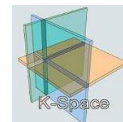




# Bringing The IPTC News Architecture into the Semantic Web

Raphaël Troncy, <[Raphael.Troncy@cwil.nl](mailto:Raphael.Troncy@cwil.nl)>

CWI, Semantic Media Interfaces



ISWC 2008: Wednesday, 29 October 2008



Wikipedia is sustained by people like you. Please donate today.

article discussion edit this page history

## Global financial crisis of September–October 2008

From Wikipedia, the free encyclopedia

It has been suggested that this article or section be merged with *Economic crisis of 2008*. (Discuss)

The **global financial crisis of September–October 2008** is a major ~~global~~ financial crisis, the worst of its kind since the Great Depression. It became prominently visible in September 2008 with the failure, merger or conservatorship of several United States-based financial firms. The underlying causes leading to the crisis had been reported in business journals for many months before September, with commentary about the financial stability of leading U.S. and European investment banks, insurance firms and mortgage banks consequent to the subprime mortgage crisis.<sup>[1][2][3][4]</sup>

Beginning with failures of large financial institutions in the United States, it rapidly evolved into a global crisis resulting in a number of European bank failures and declines in various stock indexes, and significant reductions in the market-value of equities (stock), down 27% as of October 24,<sup>[5]</sup> and commodities worldwide.<sup>[1]</sup> The crisis has led to a liquidity problem and the de-leveraging of financial institutions especially in the United States and Europe, which further accelerated the liquidity crisis. World political leaders and national ministers of finance and central bank directors have coordinated their efforts to reduce fears but the crisis is ongoing and continues to change. The crisis has roots in the subprime mortgage crisis and is an acute phase of the financial crisis of 2007–2008.

Contents [hide]

- 1 Week of September 7, 2008
- 2 Week of September 14, 2008
- 2.1 Major financial firm crisis

Suivez l'information en continu, accédez à 70 fils de dépêches thématiques. Abonnez-vous au Monde.fr: 6€ par mois + 30 jours offerts

Pourquoi le déprime continue malgré les milliards injectés ? L'analyse de l'économiste Elie Cohen.

Toutes les vidéos

videos

and what are

ises in the

IRAQ

FINANCIAL CRISIS = RELAX, I'M ON TOP OF IT?

CIVIL LIBERTIES

Noooo!

DEFICITS

KARUNA

credit crunch

African investment could be hit

China 'can be engine of growth'

Are my savings safe?

Have bail-outs worked?

CRISIS OVERVIEW

How market moves affect you

cartoons

Terminé

ISWC 2008: Wednesday, 29 October 2008

cc BY NC SA

Wikipedia is sustained by people like you. Please donate today.

article discussion edit this page history

## Global financial crisis of September–October 2008

From Wikipedia:

The global financial crisis of 2008 with the many months subprime mortgage crisis. Beginning with the subprime mortgage crisis, and the subsequent de-leveraging of the central bank and the phase of the financial crisis.

1 Week of Set  
2 Week of Set  
2.1 Major  
2.2 Minor

Suivez l'information en continu, accédez à l'abonnement au Monde.fr: 6€ par mois

WordPress.COM

Home Sign Up Features Blog Story Advanced

## Blogs about: Financial Crisis

Featured Blog

Brown: "Welfare reform will be intensified"

"responsibility on people to do all they can to help themselves" Well, the bankers have helped themselves: now it's our turn. Helene Mulholland Guardian 27th October, 2008 Gordon... more --

In These New Times

Have your say. Start a blog. See our free features --> Sign Up Now!

