

---

***EUROGI***  
***Policy Position Paper on***  
***Big Data***

***Domen Mongus***

***Member of EUROGI Executive Committee or (General Board)***

***Domen.mongus@um.si***

***INSPIRE GWF 2015***

***Lisbon, Portugal, 25(28) May2015***



# What is Big Data?

---

**Volume** is a primary attribute of Big Data, e.g. those data generated by Earth Observations.

**Velocity** describes dynamic and rapid data generation, where its real-time processing is required.

**Variety** of data sources and formats that includes unstructured and semi-structured information.

**Value** of Big Data raises from the knowledge contained within the data that holds the potential to significantly improve our understanding of the world we live in.



# Examples of sources of Big Data

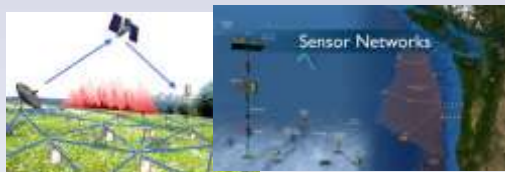


**Social media**



**Scientific instruments**

**Mobile devices**



**Sensor systems**



# GI and Big Data

---

## Volume

- Earth observations
- Monitoring systems

## Velocity

- Tracking systems
- In-situ monitoring

## Variety

- GI provides a key basis for integration

## Value

- Environmental monitoring, crisis management, and many more...!



# GI and Big Data

---

Big data is often:

- Automatically generated
- Not designed to be interoperable with other data which would be incorporated into Big Data analysis
- It is not certain that you would necessarily get value form Big Data analysis



# What is Linked Data?

---

A global graph connecting raw data to make them accessible and re-usable

- by linking records or data sets using controlled semantics (i.e. Semantic web)

Semantically linking Geographic Data would add value to geographic analysis

- Would that make new Big Data?







# Why GI community needs Big Data policy?

---

## Data-driven decision making

- make inform decisions based on the knowledge and rather than relying on intuition or personal experience.

## Enable spatial data mining

## Adding value to geographic information





# Points of discussion for Big Data

---

Should GI have a central role in creating the ability to integrate data from diverse sources?

Are there specific standards/protocols which should apply to GI in specific data fields, so that seamless integration would take place?

Which would be the priority data fields?



# Points of discussion for Liked Data

---

How should location data best be represented semantically

Are there ontology tools which best represent location information?

Wheatear every physically object should/could have its own ID and web address?

