Open Data meets Big Data

Max Craglia, Sven Schade, Anders Friis
European Commission
Joint Research Centre

www.jrc.ec.europa.eu

Serving society
Stimulating innovation
Supporting legislation

JRC is Technical Coordinator of INSPIRE

"Infrastructure for Spatial Information in the European Community"

Institutional framework	Technical standards
Fundamental data sets	Data Services

Distributed infrastructure

28 countries,
23 languages

General **rules** for establishment

Environment

34 Spatial Data Themes

Entry into force 15/5/2007

Set of European and national Legal Acts and their coordinated implementation

INSPIRE is a legal framework

- The INSPIRE Directive lays down general rules to establish an **Infrastructure for Spatial Information in Europe** for environmental policies and policies which may affect the environment
- INSPIRE is built on the SDIs established and operated by the Member States
- Implementing Rules (i.e. legislation)
  1. Metadata
  2. Interoperability of spatial data sets and services
  3. Services (discovery, view, download, transform, invoke)
  4. Data and Service sharing (policy)
  5. Monitoring & reporting

INSPIRE thematic scope

Annex I

1. Coordinate reference systems
2. Geographical grid systems
3. Geographical names
4. Administrative units
5. Addresses
6. Cadastral parcels
7. Transport networks
8. Hydrography
9. Protected sites