ISWC 2008: Wednesday, 29 October 2008

3

## News Workflow Interoperability

- No integration of media (stories, photo, animation, video)
- Little (or no) context in the news presentation
- Lack of interoperability in the current workflow

REUTERS IPTC AFP EBU-UER BBC

NAR Schema ---> Broadcaster Schema

NewsCodes ---> Controlled Vocabularies ---> User Vocabulary

ISWC 2008: Wednesday, 29 October 2008

4

## Metadata is Key

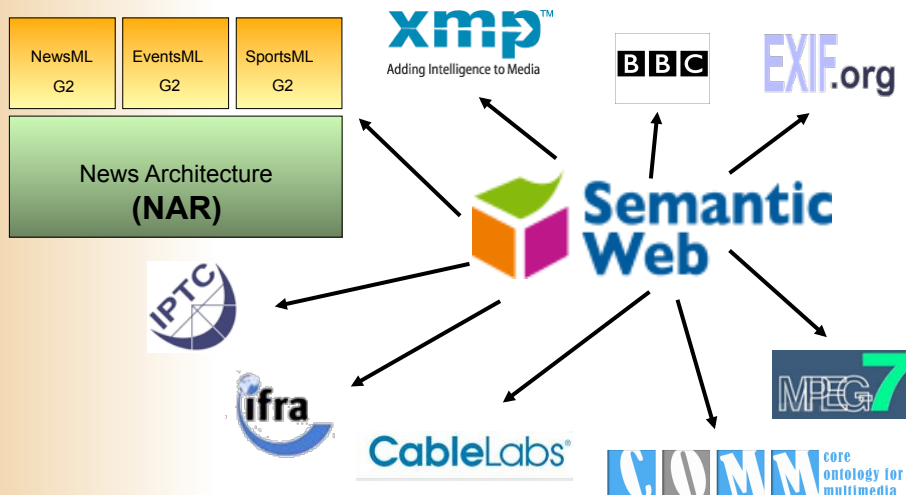
- (Ultimate) Goal:
  - Provide an environment for **searching** and **browsing contextualized multimedia news** information
- Required integration:
  - Data: various media, different forms, various sources
  - Metadata: schema integration, semantic models
- Influence and implications of UI:
  - How to **represent** semantic multimedia metadata to facilitate **presenting** information?
  - *in other words* ... What constraints do end-user interfaces put on the modeling of the metadata?

ISWC 2008: Wednesday, 29 October 2008



5

## News and Multimedia Formats



ISWC 2008: Wednesday, 29 October 2008



6

## Porting Schemas and Thesauri to the Semantic Web

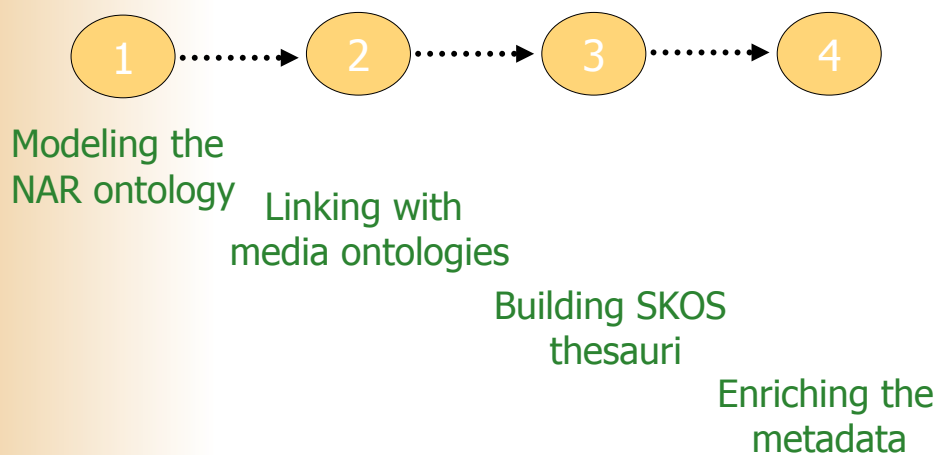
- Methodologies and tools for building ontologies:  
... from scratch
  - "SKOSification" of thesauri in the CH domain:
    - preparation, syntactic and semantic conversion,  
standardization
- ⇒ Lack of best practices for  
modeling ontologies from UML diagrams,  
integrating ontologies with various thesauri,  
while taking the end-user interface into account

