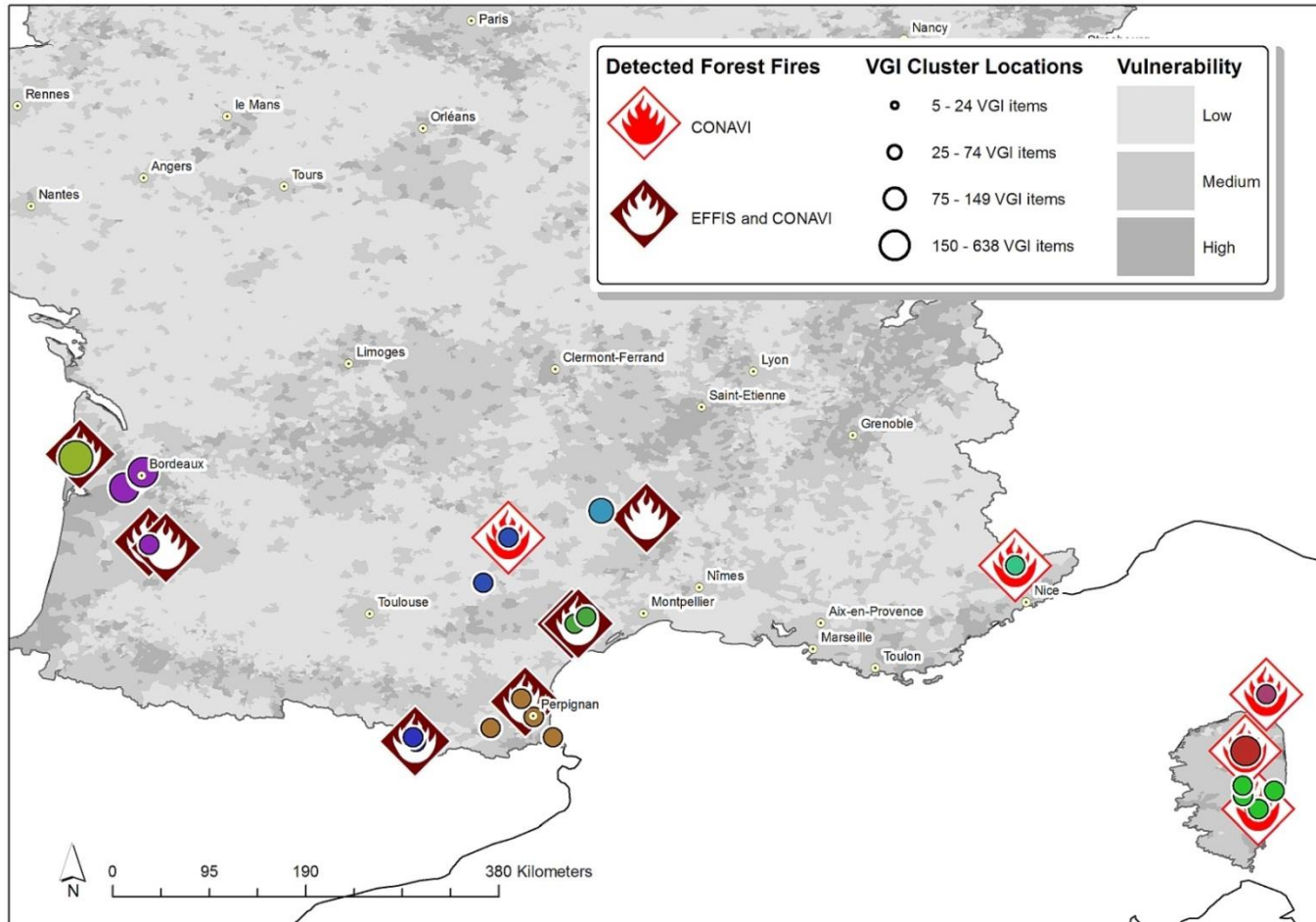- Sustainability
- Scalability

# FOREST FIRES IN FRANCE 2011
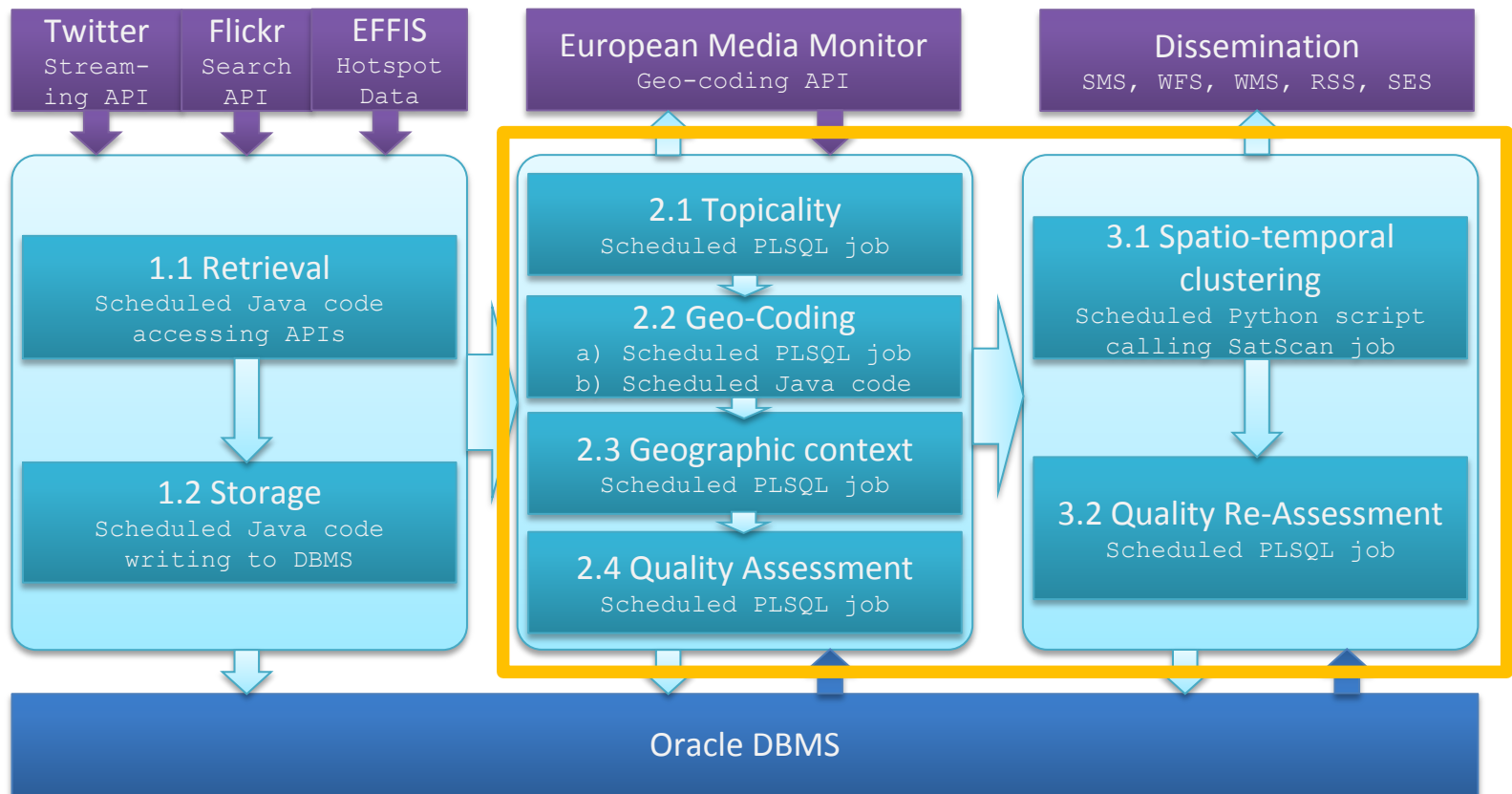EXAMPLES
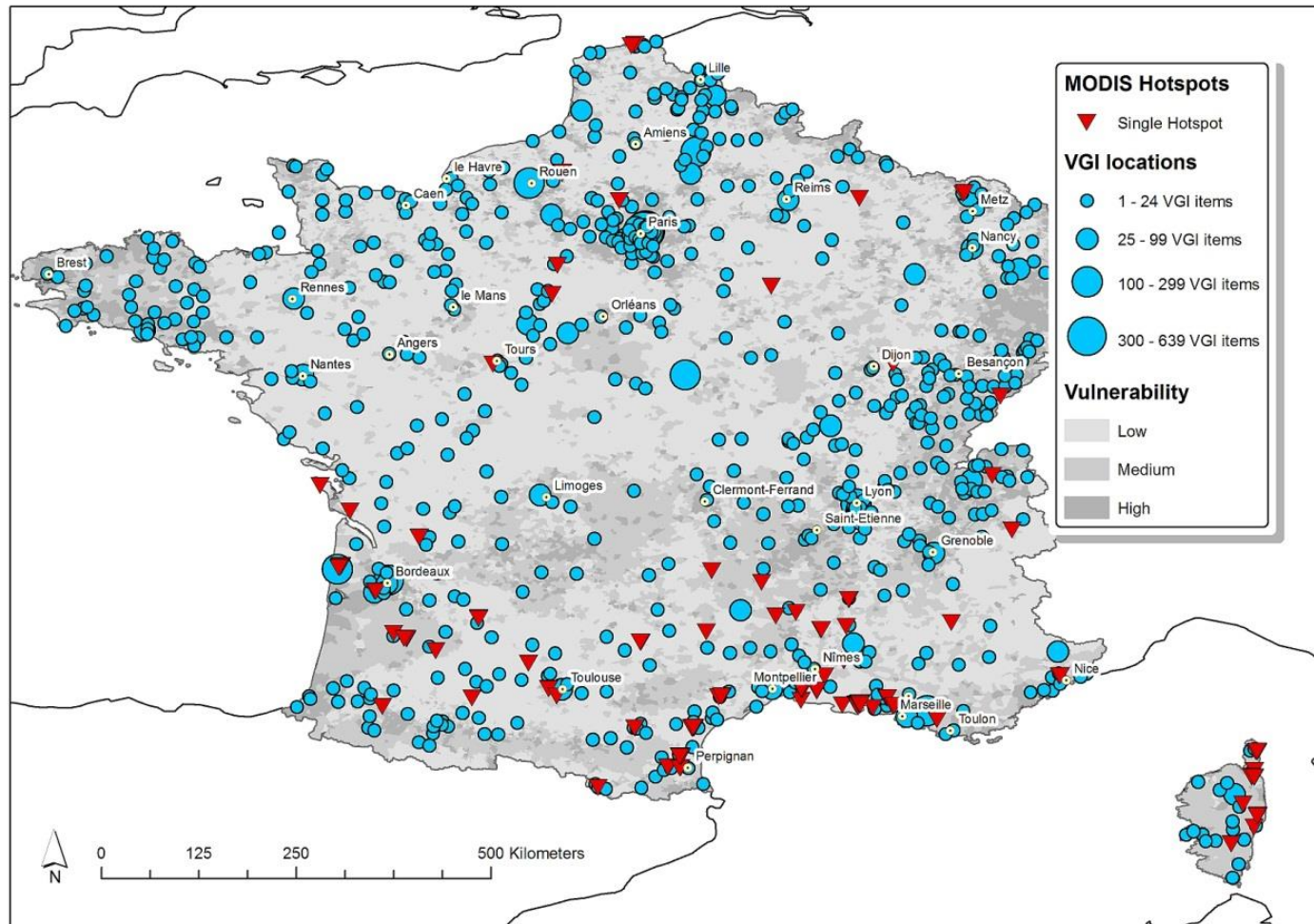


Source: [3]

# FOREST FIRES IN FRANCE BY GEOCONAVI
EXAMPLES



Source: [3]

# GEOCONAVI FIGHTING FOREST FIRES

| Twitter | Flickr | EFFIS |
|---|---|---|
| Stream-ing API | Search API | Hotspot Data |