Annex II

1. Elevation
2. Land cover
3. Ortho-imagery
4. Geology

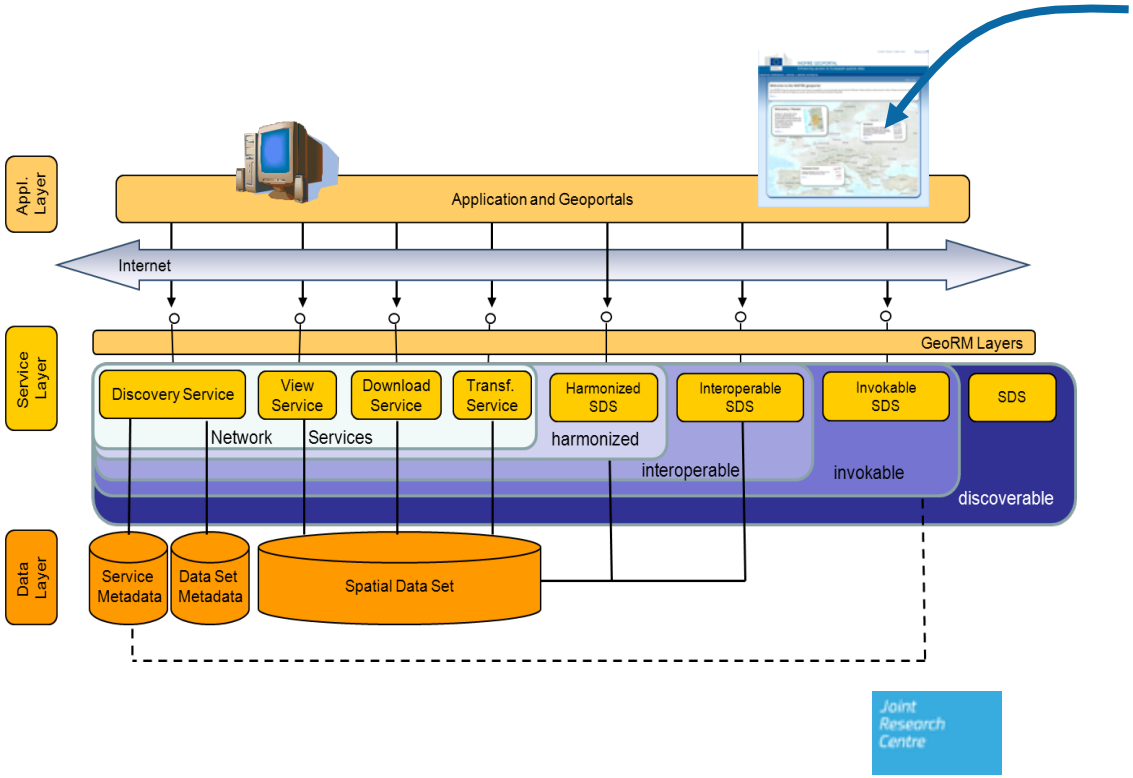
Annex III

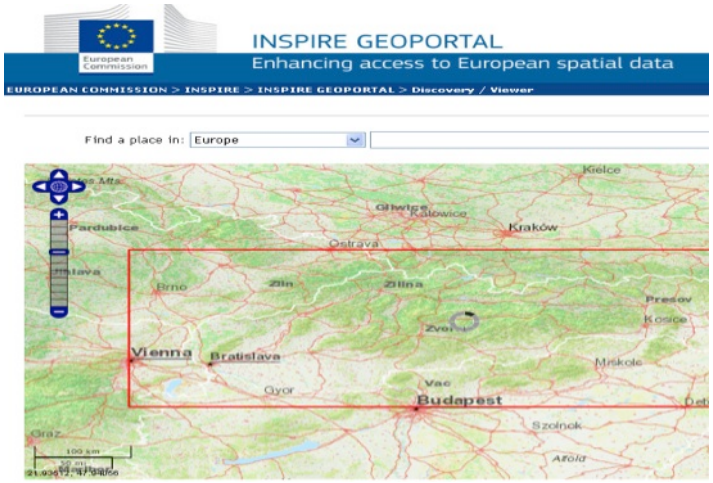
1. Statistical units
2. Buildings
3. Soil
4. Land use
5. Human health and safety
6. Utility and governmental services
7. Environmental monitoring facilities
8. Production and industrial facilities
9. Agricultural and aquaculture facilities
10. Population distribution – demography
11. Area management/ restriction/regulation zones & reporting units
12. Natural risk zones
13. Atmospheric conditions
14. Meteorological geographical features
15. Oceanographic geographical features
16. Sea regions
17. Bio-geographical regions
18. Habitats and biotopes
19. Species distribution
20. Energy Resources
21. Mineral resources

INSPIRE Geoportal

Central access point to the INSPIRE infrastructure and resources (>300.000)

→ **"The face" of INSPIRE**





- Connection to all MS network services

→ **cross-border data discovery and visualisation**

→ **support to European policy making**

Joint Research Centre

Extending INSPIRE to other policies


EU-wide Multi-Modal Travel Information

EU-wide Real-Time Traffic information


Free safety-related minimum Traffic Info

Interoperable EU-wide eCall


Information & Reservation systems for Truck Parking

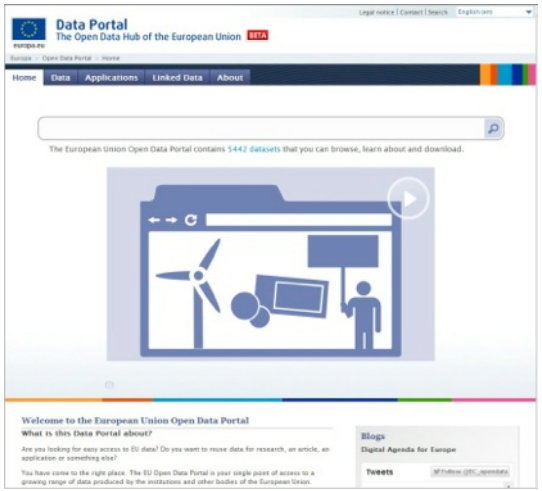
SOURCE: European Commission

INSPIRE meets Open Data (PSI)