ISWC 2008: Wednesday, 29 October 2008



7

## Building a Semantic Web Infrastructure for News

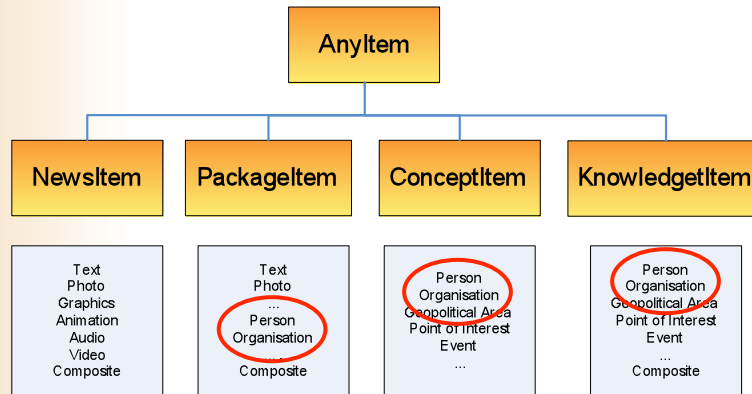


ISWC 2008: Wednesday, 29 October 2008



8

## Step 1: Modeling the NAR Ontology



⇒ focus on reuse of XML types leading to multiple repetition resulting in overly complex nested XML structures

ISWC 2008: Wednesday, 29 October 2008



9

## Step 1: Modeling the NAR Ontology

### • Flattening the XML structure

```

<!--AFP NewsML2 text-photo profile-->
<!--Processed by Xafpl-4ToNewsML2 b11-->
<newsMessage>
+ <header></header>
- <itemSet>
- <newsItem guid="urn:newsml:afp.com:20010101:DV90996" schema="0.7" version="11" xml:lang="en">
  <catalogRef href="http://iptc.org/std-dev/NAR/1.0/specification/IPTC-TempCatalog-inc_3.xml"/>
  - <itemMeta>
    <contentClass code="ccls:photo"/>
    <provider literal="afp.com"/>
    <itemCreated>2006-07-09T21:20:00Z</itemCreated>
    <modified>2006-07-11T09:14:38Z</modified>
    <fileName>DV90996</fileName>
    <edNote>MOBILE SERVICES OUT</edNote>
  </itemMeta>
  + <contentMeta></contentMeta>
  + <contentSet></contentSet>
</newsItem>
</itemSet>
</newsMessage>
  
```

NewsItem  
↑  
PhotoNewsItem

ISWC 2008: Wednesday, 29 October 2008



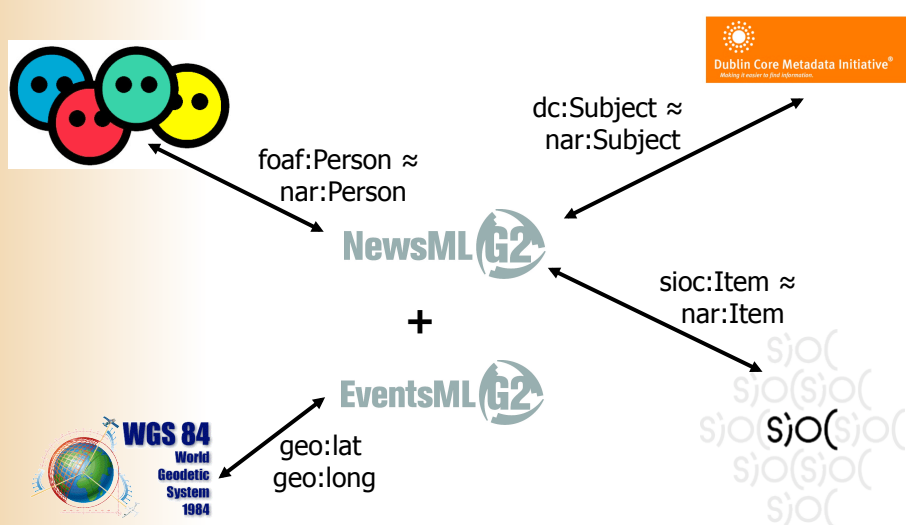
10

## Step 1: Modeling the NAR Ontology

- Modeling unique identifiers
  - Use of dereferencable URIs for any resources (news items + vocabularies)
  - Future: Use of URIs for resource fragments  
<http://www.youtube.com/watch?v=1bibCui3IFM#t=1m45s>
- Modeling the provenance of the information
  - Reification
  - Named (and Networked) Graphs

```
{<> nar:subject cat:11002000}  
      dc:creator team:md ;  
      dc:modified ``2005-11-11T08:00:00Z``.
```

