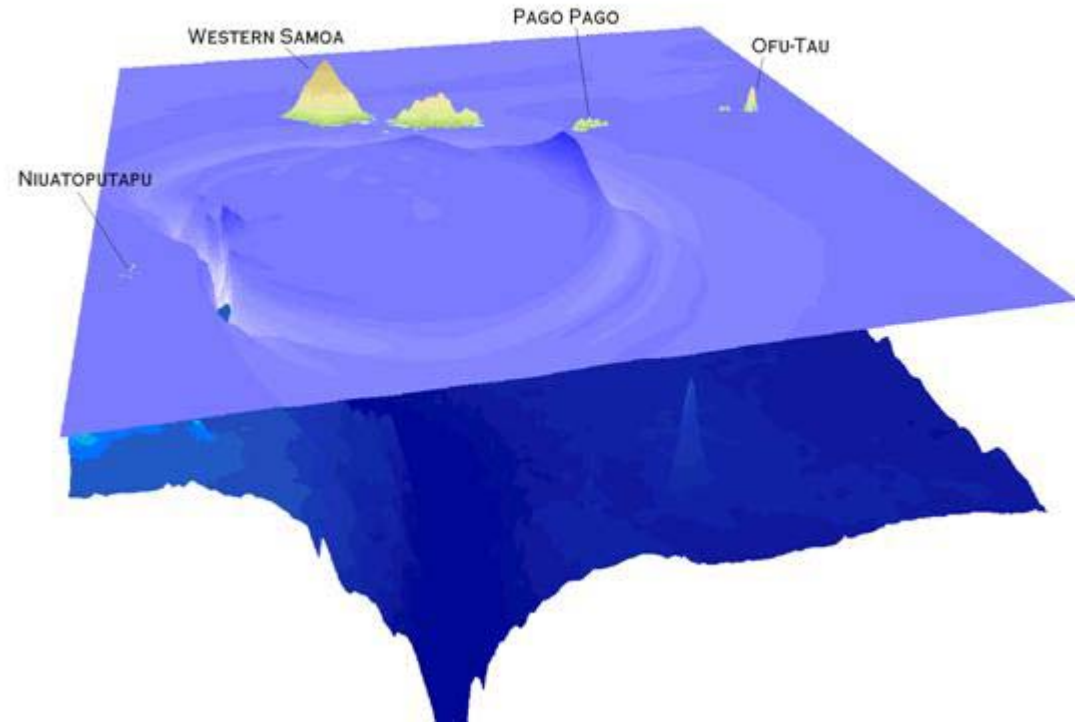


Open Source Information Mining – “It’s all on the internet”

Stephan Lechner
- Director -
European Commission,
Joint Research Centre.

<http://ipsc.jrc.ec.europa.eu/>

JRC Tsunami Model



In which way is a Tsunami model based on internet information?

Media

- Print
- Radio
- TV
- Internet (also in push mode or mobile)



Where do we get the information first?

Text news
combine speed, accuracy and coverage best!

US: English (Spanish?)

EU: 22 Languages,

Greece writes Greek, Bulgaria writes cyrillic

Global: Russian, Arabic, Mandarin ...

Additional Features:

- **Keyword detection algorithms**
- **Fusion with other sources**
- **Aggregation and presentation**
- **Expert post processing**



Аа Бб	И	α	Α	Alpha
Ёё Жж		β	Β	Beta
Кк Лл		γ	Γ	Gamma
Пп Рр Сс Тт Уу Фф		δ	Δ	Delta
Хх Цц Чч Шш Щщ				
Ъъ Ыы Ьь Ээ Юю				
Яя (İi Əə Vv ƦƦ)				

Global text mining is easier said than done!

The screenshot shows the MedISys website interface within a Windows Internet Explorer browser. The browser's address bar displays the URL: <http://medusa.jrc.it/medisys/alertedition/all/SwineFlu.html>. The website header includes the EMM MedISys logo and navigation links for Home, Diseases, Bioterrorism, and Other. The main content area is titled "Swine Flu" and features a "Latest News About - Swine Flu" section. The news feed includes several articles with their respective source logos, titles, and timestamps. Annotations with arrows point to specific features: "Update every 10 minutes" points to the "UPDATED EVERY 10 MINUTES, 24 HOURS PER DAY." text; "German" points to a German article titled "Bakterienverseuchung: Geldautomaten so dreckig wie öffentliche Toiletten"; "Bulgarian" points to a Bulgarian article titled "Млад мъж от Велинград със съмнения за свински грип"; "Norwegian" points to a Norwegian article titled "Allvarliga fall av sv"; and "Map based on press reports" points to a map of Europe in the bottom right corner, which is color-coded to show the geographical distribution of swine flu cases.