Open Data initiatives around the world

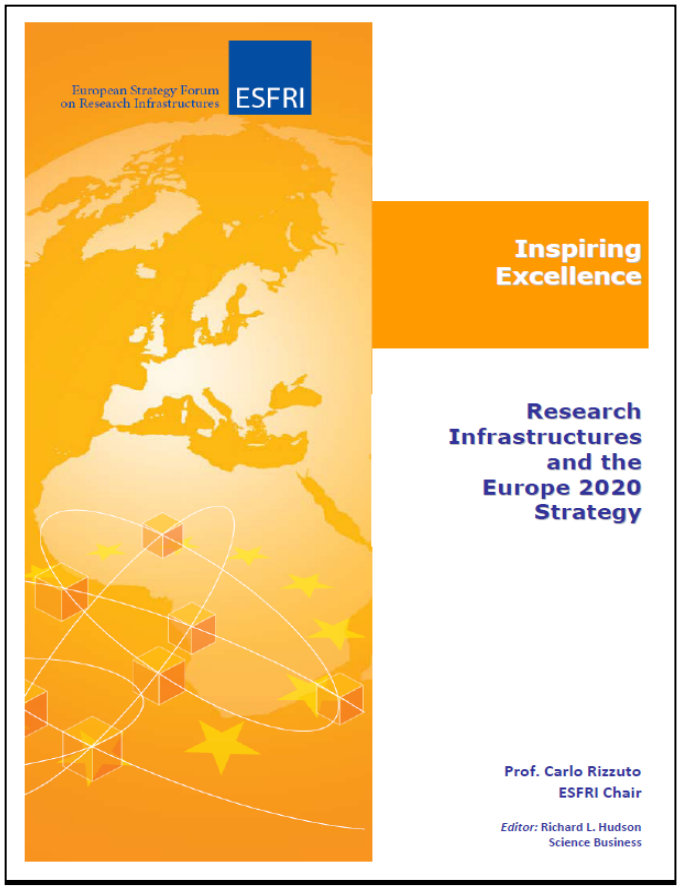
In EU, Open Data Strategy

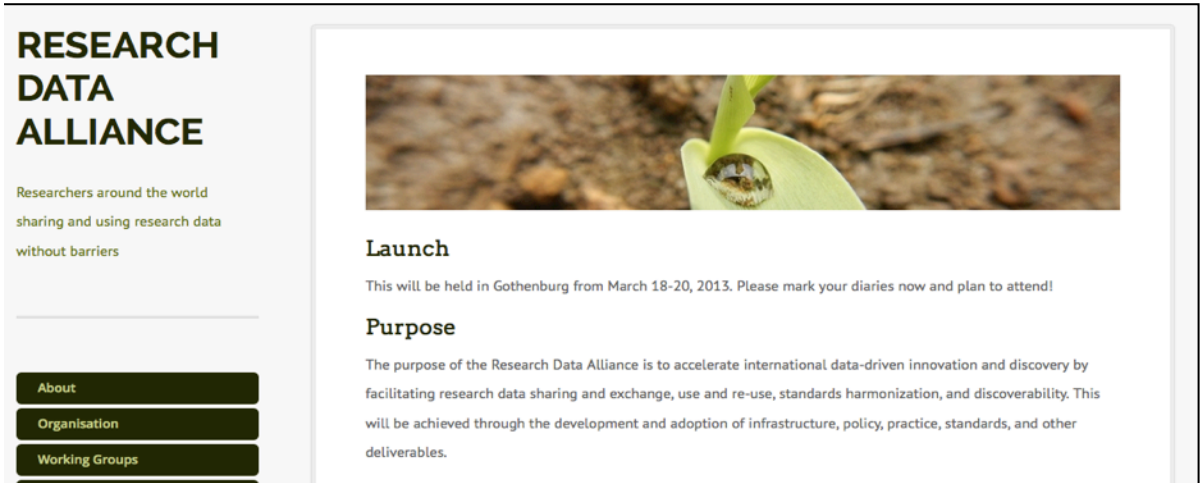
- Communication on Open Data (COM(2011)882)
- A revision of the Decision governing the re-use of Commission's own information (2011/833/EU)
- Revision of the Directive on the re-use of public sector information (2013/37/EU)
- Open data-portals
- EU Open Data portal
- pan-European portal



Open Research Data

- Open Access put forward by the Commission for scientific publications/ data within Horizon 2020
  - For a subset of thematic areas, funded projects shall include data management plans and ensure open access
- JRC has proposed a strategy for online dissemination of JRC research activities and scientific results

