The 10th International Scientific Conference
“DEFENSE RESOURCES MANAGEMENT
IN THE 21st CENTURY”
Braşov, November 13th 2015

CURRENT CHALLENGES IN DESIGN, DEVELOPMENT AND IMPLEMENTATION OF GEOSPATIAL INFORMATION SYSTEMS

Ltc. Eng. Ion Nedelcu

Ministry of National Defence / Romanian Space Agency, Bucharest, Romania

Abstract:

The value of quality geospatial information has been already proved to lay at the heart of decision making processes having impact in many fields of activity. While thiking about environment, social, climate change or security problems, the geospatial information is subject to new and more effective data acquisition, high performance computing or storage and archiving systems. However, higher and higher demand has leaded to more and more advanced systems providing access to huge quantities of data which today are estimed to reach very quickly rates of tera bytes per day. To face this obvious challenges, raised mainly by the quantity and the high speed, both technical and organizational aspects need to be considered. After analysizing the current context in which the new sources of geospatial data are presented togther with their volume, variety and speed characteristics, some technical, legal and organizational challenges are emphasized ans possible ways to treat them are also discussed.

Key words: geospatial information, geographic location, big data, geospatial analytics, Earth observation satellites, spatial data infrastructure

1. Geospatial big data and geospatial information

Traditionally, geospatial data can be categorized into raster, grid and vector data (fig.1). The raster data include geo-images typically obtained by aerial vehicles, security cameras, and satellites. Recently, the military is collecting huge amounts of raster data by utilizing drones, and the satellites keep providing the remote sensing data of the Earth. The raster data is made available to users directly or by means of map services like Google Earth.

Grid data is typically represented by elevation data used to express the height of the features above a reference surface. One of the most effective ways for collecting elevation data is by means of LIDAR - acronym for Light Detection and Ranging, a remote sensing technique that uses laser pulses to determine elevation with high accuracy, usually from an aerial survey.

The vector data consists of points, lines, and polygons that are modelling various data sources. For example, lines and polygons correspond to roads in OpenStreetMap[1], a collaborative project to create a free editable map of the world.

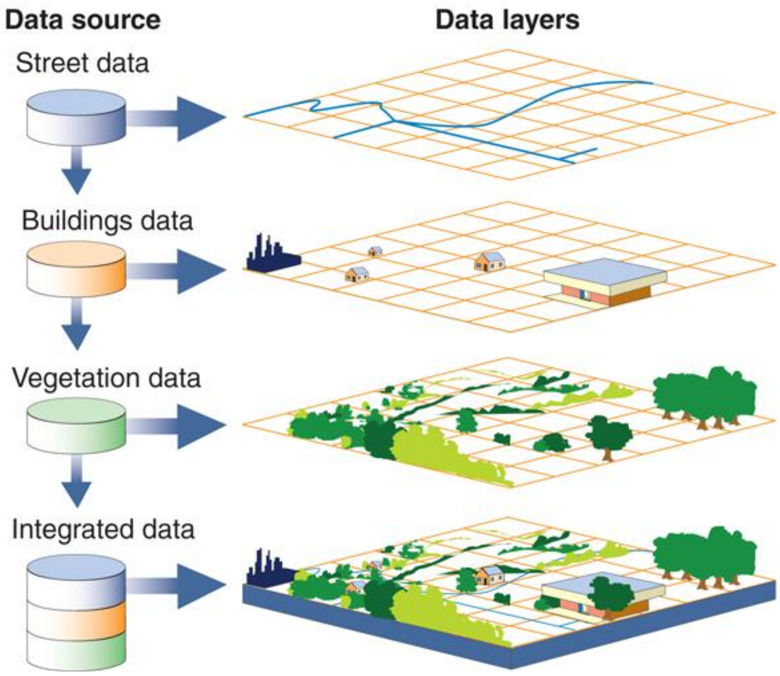

Fig. 1 Geospatial data types

1.1 Geospatial big data

With the advancements of sensor and communication technologies, new sources of geospatial data are emerging. Examples can be traffic detectors on the roads, electrical grids or environmental sensors for measuring air quality (fig. 2). These sensors are usually connected through wired or wireless communications in sensor networks.

Also, mobile devices are emerging sources for huge quantities of geo-referenced data. Smart phones became versatile device for recording trajectories or other location aware data.

Another source considered to be significantly emerging the last years is the crowdsourcing, a term referring to a process of obtaining services, ideas, or content by requesting contributions from a large group of people, and especially from an online community, rather than from traditional suppliers [2].