## Step 2: Linking with Media Ontologies





## Step 3: Getting SKOS Vocabularies

The screenshot shows the IPTC NewsCodeViewer/Editor interface. On the left is a tree view of categories, with 'television industry' selected. The right pane displays the details for this entry:

- FormalName:** 04010010 (highlighted with a red circle)
- Name:** television industry
- Explanation:** Stories related to the business of television
- Change comment:** none
- Translations (right click for language specific menu):**
  - Name:** (empty field)
  - Explanation:** (empty field)
  - First version:** 0
  - Change version:** 0
  - Deprecated in version:** 0
  - Change comment:** (empty field)

ISWC 2008: Wednesday, 29 October 2008



13

## Step 3: Getting SKOS Vocabularies

The screenshot shows a Mozilla Firefox browser window displaying an RDF/XML document. The document is an SKOS vocabulary for the 'television industry'. The following lines are circled in red:

- `<skos:Concept rdf:about="http://newsml.cwi.nl/NewsCodes/topicset.iptc-subjectcode/04010010">`
- `<skos:inScheme rdf:resource="http://newsml.cwi.nl/NewsCodes/topicset.iptc-subjectcode/">`
- `<skos:broader rdf:resource="http://newsml.cwi.nl/NewsCodes/topicset.iptc-subjectcode/04010010">`

ISWC 2008: Wednesday, 29 October 2008



14

## Step 4: Enriching the News Metadata



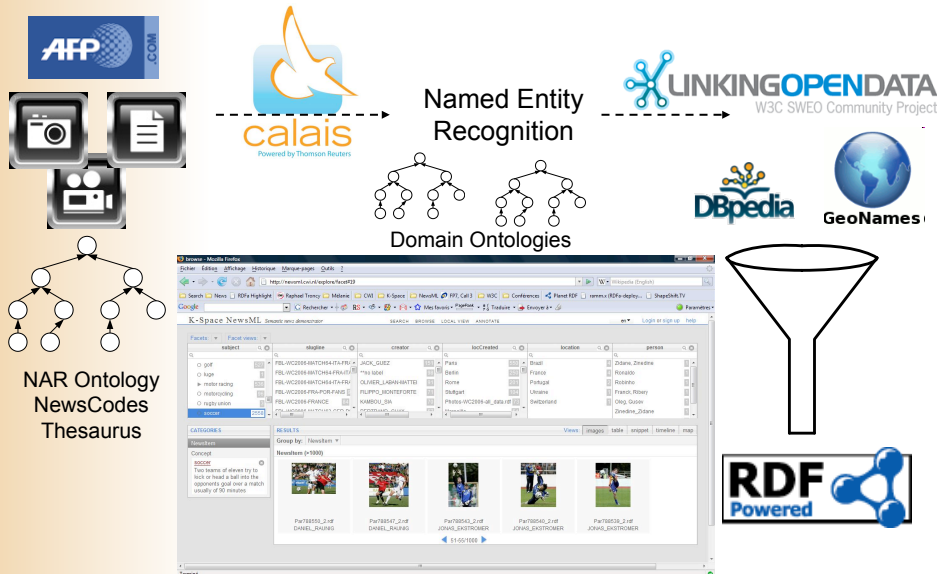
- Concepts/Entities that are subject of news
  - Thematic categories
  - People
  - Organizations
  - Geopolitical Areas
  - Points of Interest
  - Events
  - Products or artefacts

