

# OTMedia

INTERNATIONAL WORKSHOP

**Big Data for  
Media Analysis**

---

**Analyser les médias à partir  
de larges corpus de données  
numériques**

December 16th-17th, 2013

INA Issy Les Moulineaux

Paris – France

PARTNERS / PARTENAIRES



**LATTS**

École des Ponts ParisTech  
Université Paris-Est Marne-la-Vallée

## Presentation

The media are often presented as the “fourth estate”, playing a central role in the democratic debate: they investigate, verify and inform. Information is steadily growing more and more abundant, to the point of becoming hard to gather and understand. Media analysis requires today the implementation and use of tools capable of dealing with large digital corpora.

The conference aims at gathering researchers and experts in the processing, analysis, interpretation and visualization of data regarding the media. On the first day, several platforms and tools for media analytics will be presented and compared, looking for mutual strengths and potential biases. The second day will focus on how social scientists and industry experts use quantitative and qualitative methodologies to obtain results. These are some of the questions that will be addressed: Who produces information? How does information spread? What are the relationships among the different channels (radio, tv, web)? What is the role of the news agencies today? How do media landscapes evolve, and in particular the French one? What is the impact of the media on public policies?

The Transmedia Observatory (OT-Media) is a research project building a large database on the French media and developing innovative tools for data mining and visualization. More than 5 million documents have been gathered since July 2011, from over 2,000 fluxes, including tv, radio, web and AFP dispatches.

This conference, closing the OT-Media project, aims to realize an exchange of experiences on criteria of analysis and methodological stances, as well as on the evaluation procedures for these platforms.

## Présentation

Les médias sont souvent présentés comme un quatrième pouvoir, pouvoir qui joue un rôle central dans le débat démocratique : enquêter, vérifier, informer. Si l'information est essentielle, celle-ci devient de plus en plus abondante, abondante au point de la rendre difficile d'accès. L'analyse de données numériques nécessite alors la mise en place d'outils capables d'appréhender de larges corpus.

La conférence se positionne dans ce contexte et vise à réunir chercheurs et experts de la fouille de données et de l'analyse médiatique afin de dégager, à la croisée des recherches technologiques et en sciences humaines et sociales, les enjeux des évolutions techniques et des pratiques médiatiques : qui produit l'information ? Comment se propage-t-elle ? Quelles sont les relations entre les différents supports, de la radio, à la télévision, en passant par le web ? Quel rôle pour les agences de presse à l'instar de l'AFP ? Quelle est l'évolution de la diversité du paysage médiatique, spécialement français ? Qu'en est-il encore de l'impact que les médias ont sur les politiques publiques ?

L'Observatoire Transmedia (OT-Media) est un projet de recherche qui vise à développer une vaste base de données des médias français et des outils innovants de fouille de données comme de visualisation. Avec près de 2000 flux issus de la télévision, de la radio du web et encore de l'AFP, la base de données rassemble plus de 5 millions de documents depuis juillet 2011.

Cette conférence vient clôturer le projet ANR OT-Media et se propose d'échanger pendant deux jours autant sur les outils, les critères d'analyse et les mesures à mettre en œuvre, que sur les approches méthodologiques et les questions d'évaluation et de validation de ces plateformes.

## Monday 16

### **Media Analytics: Platforms and Tools**

#### **Fouille de données médiatiques: plateformes et outils**

9:00 – Welcome Coffee

9:30 – Opening

Marie-Luce Viaud (INA) and Franck Rebillard (Université Paris 3)

10:00 – First Session A

Mathieu Jacomy and Benjamin Ooghe-Tabanou (Sciences Po Paris)

#### **Projet DIME-SHS et outil Hyphe**

Brendan Jou (Columbia University)

#### **News Rover Project**

11:30 – Coffee Break

11:45 – First Session B

Jerome Thièvre (INA) and Alexis Joly (INRIA)

#### **Plateforme OT-Media**

12:30 – Presentation of the Demos

13:00 – Lunch Break & Demos

14:15 – Second Session A

Marta Severo (Université Lille 3)

#### **Projet ANR GeoMedia**

Camille Maussang (Linkfluence)

#### **Construire un moteur d'actualités à partir de milliers de sources web et basé sur les thématiques préférées des lecteurs**

Lucien Castex (INA, Université Paris Est)

#### **Du mariage pour tous à la manif pour tous, approche polysémique d'un débat de société.**

16:15 – Coffee Break & Demos

16:30 – Second Session B

Agnes Saulnier and Marion Dupeyrat (INA)

#### **Évaluation et analyse de biais du prototype OT-Media à partir de quelques exemples d'études**

17:15 – Round Table 1

J. Holcomb, B. Jou, J. Cagé, N. Hervé

#### **Computer Scientists and Social Scientists: Understandings/Misunderstandings**

18:15 – Demos

19:00 – Aperitif

Tuesday 17

**Quantitative and Qualitative Methodologies for the Analysis of the Media**

**Méthodologies quantitatives et qualitatives pour l'analyse des médias**

9:30 – Third Session A

Jesse Holcomb (Pew Research Center)

**A Healthy Tension: Advancing Media Analysis Through External Data Partnerships**

Sylvain Parasio (Université Paris Est)

**Connecting the Dots: An Ecological Perspective on How Publics Gather Around Criminal Data**

11:00 – Coffee Break & Demos

11:30 – Third Session B

Julia Cagé (University of Harvard) and Nicolas Hervé (INA)

**The Production Function of Information in an Online World**

David Strömberg (Institute for International Economic Studies)

**The Public of the Media**

13:00 – Lunch Break & Demos

14:15 – Fourth Session

Cécile Méadel (Mines ParisTech)

**L'usage des techniques de communication**

Eric Lagneau (EHESS Paris)

**Observation ethnographique de rédactions à campagne en 2012: quelques réflexions méthodologiques**

Franck Rebillard & Dario Compagno (Université Paris 3)

**Media Analytics for the Modeling of Long Events**

16:15 – Coffee Break & Demos

16:30 – Round Table 2

D. Strömberg, C. Grasland, M. Severo, M. Jacomy, C. Meadel, S. Parasio

**Possibilités et limites des Big Data pour l'analyse des médias**

17:30 – Wrap Up and Conclusions

Franck Rebillard (Université Paris 3) and Marie-Luce Viaud (INA)



21-23 rue Camille Desmoulins  
Issy-les-Moulineaux

### **Directions**

RER C – Station ISSY-VAL DE SEINE  
Metro 12 – Station Corentin Celton, Mairie d'Issy  
Metro 8 – Station Balard  
T2 – Station ISSY-VAL DE SEINE

[www.otmedia.fr/conference](http://www.otmedia.fr/conference)

Observatoire Transmédia  
OT-MEDIA

Projet ANR CONTINT (2010-2013)

### **Organizing Committee**

Lucien Castex (INA, Université Paris Est)  
[contact.castex@gmail.com](mailto:contact.castex@gmail.com)  
Dario Compagno (Université Sorbonne Nouvelle Paris 3)  
[dario.compagno@gmail.com](mailto:dario.compagno@gmail.com)