**1.1 Retrieval**
Scheduled Java code accessing APIs

**1.2 Storage**
Scheduled Java code writing to DBMS

### European Media Monitor
Geo-coding API

**2.1 Topicality**
Scheduled PLSQL job

**2.2 Geo-Coding**
a) Scheduled PLSQL job
b) Scheduled Java code

**2.3 Geographic context**
Scheduled PLSQL job

**2.4 Quality Assessment**
Scheduled PLSQL job

### Dissemination
SMS, WFS, WMS, RSS, SES

**3.1 Spatio-temporal clustering**
Scheduled Python script calling SatScan job

**3.2 Quality Re-Assessment**
Scheduled PLSQL job

### Oracle DBMS

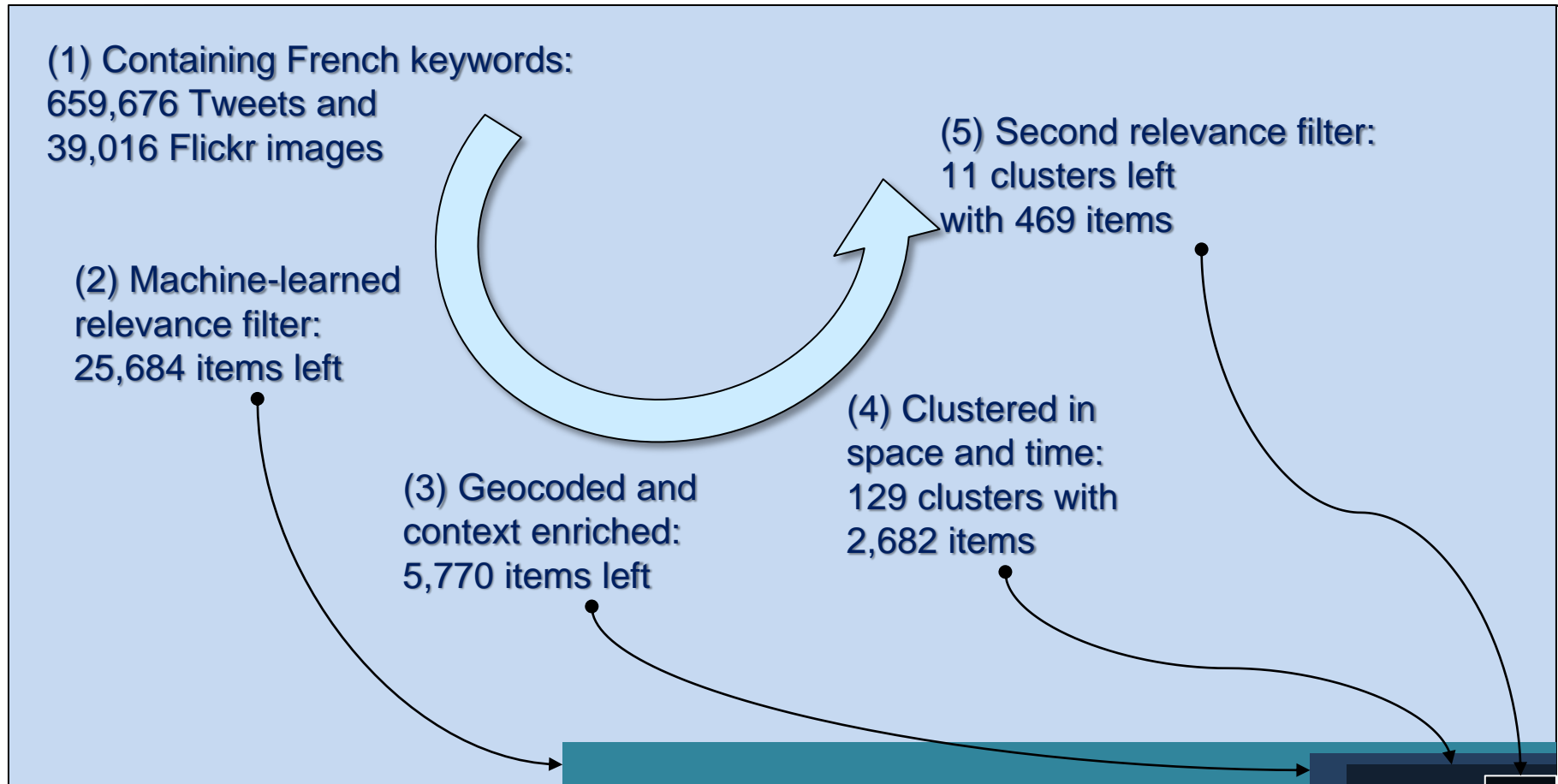# GEOGRAPHIC CONTEXTUATLIZATION

Choice of dataset

- Talk to the domain experts

- Talk to the data experts

- Make a choice

For this case study

- MODIS hotspots

- Population density (vulnerability, reliability)

- Forest cover (risk, reliability)

# FRENCH FOREST FIRE SOCIAL MEDIA
## PAST RESEARCH

(1) Containing French keywords:
659,676 Tweets and
39,016 Flickr images

(2) Machine-learned
relevance filter:
25,684 items left

(3) Geocoded and
context enriched:
5,770 items left

(4) Clustered in
space and time:
129 clusters with
2,682 items

(5) Second relevance filter:
11 clusters left
with 469 items

# SEMANTICS OF PLACES ACROSS GEO-SOCIAL MEDIA
OVERVIEW

- Theory-guided research and local case study:
  - How to people see and understand the places they frequent?
  - What is different across media sources?
- More than one (volunteered) data source
- Identification of places and their semantics
- Comparison of places between data sources
- Comparison of places with geographic features and authoritative data sources

# SEMANTICS OF PLACES
IMPLEMENTATION

- Shatford-Panofsky and Agnew

- Greater London Area

- From Twitter to Flickr

- Data Mining (Spatio-temporal clustering) -> Semantic Analysis (Cosine Similarity, …)