The geospatial data collected from all these sources is characterized by 3 properties - volume, velocity, and variety – the big data characteristics, the velocity being the most obvious one when talking about the data coming from these new, emerging sources. For this data, instead of storing it analysing it later, it is a clear need to look at the data on the fly and make decisions in the shortest time. This is why the data analytics capacity and performance will play a more important, even crucial role.

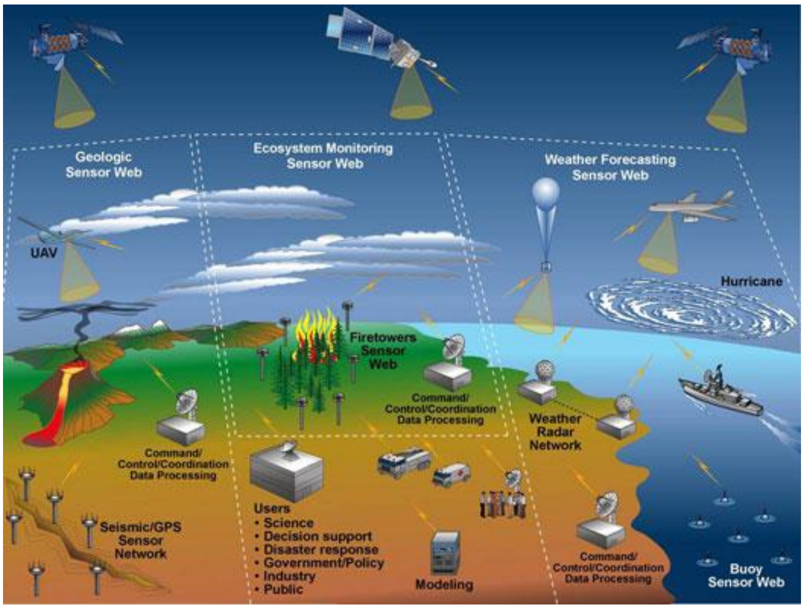

Fig. 2 Sensor networks [3]

1.2 The wide use of geospatial data

In a recent post, the European Space Agency shows how Copernicus satellites support research on marine litter (fig. 3). This is done by providing the teams organizing expeditions trying to identify ocean garbage patches with forecasts of sea currents and sea-surface heights helping this way to identify most probable plastic convergence areas.

Fig. 3 The marine litter – a global environmental problem to solve with the help of Earth Observation satellites [4]

The satellites cannot detect marine litter so far by direct methods. In change, one can use satellite observations data to derive key information for feeding or calibrating models that predict its accumulation (fig. 4). Relevant parameters are altimetry, sea-surface salinity, sea-surface temperature, ocean colour and sea-ice data.

This fact is an example of effectively using big data coming from space assets becoming even more relevant now in an era in which Earth Observation is boosted by the launch of the Copernicus satellites, a series of satellite missions developed in the framework of the European Earth Observation programme Copernicus [5].

Other major achievements like Digital Globe recent opening of commercial services making available 30 cm spatial resolution optical imagery [6]or Airbus Defence and Space TerraSAR-X data 25 cm spatial resolution [7] are strongly contributing to a wide range of application ranging from security to environmental.

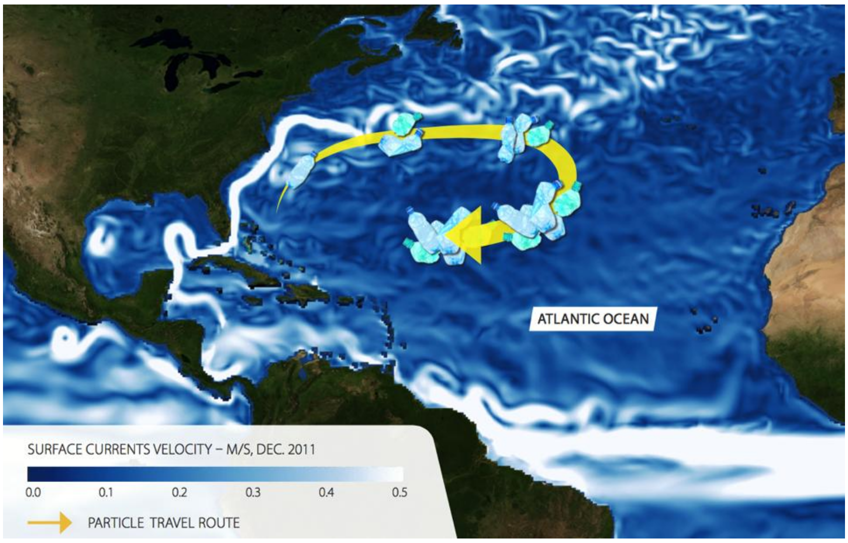

Fig. 4 Marine litter accumulation area predicted using satellite data feeding marine currents models

Together with the other types of geospatial data, the Earth observation data is considered and proved a tremendous capacity for contributing to a big number of global challenges. The Millennium Project [8] identified 15 global challenges that the human kind is facing [9].

