



# JRC Optima Activity

THE META PRIZE IS PROUDLY AWARDED TO

# JRC Optima Activity Language Technology Language Technology

*for outstanding use of Language Technology  
to fight information overflow and to overcome language barriers.*

*The META Technology Council recognises that this work is an important cornerstone on  
the way to enabling the European Multilingual Information Society.*

On behalf of the META Technology Council

**Hans Uszkoreit**  
Co-ordinator of META-NET  
DFKI, Germany

On behalf of the META Technology Council

**Georg Rehm**  
Network Manager of META-NET  
DFKI, Germany

[www.meta-net.eu](http://www.meta-net.eu)

# Joint Research Centre

The European Commission's in-house science service

[www.jrc.ec.europa.eu](http://www.jrc.ec.europa.eu)

*Serving society  
Stimulating innovation  
Supporting legislation*



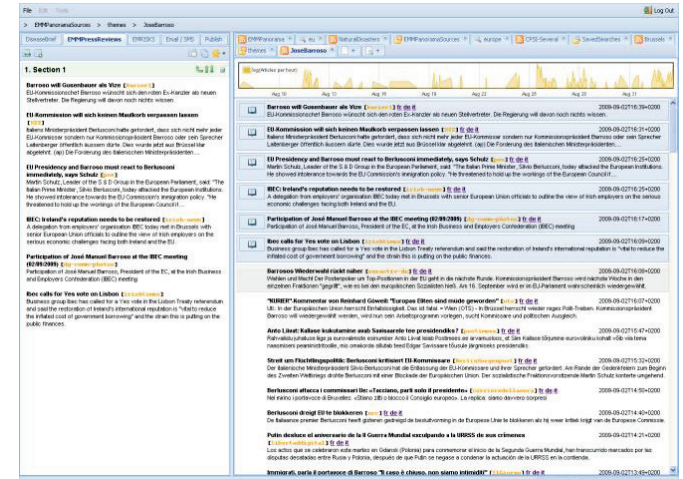


# Europe Media Monitor (EMM)



Erik van der Goot & Optima team

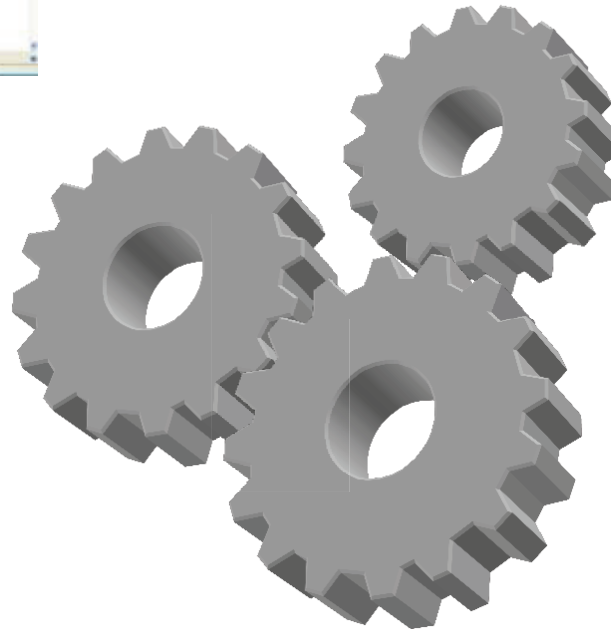
# EMM System Overview



EMM Newsbrief  
EMM NewsExplorer  
MediSys

NewsDesk Service  
Editorial Interface

EMM Open Source  
Monitoring Engine



# Some EMM statistics



- Gathers approximately **150,000 new news articles per day**
- In **~ 50 languages**
- **From ~ 3750 news portals** world-wide (roughly 10.000 pages/RSS feeds)
  
- Classifies all news according to **hundreds of subjects and countries.**
- (~1400 category definitions, ~35.000 key word patterns and pattern combinations)
- **24 / 7:** Runs 24 hours per day, 7 days a week.
  
- Started in 2002
- **Up to 1.5 Million hits per day** for public websites (EMM NewsBrief, MediSys, NewsExplorer)
- But more importantly, 25.000 visitors/day on the public sites
  
- Developed, built, maintained and run by the JRC.