- Geo-demographic data
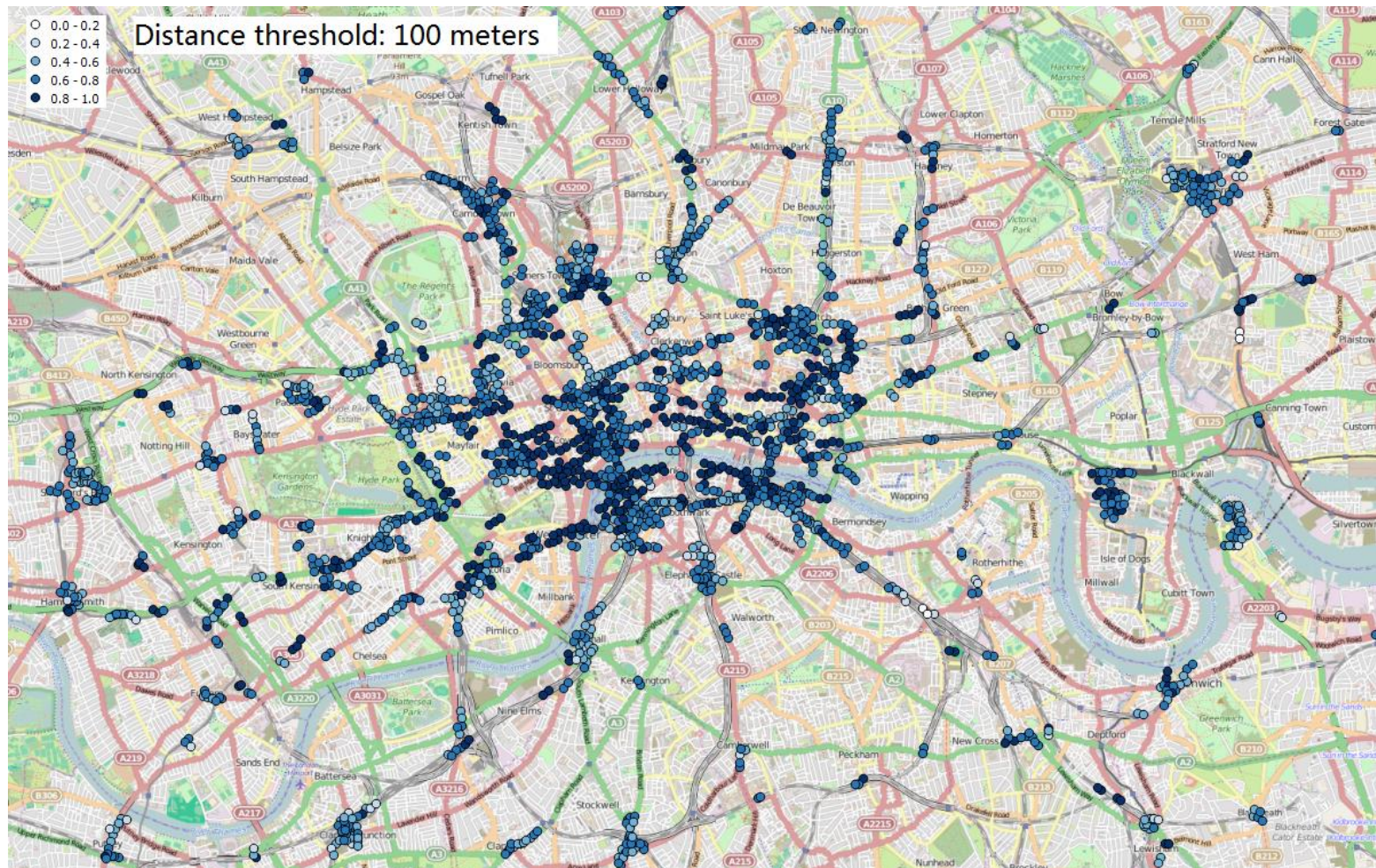
# SEMANTICS OF PLACES
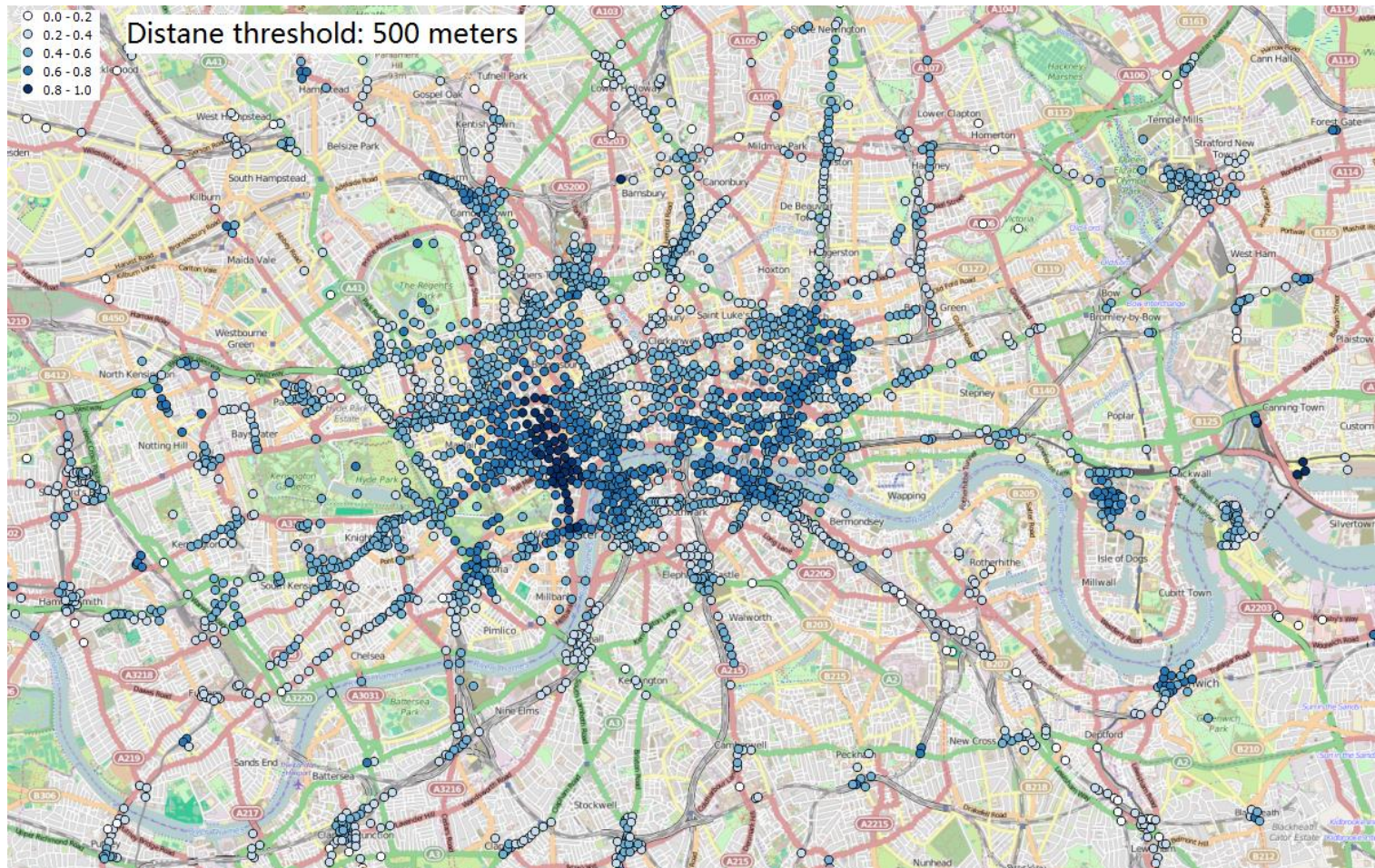IMPLEMENTATION : COSINE