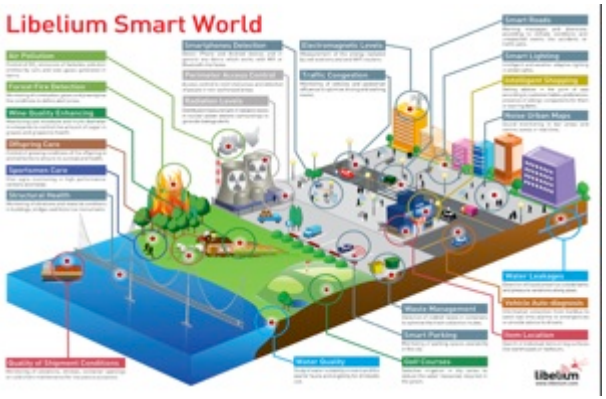




New Data Sources = Volume and Variety

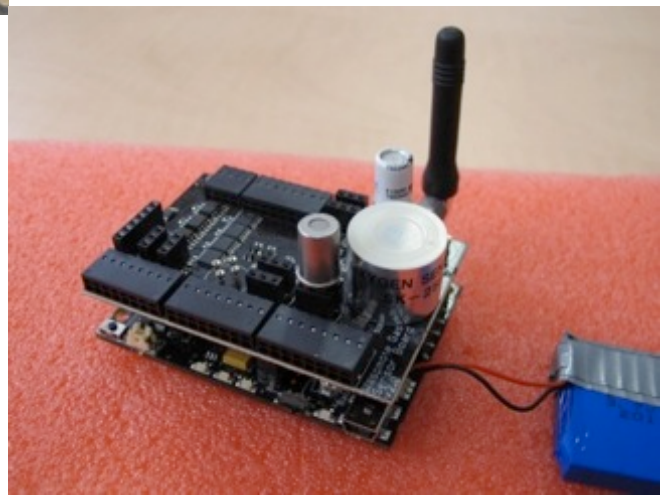
- 2 Terabytes per day once in operations
- Raises issues of where and how to store the data, how to provide easy and rapid access, how to process and analysis, and maintain over time.

Massive diffusion of cheap sensors provides new opportunities and challenges



Publiclaboratory.com

App Store Play Store



- Mobile phones sensing better in some fields than others (e.g. noise)
- Drones some limitations relating to regulatory framework
- Waspmotes need programming and issues of calibration and response time but opportunities high..

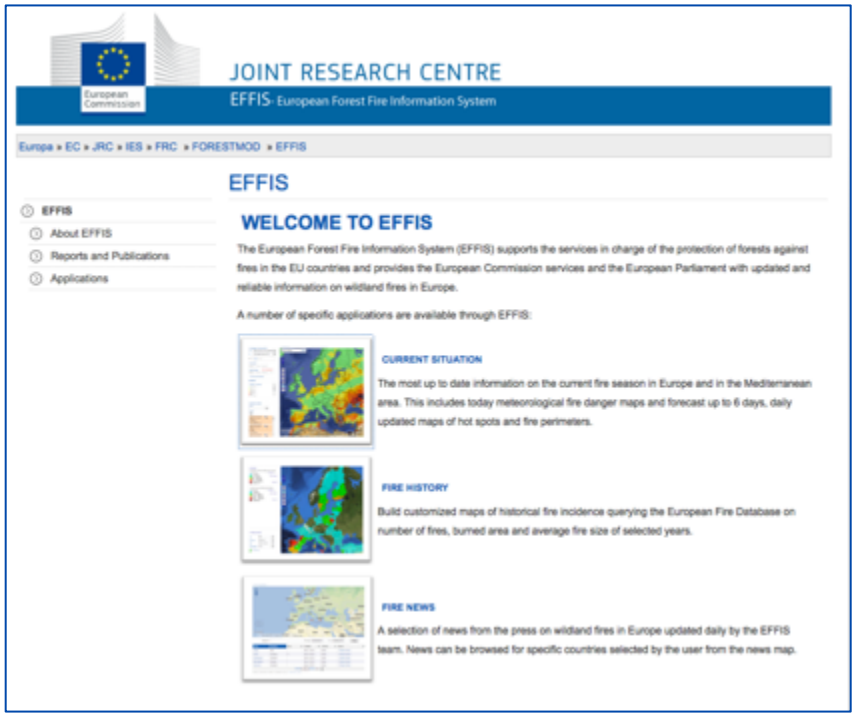
European Commission

Combining Remote Sensing with Social Sensing

Joint Research Centre

Large scale experiment at JRC to assess quality of social network data

- Project at JRC 2010-12 to develop automatic workflow and extract and assess data from Flickr and Twitter related to forest fires and compare to official data from European Forest Fire Information System managed by JRC.



European Commission

FOREST FIRE CASE STUDY

RETRIEVAL
Collect the volunteered information from data publicly available on social media platforms as text messages and photo tags.