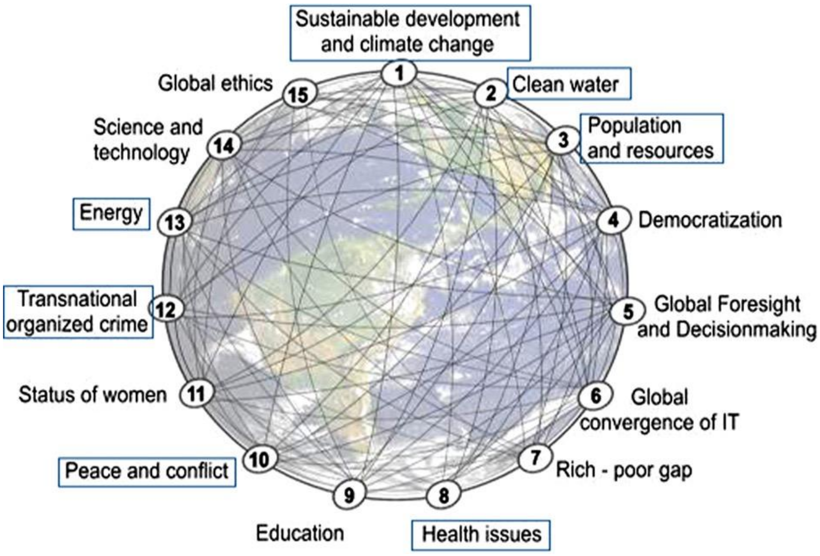

Fig. 5 Global challenges according to the Millennium Project

2. Geospatial technology

Geospatial technology refers to equipment used in visualization, measurement, and analysis of earth's features, typically involving such systems as GNSS (global navigation satellite systems), GIS (geographical information systems), and RS (remote sensing). Its use is well known and widespread in the military and in homeland security, but its influence is everywhere, even in areas with a lower public profile, such as land use, flood plain mapping and environmental protection.

RS produce imagery and data collected from space or airborne camera and sensor platforms. Some commercial satellite image providers now offer images showing details of one-meter or smaller, making these images appropriate for monitoring humanitarian needs and human rights abuses.

Fig. 3 Geospatial technology components [10]

GIS refers to a suite of software tools for mapping and analysing data, which is geo-referenced (assigned a specific location on the surface of the Earth - geospatial data). GIS can be used to detect and analyse geographic patterns in support of, for example, urban planning and water management. As part of the GIS, the Internet Mapping Technologies are changing the way geospatial data is viewed and shared. The developments in user interface are also making such technologies available to a wider audience whereas traditional GIS has been reserved for specialists and those who invest time in learning complex software programs.

A GNSS is a network of satellites, which can give precise coordinate locations to civilian and military users with proper receiving equipment. The most known GNSS is the American GPS (Global Positioning System). Similar systems are developed by Russia (GLONASS), European Union (Galileo), China (Beidou & Compass). Regional navigation satellite systems are developed by India, France and Japan.

3. Challenges and possible approaches

While the reality and status of technology development shows a clear need to face the new problems arising from the availability of new types of geospatial data in huge quantities and at higher speed, the existing technologies and infrastructure in place are not yet prepared to answer these increasing needs.

Geospatial data growing rates are considered to be at least by 20% every year. According to the estimation by United Nations Initiative on Global Geospatial Information Management (UN-GGIM), 2.5 quintillion bytes of data are being generated every day, and a large portion of the data is location-aware. Also, in Google, about 25 Petabytes of data is being generated per day, and a significant portion of the data is considered to be spatio-temporal data [2].

Along with this exponential increase of geospatial big data, the capability of high performance computing is being required for modelling and simulation of geospatially

enabled contents. However, because of limited processing power, it is hard to fully exploit high-volume or high-velocity collection of geospatial data in many applications.
Recently, distributed, parallel processing on a cluster of computers or a cloud became widely available for use, breaking the existing limitations on processing power.