SIMILARITY NEAREST NEIGHBORS
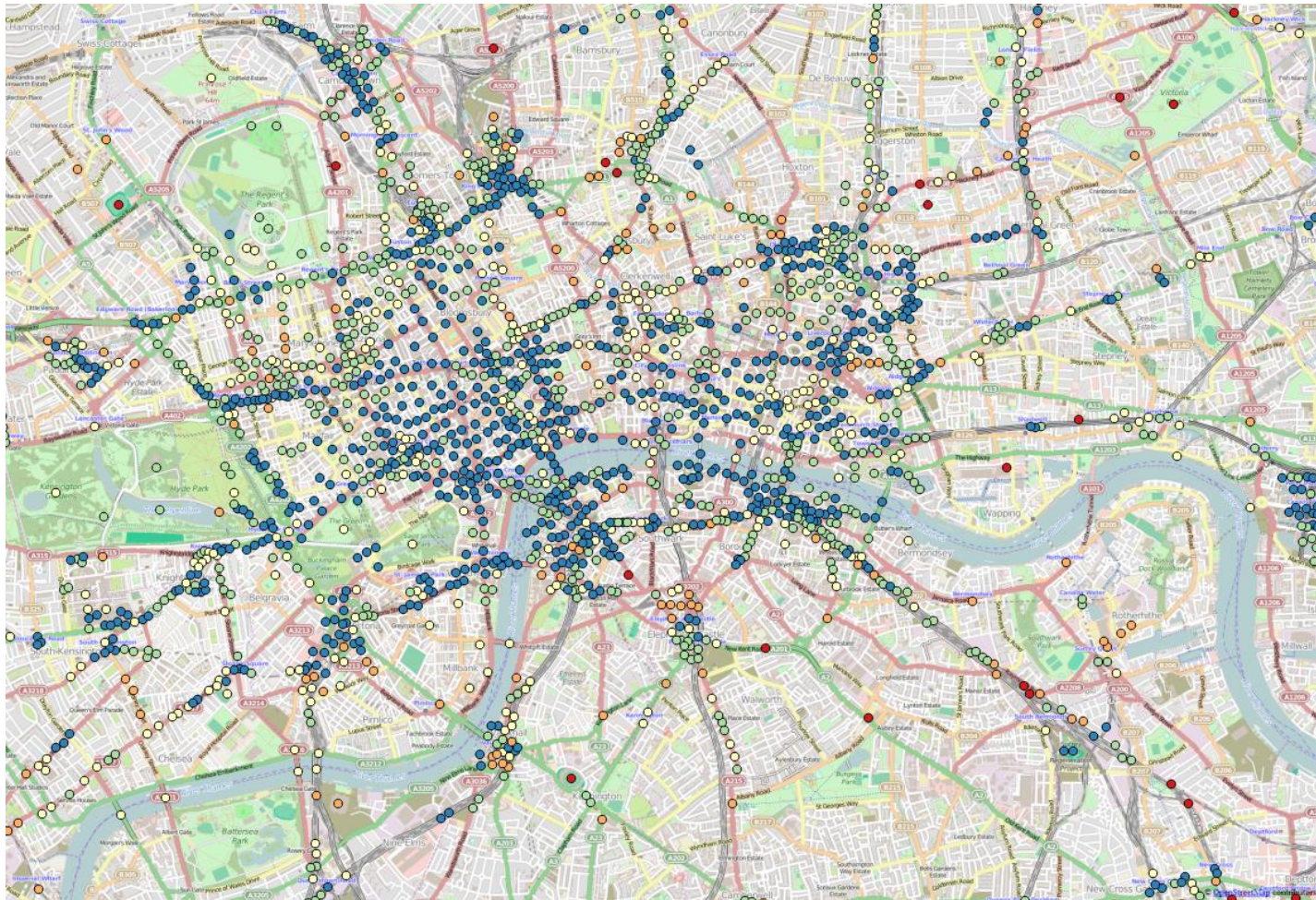
# SEMANTICS OF PLACES
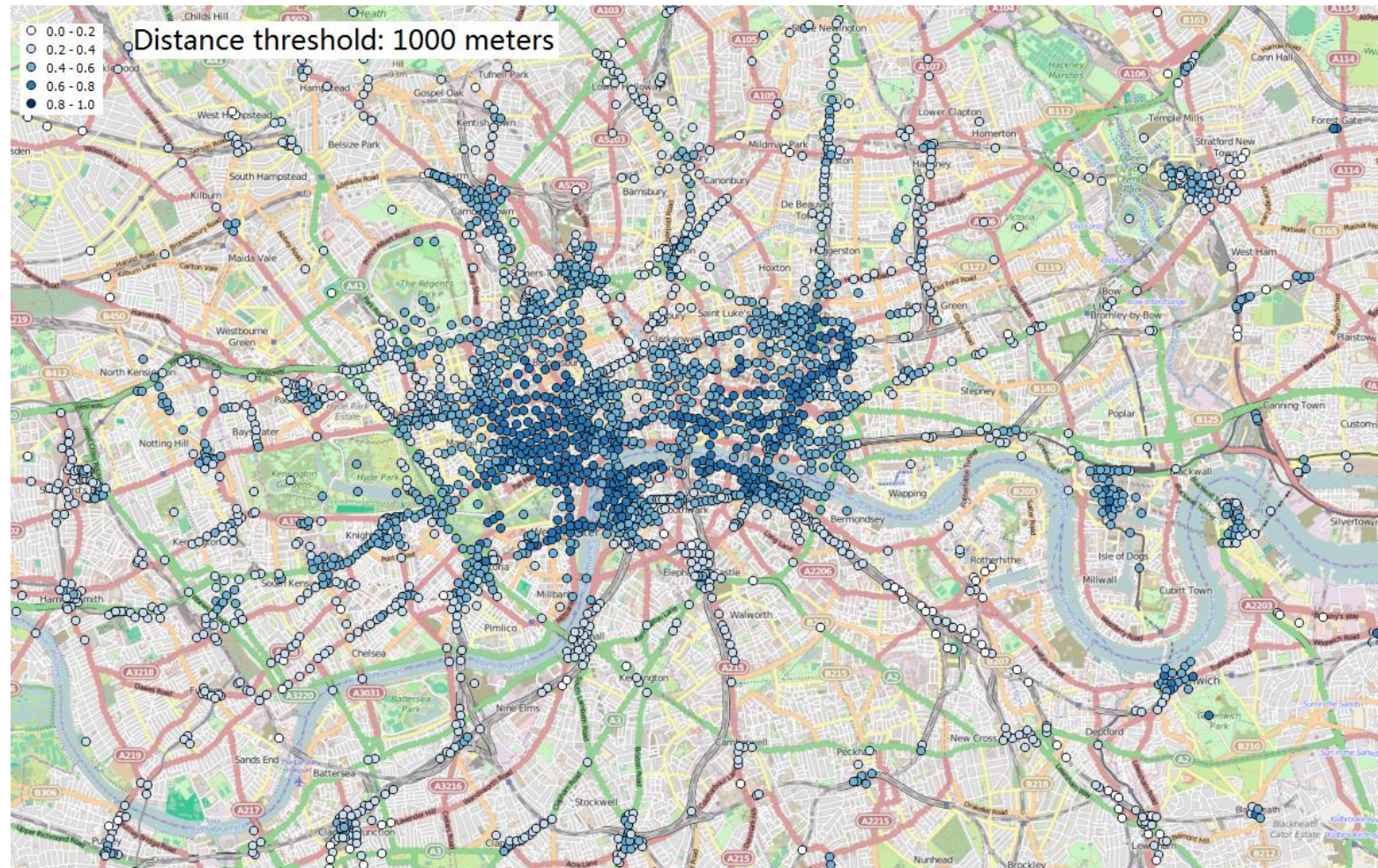## IMPLEMENTATION : CORRELATION DISTANCE & SIMILARITY
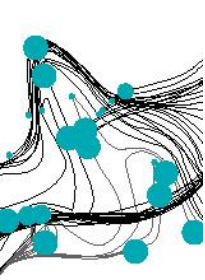
# SEMANTICS OF PLACES
## IMPLEMENTATION : CORRELATION DISTANCE & SIMILARITY