GEOCODING
Only a fraction of the VGI is explicitly geocoded with coordinates, while the majority is implicitly geocoded with place names.
We extract the place names and their spatial reference and geocode them to make these available for analysis.

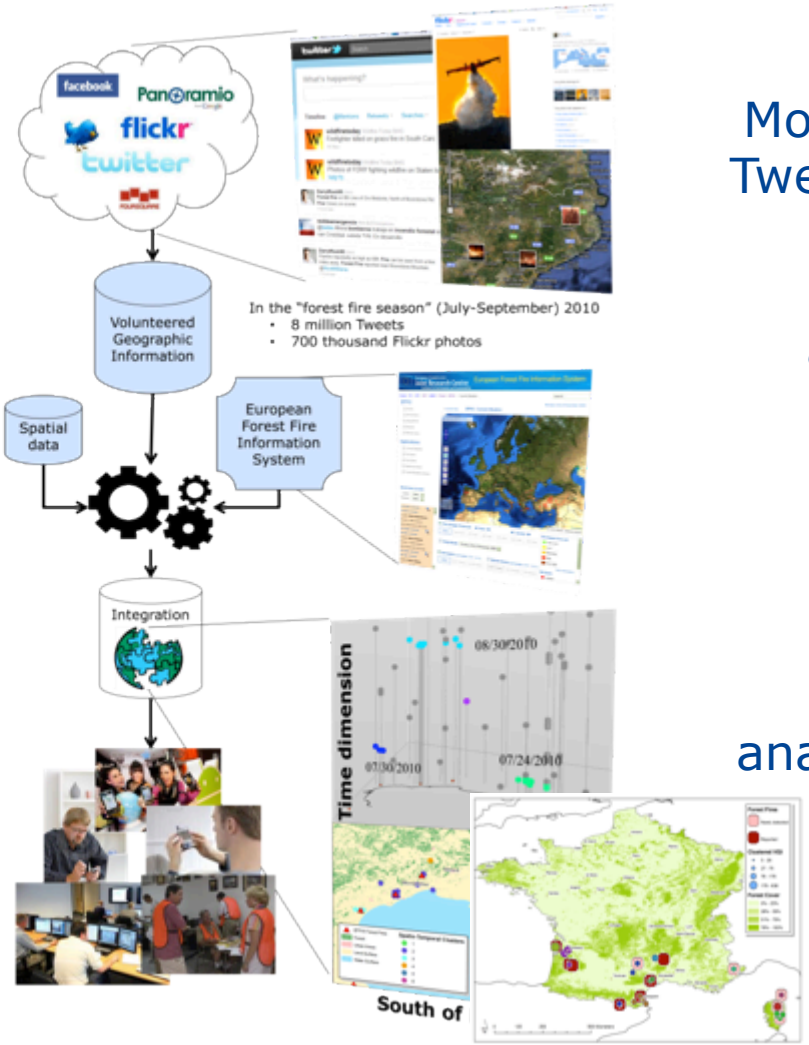
QUALITY ASSESSMENT
The crucial aspect for using volunteered information in the context of crisis management is to assess its quality.
For this task, we explore several methods, including:
- Syntactical validation of the data
- Cross-referencing with other data
- Spatial and temporal clustering

The two main metrics for assessing the quality will be the **credibility** and the **relevance** of data with respect to a specific event.

INTEGRATION
The quality-assessed VGI needs to be integrated within the disaster management process and other official data such as remote sensing images and spatial data infrastructures.

DISSEMINATION & EVALUATION
The two main target audiences for the results are:
1. The general public
2. Decision makers

They have very different needs, that will be addressed in a final step, a task- and user-centered design process.
The added value of the VGI will be rigorously evaluated.