© IPTC – www.iptc.org



15

## Step 4: Enriching the News Metadata



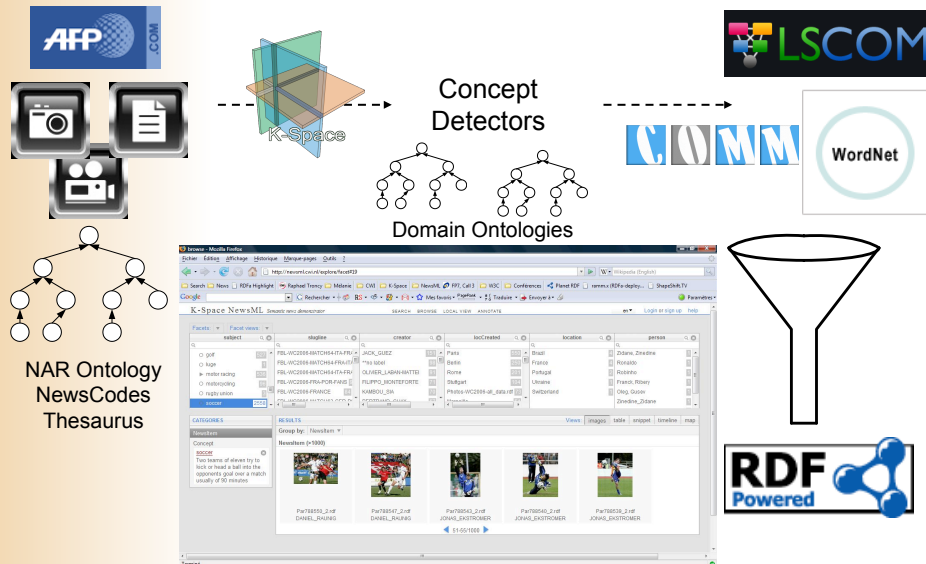
ISWC 2008: Wednesday, 29 October 2008



16



## Step 4: Enriching the News Metadata

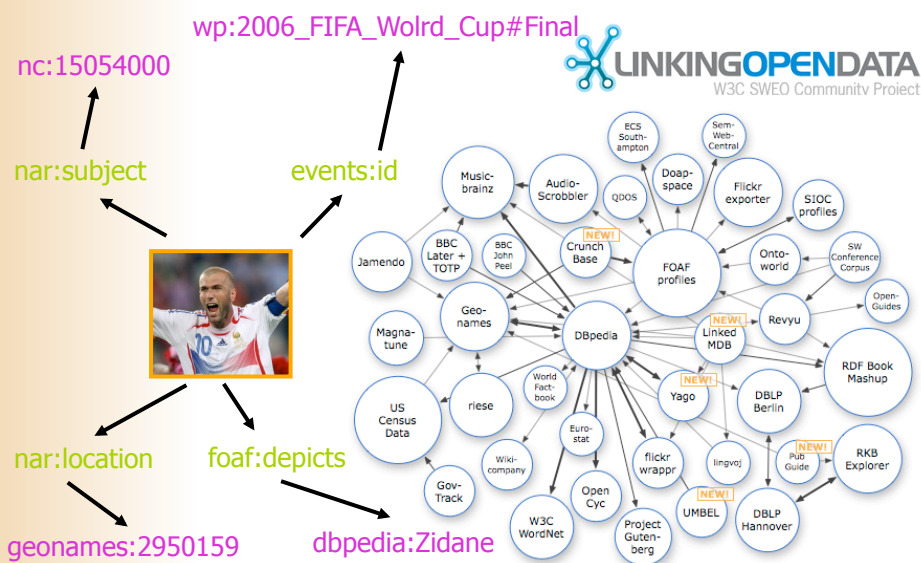


ISWC 2008: Wednesday, 29 October 2008



17

## Web of Data and Linked Data



ISWC 2008: Wednesday, 29 October 2008



18

## Presenting News Information



- Dimensions used for searching news items

- When *time* 10/07/2006
- Where *location* Paris
- What *is depicted* J. Chirac, Z. Zidane
- Why *event* WC 2006
- Who *photographer* Bertrand Guay, AFP

**Metadata**

ISWC 2008: Wednesday, 29 October 2008



19

## Semantic Search of Multimedia News

Description	Number of RDF Triples
General Ontologies: NAR, DC, FOAF	7,336
Domain Specific Ontologies: football	104,358
Thesauri: newscodes	34,903
DBpedia, Geonames	53,468
AFP News Feed (June/July 2006)	804,446
AFP Photos (June/July 2006)	61,311
INA Broadcast Video (June/July 2006)	1,932
<b>Total</b>	<b>1,067,754</b>

Powered by ClioPatria  
1.0 alpha 3

ISWC 2008: Wednesday, 29 October 2008



20

K-Space NewsML


Semantic news demonstrator

[Login](#)
[help](#)
[English](#)


[search](#)
[browse](#)
[local view](#)

This news search engine will give you access to news items kindly provided by AFP  
Type a keyword, for example: Amsterdam, Lyon, Beyrouth, Zidane or G8.