MedISys - Windows Internet Explorer
http://medusa.jrc.it/medisys/alertedition/all/SwineFlu.html

Home | Diseases | Bioterrorism | Other

Swine Flu
UPDATED EVERY 10 MINUTES, 24 HOURS PER DAY.

Home | Latest News About - Swine Flu

Update every 10 minutes

German

Bulgarian

Norwegian

Different source article types

Map based on press reports

Situation assessment far faster than official statistics



GDACS: Global Disaster Alert and Coordination
for more than 10.000 first responders

<http://www.gdacs.org/>



EMM News Brief: internet news
analysis in 22 languages

<http://press.jrc.it>

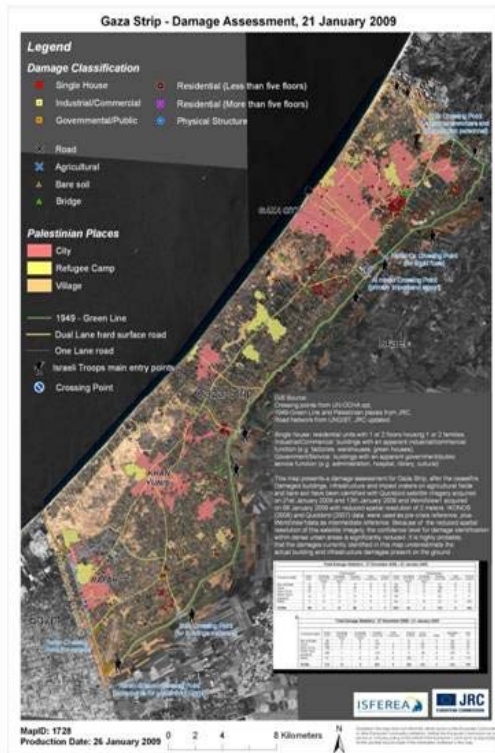


EMM MediSys: Disease Spread
Monitoring and forecast

<http://medusa.jrc.it/>

Gaza Strip January 2009:

- How much damage was done where?



Situation map based on automated image analysis

Georgia / South Ossetia 2008:

- What was the real target?



TV news



Reality

Independent assessment for clear diplomatic briefings

Event

Sat 27 Feb 2010, 3:34h local time

- 01:34h New York → night
- 08:34h Brussels → new shift
- 15:34h Tokyo
- 22:34h (Friday) LA → night

Magnitude: 8.8M (very strong)

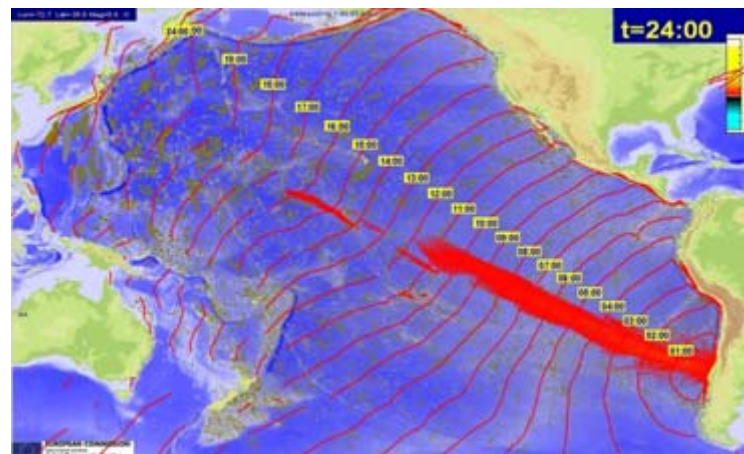
- Depth: 35km (shallow)
- Tsunami was generated
- Several strong aftershocks

Location

- 113km from Concepción
 - 500000 people within 100km
 - 3 million people within 200km
- Mountain area
 - Landslide risk, dams

GDACS system response

- **after 23 min** Detection of earthquake. Tsunami wave height and travel times calculated
 - **after 43 min** EQ stronger than originally estimated
 - **after 53 min** GDACS **Red Tsunami Alert**;
Online report
SMS, Email, Fax alert sent to 11000 users.
- JRC Tsunami model confirms
no major problem for Pacific Islands



JRC Tsunami model

- Coastal settlements
- Underwater surface
- Steepness of coastline
- Elevation of islands
- Pre-calculated propagation



- **Open source data mining is an important input for situation assessment**
- **Data need to be fused with other sources (satellite images, model calculation, population databases, ...)**
- **Pre-calculation wherever possible**
- **Intelligent and automated impact analysis: Better fast than perfect!**
- **New media will become more interesting (“Community remote sensing”)**

Intelligentu mechanisms, not intelligencee mechanisms!