# SEMANTICS OF PLACES
## IMPLEMENTATION : CORRELATION DISTANCE & SIMILARITY

# CHALLENGES AND OPPORTUNITIES OF GEO-SOCIAL MEDIA

## EARTH OBSERVATION WITH UNCALIBRATED IN-SITU SENSORS

- Introduction: Using geo-social media APIs as sensors

- Opportunities and challenges: Practical examples

- **Outlook on future research directions**

# UNSOLVED PROBLEMS FROM FRENCH CASE STUDY
RESEARCH QUESTIONS

Relevant datasets for contextualization

- Choice

- Integration

Geospatial Semantic Web
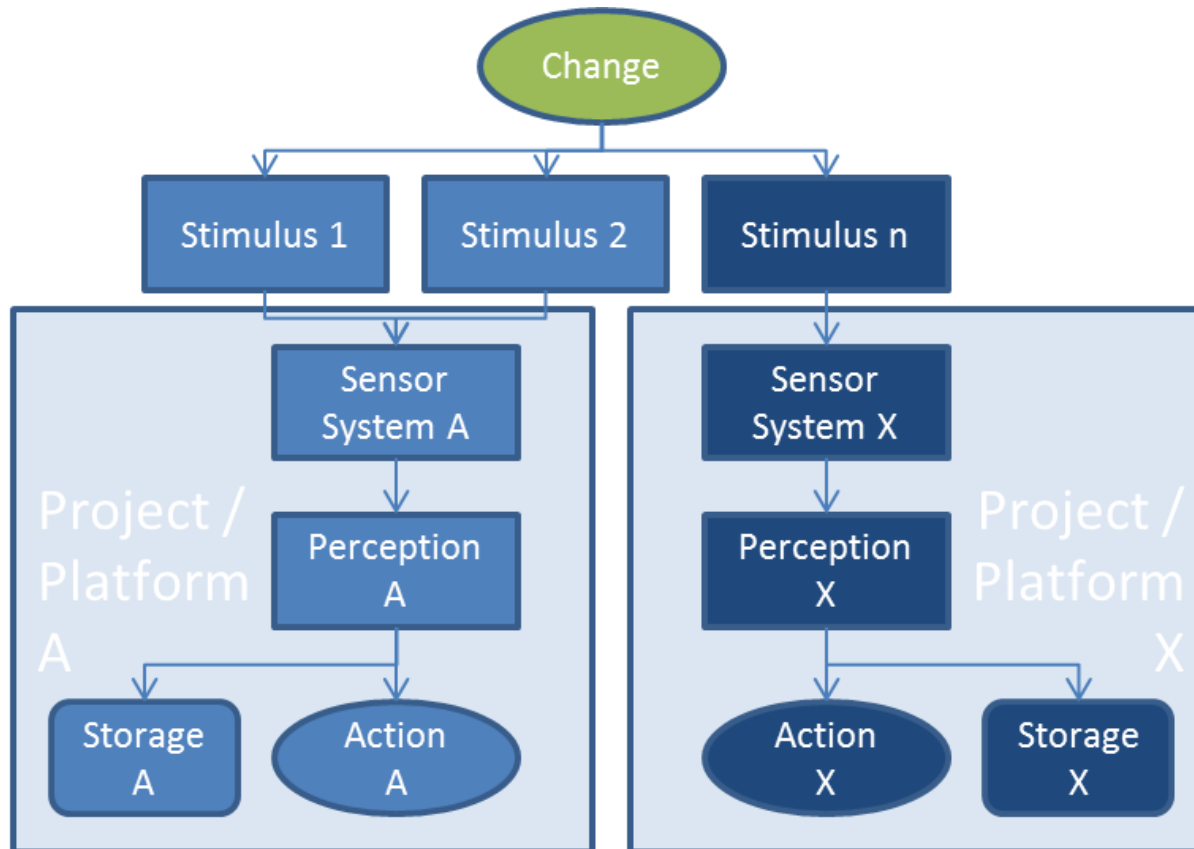Multi-Sensory Integration

Settings for data mining and machine learning
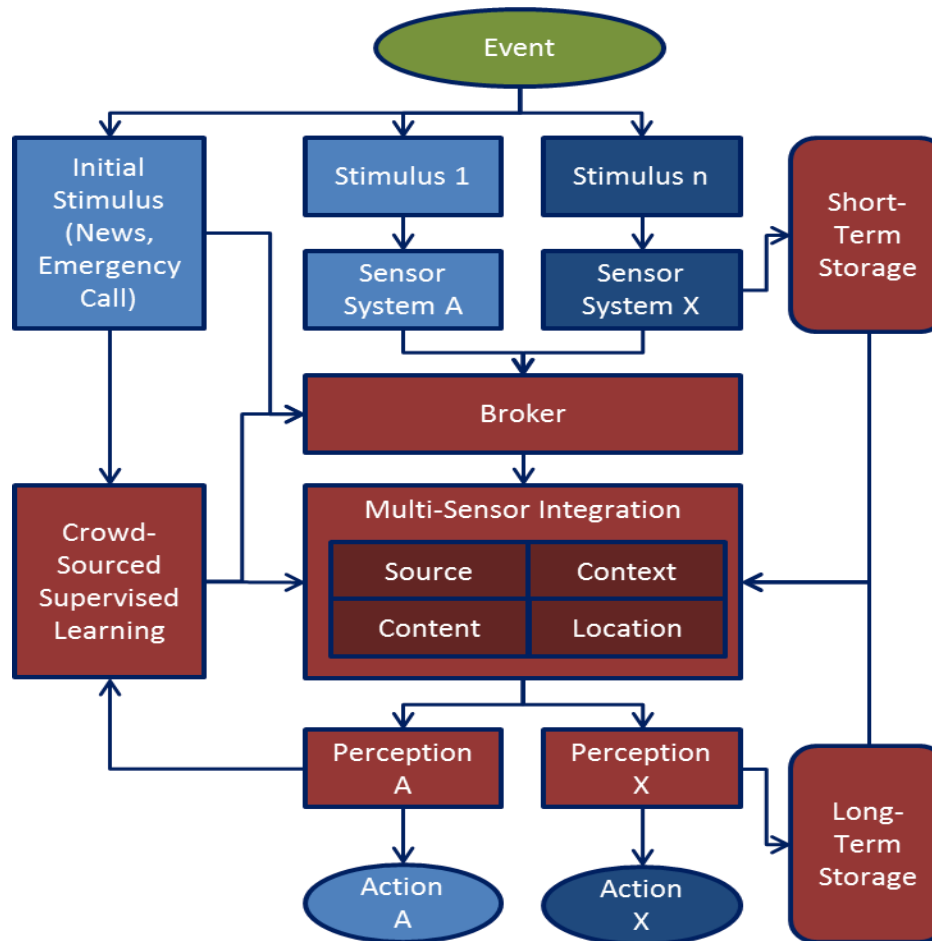
- Method

- Parameters

Crowdsourced Supervision

# INTEGRATING GEO-SOCIAL MEDIA
## WHAT'S HAPPENING NOW

# INTEGRATING GEO-SOCIAL MEDIA
## FUTURE IDEAS

# GEO-SOCIAL MEDIA FUSION
## NEUROSCIENCE PERSPECTIVE

Multi-Sensory Integration

- Combines the information from different sensory systems
- Results in coherent representation of the environment
- Is prerequisite for adaptive behavior and response to the environment
- Decreases sensory uncertainty and reaction times

Characteristics

- Mutual feedback between sensory systems
- Spatial proximity, temporal proximity, and inverse effectiveness

# GEO-SOCIAL MEDIA FUSION
## ENGINEERING PERSPECTIVE

- Sensor data or information fusion

- Focus on

  - Low-level abstracted sensor data

  - Data fusion from several but similar sensors

  - Different but related sensors in close spatial proximity (e.g. robotics)