Collections





AFP.com (> 100.000 objects)





INA.fr (> 500 objects)

Vocabularies & Thesauri










ISWC 2008: Wednesday, 29 October 2008



21

## Conclusion

- 4-Steps methodology for building an ontology-based news infrastructure
  - UML-2-OWL: Flatten XML structure, Identify all resources
  - SKOS-ify existing thesauri and use the Web of Data
  - Reuse what is there ... Expose what you make
- Enrich metadata with text and visual analysis
  - Provide new dimensions (facets) for browsing the data
    - Ex: distinguish field images vs stadium and street images with a grass detector for the World Cup dataset

SOM Clustering Example with PSO Training - Mozilla Firefox

http://localhost:8080/cbmi-demo/cv/NewsData.do

ISWC 2008: Wednesday, 29 October 2008

CC BY-NC-SA

23

## Future Work

- Data Modeling
  - Events Model
- End-user Interfaces
  - Yahoo! Search BOSS
- Data Quality
  - Named Entity Recognition (Calais),
  - Disambiguation algorithms (SemanticProxy),
  - Visual clustering, Video segmentation

ISWC 2008: Wednesday, 29 October 2008

CC BY-NC-SA

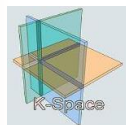
24

# Credits

- Datasets:



- People:



- More info: <http://newsml.cwi.nl>

K-Space NewsML Semantic news demonstrator

[search](#)
[browse](#)
[local view](#)

[login](#)
[help](#)
[English](#)

results : Lyon

TextNewsItem <- title <- match (3)

view all

TextNewsItem <- headline <- match (16)

view all

Lyon welcome Ribery interest

Lyon welcome Ribery interest-LYON, France, June 10, 2006 (AFP) - Lyon confirmed Saturday night their interest in signing highly-rated Marseille forward Franck Ribery amid morning reports the player said he wants to join the French champions.The Saturday edition of newspaper Le Parisien-Aujourd'hui had claimed that Ribery wants to join Lyon but the champions are

Lyon close the net on Ribery

Lyon close the net on Ribery-LYON, France, June 22, 2006 (AFP) - Lyon stepped up the pressure on Marseille to allow star forward Franck Ribery to move to the champions before the start of next season by claiming Thursday they expect to thrash out a deal after the World Cup.Ribery has been quoted in the press over the last few weeks saying he wants to join Lyon for next

Lyon's Frau joins PSG

Lyon's Frau joins PSG-PARIS, July 12, 2006 (AFP) - Lyon striker Pierre-Alain Frau signed for Paris Saint Germain on Wednesday.Frau, 26, agreed a four-year contract with PSG after deciding to quit the French champions in search of regular first-team football.The former Sochaux star, who joined Gerard Houllier side's in 2004, had struggled to break into the Lyon team

PhotoNewsItem <- slugline <- match (6)

view all

Par805798\_2.rdf

JEAN-PHILIPPE\_KSIAZEK

Par805806\_2.rdf

JEAN-PHILIPPE\_KSIAZEK

Par811606\_2.rdf

JEAN-PHILIPPE\_KSIAZEK

Par827009\_2.rdf

MARTIN\_BUREAU

Par827013\_2.rdf

MARTIN\_BUREAU

Par827018\_2.rdf

MARTIN\_BUREAU

PhotoNewsItem <- locCreated <- Feature <- alternateName <- match (24)

view all

PhotoNewsItem <- person <- natural-person-denomination <- NAME <- match (1)

view all

PhotoNewsItem <- person <- FootballPlayer110101634 <- currentclub <- match (1)

view all

ISWC 2008: Wednesday, 29 October 2008

26

13

ISWC 2008: Wednesday, 29 October 2008

27

ISWC 2008: Wednesday, 29 October 2008

28



browser - Mozilla Firefox

http://newsml.cwi.nl/explore/facet44

Search News RDFa Highlight Raphaël Troncy Melanie CWI K-Space NewsML FP7, Call 3 W3C Conférences Planet RDF ramm.x (RDFa-deploy...) ShapeShift.TV