Also, at organisational level, new challenges can be identified when thinking about the operation and maintenance of data archives and information systems in this context. For example, related to the European Copernicus programme mentioned in the previous paragraphs, the European Commission has launched the Collaborative Ground Segment action and tasked the European Space Agency to coordinate EU and ESA member states actions towards reducing the burden of managing the Copernicus satellites data archives which for the moment is assumed by ESA.

3.1 The technical perspective

Considering a system engineering view, current research shows a mostly unanimous trend for the systems to include at least three layers:

- geospatial big data integration & management,
- geospatial big data analytics,
- and geospatial big data service platform.

The first layer is responsible for quickly storing, retrieving, indexing, and searching geospatial big data.

The second layer is responsible for performing data analytics on the data. This layer is further decomposed into the module of interactive analytics for real-time or dynamic data and the module of batch analytics for static or archived data. In geospatial visual analytics the most important distinction is whether the reasoning is done primarily by the human analyst, with the support of interactive visual interfaces or the reasoning is primarily computational, with the interactive visual interfaces enabling control and interpretation of the computational methods.

One of the current problems in visual analytics is that, given the variety of the data and the problems to solve, the automatic data analysis community (machine learning, data analysis, statistics, etc.) is not represented sufficiently in the visual analytics research community to enable a fast progress of the field. While fully automatic techniques only work if the problem is clearly specified, they are essential for the success of the field of visual analytics. The visual analytics community has to make sure that more researchers from the above domains join the community and help to develop visual analytics systems. Even a number of successful applications of visual analytics have been developed over the last five years, the development of tightly integrated data analysis and visualization methods is still in the beginning and more research is needed to make progress in this respect [1].

Collection and analysis of data about individuals is vital for progress in many areas such as public health, transportation or security. Technologies enabling collection and analysis of various kinds of personal data have developed rapidly. A negative side of these developments is the growing threat to personal privacy [1]. This particularly applies to data containing locations of people. Analysis of such data may conflict with the individual rights to prevent disclosure of the location of one’s home, workplace, activities, or trips. Visual analytics can contribute to privacy protection in two ways. First, visual analytics researchers can identify what kinds of information can be extracted from various types of data by means of visually supported analysis and consider potential implications to personal privacy. These findings can be communicated to privacy protection researchers for developing methods to remove or decrease the detected privacy threats. Second, to allow humans to deal with large datasets, visual analytics researchers often employ techniques for data generalization and abstraction.

3.2 The organizational perspective

From the organizational perspective, the following questions can be of particular interest [4]:

- What is the best way to organize and manage the vast array of geospatial information that is acquired at many levels and that has a variety of potential uses?
- What is the best way to share data, particularly among central to local government stakeholders?
- What is the best way to coordinate among agencies?

A possible answer while trying to solve the organizational problems was given by the states all over the world by starting to develop Spatial Data and Information Infrastructures (SDI). Similar initiatives are sustained and have reached a more or less advanced status in USA, Canada, Australia as well as European Union and the Asian countries.

When developing SDIs, at least the following components are supposed to be operationally functional:

- Data themes: geodetic control, orthoimagery, elevation and bathymetry, transportation, hydrography, cadastre, and governmental units.
- Metadata: information about the data, its content, source, accuracy, method of collection, and other descriptions that help ensure the data are used appropriately.
- National Spatial Data Clearinghouse: an electronic service providing access to documented spatial data and metadata from distributed data sources. The Clearinghouse is intended to provide access to NSDI for spatial data users.
- Standards: common and repeated rules, conditions, guidelines or characteristics for data, and related processes, technology, and organization.
- Partnerships: promote cost-effective data collection, documentation, maintenance, distribution, and preservation strategies; include private sector geographic, statistical, demographic, and other business information providers and users.

In Europe, the spatial information infrastructure (INSPIRE) development is based on a number of principles[11]:

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Geographic information needed for good governance at all levels should be readily and transparently available.
- Easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.

INSPIRE is established by Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community. Since this is a EU directive, it was transposed by all EU member states into their national legislation. The directive is complemented by a series of regulations related to publishing data, metadata and associated network service required as part of the implementation process.