More than 20million Tweets and 1 million Flickr images retrieved and analysed for fires South of France

Spatio-temporal clustering and analysis shows 80% of fires correctly detected

Joint Research Centre

From Data to Processes

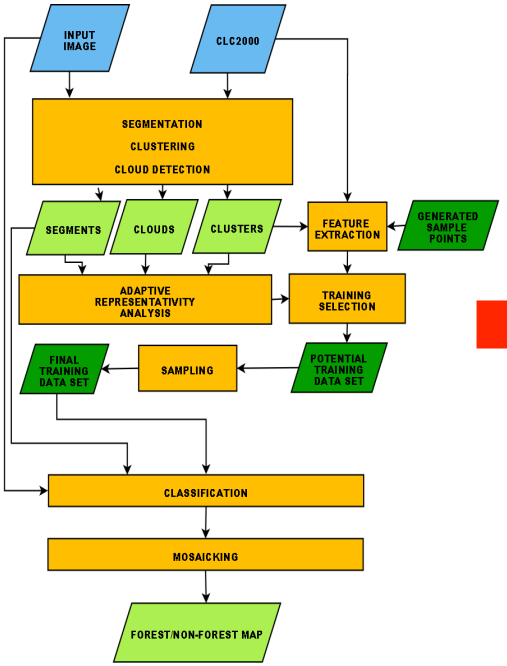
If you have BIG Data from multiple sources, you cannot move the data for processing, need to move the analysis and processing to the data.

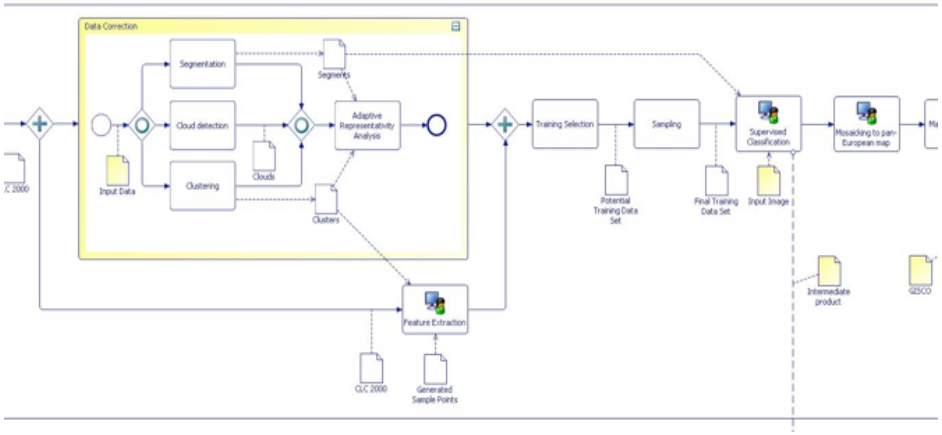
To support multi-disciplinary research we need also to develop a shared understanding of *what do you do with the data*?

How do you frame a problem and possible solution according to different disciplinary approaches.

This quest requires to describe not just the data, but also processes or workflows, leading to new executable web services that are understood across disciplines.

Workflows described in BPMN





Simple analytical functionalities can be expressed we web services. Executable workflows and models can then be integrated and interfaced with other services to develop chains.

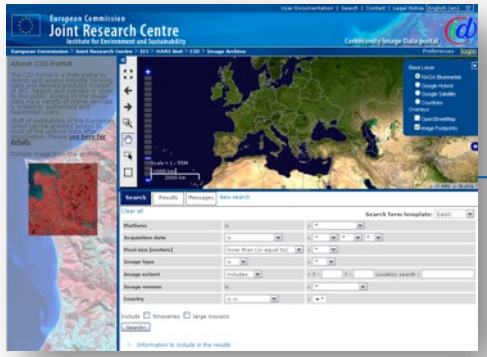
Building blocks of the Open Data project at JRC

Reference Data and Services Infrastructure (RDSI)

Data Catalogue

MIDAS

Model Inventory Database and Access service (MIDAS)



CID

Community Image Data Portal (CID)

Operational Services

JRC Knowledge Portal

JRC Facts & Figures Service

EC OpenData Portal

INSPIRE@EC

EU Climate Adapt Platform

Joint Research Centre

Conclusions and Points for Discussion

- JRC is starting to hit the Big Data Issues
- Task Force set up by our DG on **Big Data and Computing Strategy for Scientific Support**
- We welcome exchange with CERN on some key issues for us:

1) Your experience on Data Management Principles and Plans covering the full data life cycle e.g., how to deal with versioning, citation, permanent IDs, publishing, and preservation
2) Your experience on data storage and processing: in particular advice on different platforms Grid, HPC, Hadoop, Cloud
3) Your experience on Big Data analytics and data visualisation

Thank you for your kind attention

massimo.craglia@jrc.ec.europa.eu