Google

Rechercher

Mes favoris PageRank Traduire Envoyer à Paramètres

soccer -1000 FBL-WC2006-MATCH84-ITA-FRA-FANS World Cup warm-up result: Sweden 1-0 Chile 1  
labour FBL-WC2006-MATCH84-FRA-ITA-FAN World Cup warm-up result: Liechtenstein 0 Togo 1  
FBL-WC2006-MATCH84-ITA-FRA World Cup statistics  
FBL-WC2006-FRA-POR-FANS Swiss football results  
FBL-WC2006-FRANCE Real still want Kaika and Van  
FBL-WC2006-MATCH84-ITA-FRA

JACK\_GUEZ  
\*\*no label  
OLIVER\_LABAN-MATTEI  
FILIPPO\_MONTEFORTE  
KAMBOU\_SIA  
GORDIAN\_CIVAN  
French football fans celebrate on the Champs Elysees avenue in Paris after the World Cup 2006 semi-final football match France vs Portugal 05 July 2006 at the Allianz Arena stadium in Munich, Germany France won 1-0. AFP PHOTO JACK GUEZ  
Italian football fans celebrate in

CATEGORIES

NewsItem

Concept

soccer

Two teams of eleven try to kick or head a ball into the opponents goal over a match usually of 90 minutes

RESULTS

Group by: NewsItem

Munich (locCreated), 16 items

Views: images snippet map timeline

Plan Satellite Mixte

miné

ISWC 2008: Wednesday, 29 October 2008

CC BY NC SA

29

K-Space NewsML Semantic news demonstrator

search browse local view

display cell format

local view

G8Video\_002.rdf

http://newsml.cwi.nl/APP/G8Video/G8Video\_002.rdf

ATEINTE A L'HONNEUR

SUSPENSION

ZIDANE PENDANT AU MOINS 2 MATCHS

11 09:52 20:09 4:11

Links

- permanent link
- annotate
- tcin
- tcout

Property	Value
SecondOfMoreBroadcast	•
audiencePresent	•
availableMaterial	• [COM] - FPVDB06071308 01/01 - Stat. Num.: Numérisé et en ligne - TC IN: 00:13:37:14 TC OUT: 00:19:29:23 - Format: MPEG2 - Définition: CIS - Son: STEREO - Coul.: COULEUR - Filère: FP - Type Mat.: COP - Durée: 00:37:11 - Stat Vers.: Versé Localisation: NAP (02/08/2006) -
broadcastGoal	•
broadcastInformation	• 13/07/2006 - type date: Diffusé - heure: 20:16:01 - canal 2eme chaine (A2) - ext. géo. Nationale
broadcastMode	•
collectionTitle	• 20 heures le journal
corpus	•
descriptors	• DET: alpinisme ; DET: accident ; DET: expédition ; DET: cadavre ; DET: recherche ; DET: retour (rattachement) ; DEL: Népal ; DEL: Himalaya ;
duration	• 00:05:27
eventDate	•
foreignOriginalVersion	•
generic	• JOU Jacquier, Gilles ; PAR Koenig, Serge ; PAR Baud, Alain ;
genre	• Journal télévisé ; Reportage ;
id	• 3129199001020
material	• [COM] - FPVDB06071308 01/01 - Stat. Num.: Numérisé et en ligne - TC IN: 00:13:37:14 TC OUT: 00:19:29:23 - Format: MPEG2 - Définition: CIS - Son: STEREO - Coul.: COULEUR - Filère: FP - Type Mat.: COP - Durée: 00:37:11 - Stat Vers.: Versé Localisation: NAP (02/08/2006) -
materialID	• FPVDB06071308 01

ISWC 2008: Wednesday, 29 October 2008

CC BY NC SA

30