# The importance of multilingual information gathering



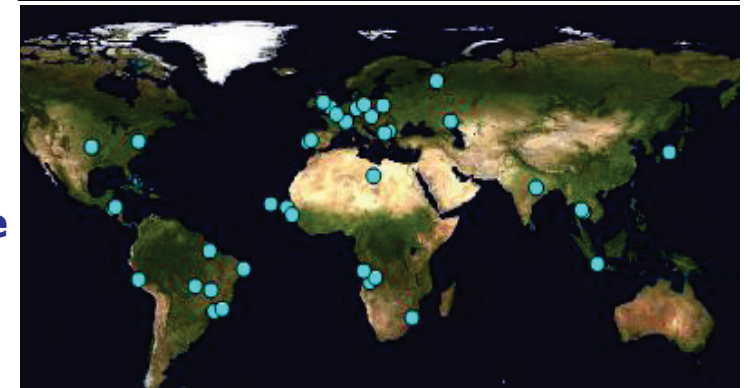
**Italian - German**



**English - French**



**Spanish  
- Portuguese**





- Various **multilingual parallel corpora** in up to 22 languages (23 official EU languages minus Irish), e.g.:
  - JRC-Acquis (total of over 1 billion words; available since May 2006);
  - DGT-TM (total of 57 million translation units; developed by DGT; since 2007, extended in May 2012);
- **Multilingual multi-label categorisation software** JRC EuroVoc indexer:
  - Trained for 22 languages (released in May 2012)
- **Multilingual named entity resource JRC-Names:**
  - 27 different scripts and many more languages;  
Up to 413 spellings for a single name;
  - Collected by analysing up to 100,000 news articles in up to 20 languages since 2004; partially augmented through Wikipedia.



## Muammar Gaddafi

معمر القذافي; Mouammar Kadhafi; Muammar al-Gaddafi; Moammar Gadhafi; Muammar Kadhafi; Muammar Gheddafi; Muammar Kaddafi; Муамар Кадафи; معمر قذافي; Muammer Kaddafi; Muamar Gadafi; Moamerja Gadafi; Moamar Gaddafi; Muammar Kadafi; Muamar el Gadafi; Муамар Каддафи; Muammar Khadaffi; Muammar el Gaddafi; Moammar Gaddafi; Muammar Gadafi; Moammar Kadhafi; Moamer Gadafi; Muamar Kadafi; Moamer Kadhafi; Mouammar Khadafi; Muammar Gadaffi; Муаммар Каддафи; Muammar Khaddafi; Muammar al Gaddafi; Muhammad Gheddafi; Muammar Qaddafi; Moammar Gadaffi; Muamar al Gadafi; Muammar Khadafi; Muamar Kadhafi; Moamer Gadhafi; Muammar Kadaffi; Muammar al-Ghadhafi; Muammar Gadhafi; Mouammar Kaddafi; Muammar Gadafy; Muammar el-Qaddafi; Muamar Khadafi; Muamar Gaddafi; Muammar al-Gadafi; Moamer Gaddafi; Muammar Ghaddafi; Muammar al-Qadhafi; Muammar Ghadhafi; Mouammar El Kadhafi; Muammar Qadhafi; Muammer Gaddafi; Muammer Gadaffi; Moamar Gadafi; Moammar Gheddafi; Mouamar Kadhafi; Moamar Kadhafi; Mouamar Khadafi; Moamer Qadhafi; Moammar al-Kadhafi; Moammar Qaddafi; Moamer Kadaffi; Moammar Khadafi; Moammar al-Qadhafi; Muammar Ghadaffi; Moeammar Kadhafi; Muammar Al Gaddafi; Muammar el-Gaddafi; Mummar Gaddafi; Muammar al-Qathafi; Muammar Gadaffi; Muammar al-Qaddafi; Muammar Al-Qadhafi; Muammar al-Kadhafi; Muammar Al-Kaddafi; Moammar Khadaffi; Moammar Kadafi; Muammar Al-Gaddafi; Moammar Ghadafi; Mouammar Al Kadhafi; Moammar Kaddafi; Mouammar El-Kadhafi; Moammar Khaddafi; Muammar el Gadafi; Muhammad Gaddafi; Moamer Gadhafi; Muammar al-Gadhafi; Муамар Кадафи; Muamar El Gadafi; Moammar Qadhafi; Muammer el Gaddafi; Mouhammar Kadhafi; Muamar Gadaffi; Moamar Gadhafi; Muammar Khadafy; Muammar al-Kadafi; Muammar al Ghadhafi; Muhamar Kadafi; Muammar Abu Minyar al-Gaddafi;