- Less activity on

  - Integration of heterogeneous sensors covering irregular areas
    (hard/soft data integration from disparate sensors)

# GEO-SOCIAL MEDIA FUSION
## SO WHAT?

- Two major principles from neuroscience and cognitive science align with core GI-principles: what is near in space and time is related

- Inverse effectiveness hints at why outliers might be important

- Engineering provides methods and algorithms

# HYBRID GEO-INFORMATION PROCESSING
## RESEARCH QUESTIONS

**Developing hybrid quality assurance mechanisms for near real-time geo-information streams**

- *How can crowd-sourced supervised machine-learning improve information quality?*

- *Which are feasible approaches and implementations of crowd-sourced and cloud-based real-time processing and information dissemination?*

- *How can we crowdsource the analysis of model outputs and data mining processes?*

Key Technologies

- Apache Spark / Storm

- Active Learners

- Cloud Computing

# HYBRID GEO-INFORMATION PROCESSING
## WORKFLOW
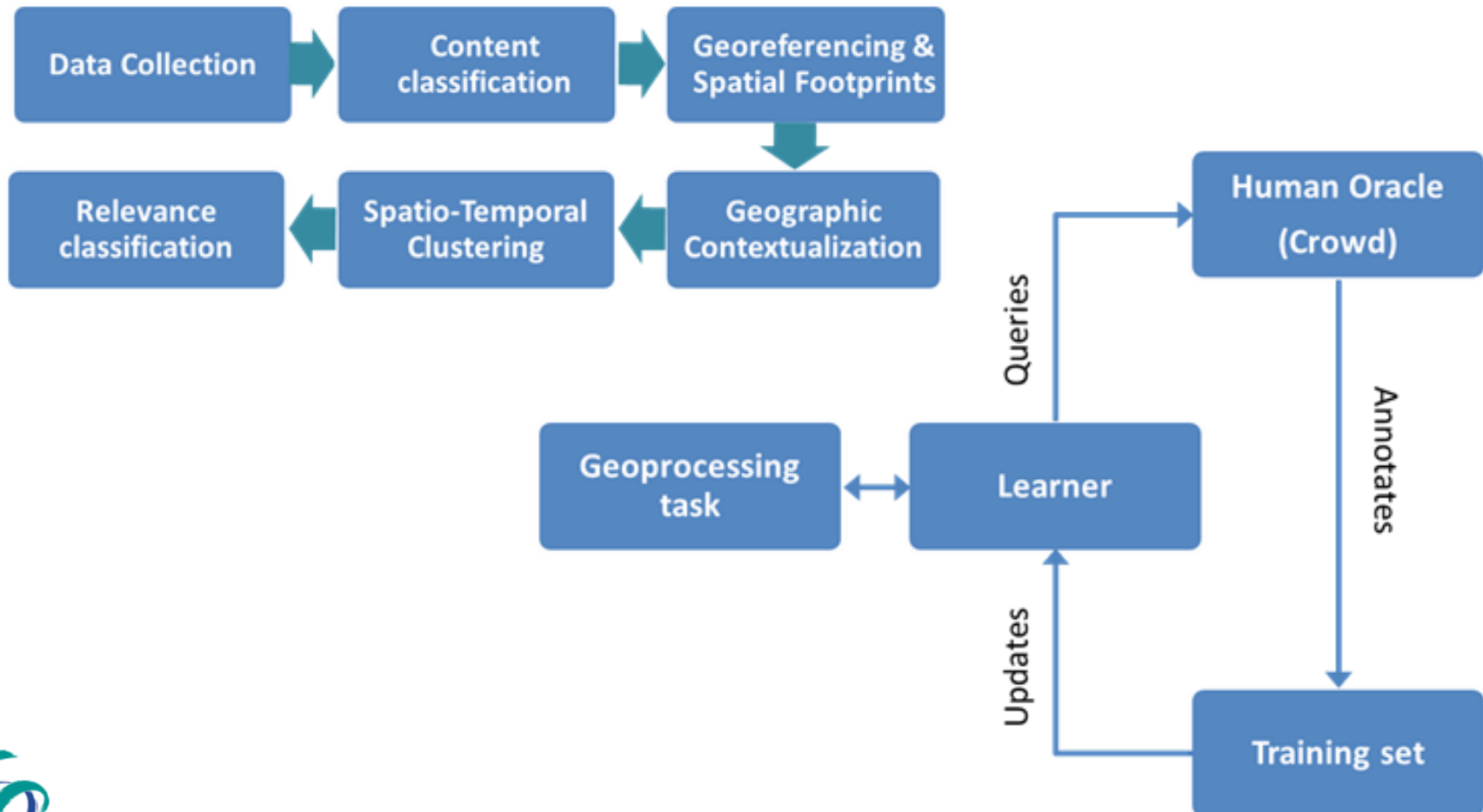
# MORE INFORMATION

- [1] Craglia, M., Ostermann, F., & Spinsanti, L. (2012). Digital Earth from vision to practice: making sense of citizen-generated content. *International Journal of Digital Earth*, *5*(5), 398–416.

- [2] Ostermann, F., & Spinsanti, L. (2012). Context Analysis of Volunteered Geographic Information from Social Media Networks to Support Disaster Management: A Case Study On Forest Fires. *International Journal of Information Systems for Crisis Response and Management*, *4*(4), 16–37.

- [3] Spinsanti, L., & Ostermann, F. (2013). Automated geographic context analysis for volunteered information. *Applied Geography*, *43*(9), 36–44.

- http://www.slideshare.net/jrc_vgi_ff/geographic-context-analysis-of-volunteered-information

- https://sites.google.com/site/geoconavi/

- http://geocommons.com/maps/183605

# CHALLENGES AND OPPORTUNITIES OF GEO-SOCIAL MEDIA
## EARTH OBSERVATION WITH UNCALIBRATED IN-SITU SENSORS

Thank you!

f.o.ostermann@utwente.nl
@f_ostermann
nl.linkedin.com/in/foost