Also, from the organizational perspective, experience showed that when establishing coordinating bodies, such as those established inside NSDIs, one should consider a number of criteria:

- strategic and business plans;
- a GIS coordinator and staff;
- clearly defined authority and responsibility for coordination;
- a relationship with the upper and lower level chief information officers;
- a political or executive actor for coordinating GIS;
- a connection to the national spatial data infrastructure and clearinghouse programs;
- the ability to work with local governments, academia, and the private sector;
- sustainable funding, especially for producing and maintaining geospatial data;
- the authority for the GIS coordinator to enter into contracts.

3.3 Licensing and liability

When addressing the liability related to geospatial data and associated services, there is a number of complexities and uncertainties that need to be tackled related to information products and services generally, as well as by legal theory uncertainties surrounding liability. Application of geospatial technologies may require integration of different types of data from multiple sources, assimilation of attributes, adherence to accuracy and fitness-for-use, and selection of processing methods. All of these actions may be affected by possible errors. A variety of software programs may be run against a single geographic database, while a wide range of users may have very different use objectives.

The complexity of the legal questions surrounding liability for geospatial data, combined with the diversity of problems to which geospatial data and technologies may be applied and the continually changing technological environment, have created justified concerns over liability for geospatial technology development and use.

Because is unlikely to reach a uniformity in recording the licensing or use right associated to geographic data sets that would allow legally sharing of this data the liability for the quality of the data still remain an issue.

While a lot of people and organizations are adhering to free and open source types of licensing or crowdsourcing, the geospatial data made available in this way remains a valuable resource and, if documented with appropriate metadata, this can represent a web-wide resource providing legal authorization use of datasets, extract from databases, provide web mapping and web feature services and engage in data mining.

4. Conclusions

The analysis presented in this paper refers to geospatial data - data assigned to a specific location on the surface of the Earth – being equally represented in vector, raster or grid format.

Because of sustained investment in space technologies, the Earth Observation satellites are today producing huge quantities of data that are used for solving complex problems at global level. On another hand emerging technologies are creating a large variety of geo-referenced data allowing location aware analysis and complex data analytics. However, due to the lack of experience or the absence of a legal framework, the use of a significant quantity of data made available free of charge over the Internet may be affected by the missing liability for its quality.

While data storage is today no more a problem, the very high rates or the speed characterising the new data flows are creating a context in which the need to effectively retrieve and process large quantities of data become obvious, which makes the data analytics research and development activities very much needed.

Based on previous experience acquired during initial development of spatial data infrastructures at different levels, regional initiatives adopted for example at European level towards developing spatial data infrastructures seem to create strong premises for a more effective management and dissemination of this type of data. Policy adopted at regional level, transposed into the regulatory framework of the participating actors and then implemented at the appropriate levels by applying the subsidiarity principle is already showing promising results but still need further development.

The technical perspective shows a clear need for a multi-disciplinary approach in research and development activities towards developing effective geo-spatial data analytics – algorithms and workflows capable to deal with huge quantities of data arriving at very high speed.

The same time, on the organizational side, the collaboration becomes a keyword associated with the need to share various tasks related to data management as well as with processing and data access services.

5. References:

[1] http://www.openstreetmap.org/
[2] https://en.wikipedia.org/wiki/Crowdsourcing
[3] http://cert.ics.uci.edu/
[4] Gennady Andrienko, Natalia Andrienko, Daniel Keim, Alan M. MacEachren, and Stefan Wrobel, *Challenging Problems of Geospatial Visual Analytics*, Elsevier, Journal of Visual Languages & Computing, 2011, Pages 251-256.
[5] http://copernicus.eu/
[6] https://www.digitalglobe.com/30cm/
[7] http://www.geo-airbusds.com/en/5054-discover-terrasar-x-new-imaging-modes
[8] http://www.millennium-project.org/
[9] Jae-Gil Lee, Minseo Kang, *Geospatial Big Data: Challenges and Opportunities*. Elsevier, Big Data Research, 2015, Pages 74-81.
[10] http://wildfire.blog.nfpa.org/
[11] http://inspire.ec.europa.eu
[12] Peter Folger, *Geospatial Information and Geographic Information Systems (GIS): Current Issues and Future Challenges*. Congressional Research Service, 2008.
[13] Peter Folger, *Issues and Challenges for Federal Geospatial Information*, Congressional Research Service, 2012.
[14] ***. *Looking for Great Ocean Garbage Patches. European Space Agency*, Copernicus Factsheets, Issue 52, May 2015.
[15] Harlan J. Onsrud, *Liability for Spatial Sata Quality*, University of Maine, May, 2009.