Geodata – Not Just for Maps

Erik Kjems

PhD., Graduate Engineer, MBA. Associate Professor at Aalborg University, Department of Civil Engineering. Former head of the Centre for 3D GeoInformation and the VR Media Lab at Aalborg University. Works with spatial, interactive models of cities and landscapes with a particular focus on roads.

As a lot of the information that we aim to retain by means of geodata, among other things, becomes accessible as real-time information, it is necessary to adapt our systems and data descriptions to these new challenges. Our desire for ever more information increases the demands on the systems and the organisation around geodata. In terms of 'Big Data' and 'the Internet of Things', georeferences can play a decisive role, as they can contribute to keeping track of the data traffic. Is our unified national geodata basis ready for this role?

Keywords: Spatial information, big data, city models, landscape models

Introduction

When I read in the February issue of Geoforum's members' magazine about FOTdanmark (Schielder, 2014), and about whether the task of establishing a unified national data basis had now been solved, I felt an urge to write the following article. The article does not contain any great reflection on FOTdanmark; instead, it focuses on the geo-related tasks that lie ahead. The above article mentions in conclusion that "our geodata are, to use a popular phrase, undergoing rapid change, but what will this lead to". At the Centre for 3D GeoInformation at Aalborg University, we have worked for almost 15 years on the link between geodata as represented in the system and in the real world. This approach focuses on geodata as living and changeable data, considering that the world they describe is changeable. As a lot of the information that we aim to retain by means of geodata, among other things, becomes accessible as real-time information, it is necessary to adapt our systems and data descriptions to these new challenges. Our desire for ever more information increases the demands on the systems and the organisation around geodata. In terms of 'Big Data' and 'the Internet of Things', georeferences can play a decisive role, as they can contribute to keeping track of the data traffic. Is our unified national geodata basis ready for this role?

From 3D city models to... ?

For a long time, there has been a focus on the data about the 'map' surface, commonly referred to as 3D city models, as it was the cities that used 3D models for visualisation purposes. Many of the models that were produced at the time are out-of-date and require some tender loving care to become useful again, while other cities have chosen to maintain the costly city models continually. Unfortunately, we do not have a common definition for the parts that are included in city and landscape models. It would be great to have a unified nationwide 3D basic model available with a simple CityGML level2 definition of elements included. However, the work has practically ground to a halt, and not just in Denmark. On the whole, the production of city models has stagnated internationally (Morton, Horne, Dalton and Thompson, 2012). In years gone by, any city of a certain sized had to have a city model, sort of "I've got to get one of these", and, to put it bluntly, city models were produced that did not give any consideration to any form of underlying organisation with competences within the use of the 3D city model. The people who had the organisation in place and paid the bill, got a valuable well-maintained model that could be included in the urban planning and used for excellent presentations where citizens were involved and similar events. There are also many examples of usage that goes beyond pure project visualisations, where the model is used for the string of applications that have been outlined over time, including here in Denmark (Batty et al, 2001; Flemming et al, 2011). For instance, they are used for spatial noise mapping and visualisation of air quality. To many cities, the city models have been an expensive adventure, and they may even be considered bad investments, but the problem is not the model, but the lacking competence surrounding it.

Nevertheless, I believe that this is the time for us to spring into action in earnest. This may seem odd considering that in actual fact, no new application possibilities have emerged other than those that have already been identified years back. But I would like to look ahead and try to ascertain what the future scenario may look like, and what contours can be seen even now. I presume that a lot of people have heard the expressions 'Big Data' (Mashey, 1997) and 'the Internet of Things', which has already got its own acronym (IoT) (Atzori, Iera and Morabito, 2010; Sarma, Brock and Ashton, 2000; Weiser, 1991), and maybe also the concept 'Data Fusion' (Waltz, Llinas and others, 1990), which are quickly gaining ground. The expressions are not new at all, and development within the areas has been going on for a couple of decades. All three areas have the potential to become very significant for the geodata area, but let me deal with them one at a time.

The concept of *Big Data*, large amounts of data, is mentioned most often in connection with the many social websites, such as Facebook, Twitter, Instagram LinkedIn and others, but Big Data also include data related to transport, roads legislation, biological phenomena etc. Big Data are characterised by data being unstructured in connection with their origin, and often hard to organise, which makes them demanding in informatics terms ('Big Data Definition', n.d.). The reason why the concept is gaining ground right now is partly the connection to the IoT area.

IoT is characterised by things being given IP addresses and linked to the Internet. The expansion of the address area in the Internet Protocol at the change from version IPv4 to IPv6 has increased the number of possible units on the the Internet from approx. 4 x 109 to 3.4 x 1038. The very purpose was to make room for the many 'things' or units that are expected to be connected to the Internet in coming years. This is not to do with those that we have control of ourselves, such as smartphones, tablets etc., but those units that are connected to the Internet and gather information from our surroundings, right from the washing machine and the heating system in our home to advanced weather stations or traffic portals. This type of information is popularly linked with the 'Sensor City' concept, which originated at Harvard in 2007, where 1,000 air pollution sensors had been set up across the Cambridge area. The amount of data that comes from these types of units has seen exponential growth in recent years. The analysis firm Gartner claims that IoT will constitute 26 billion units by 2020, which is more than 30 times of what we had in 2009 (Middleton, Kjeldsen and Tully, 2013). Please note that the figure still only contains things on the Internet, not interactive units. IBM states that the very 'Smart City' concept with a growing number of sensors in the city will be one of the greatest players in this development (IBM, n.d.). One of those who must really know about developments in the amount of information streams is Eric Schmidt, chairman of Google's board and director at the time. At a Techonomy Conference in 2010, he said as follows (http://techonomy.com/conf/): "Every two days now we create as much information as we did from the dawn of civilization up until 2003, that's something like five exabytes of data". One can only guess at what the figure must be today. Now all we need to do is combine the vast amount of data that we collect, and that is where 'Data Fusion' enters the picture.

Data Fusion is referred to in two different contexts, where one in particular relates to a well-known phenomenon in the geodata business. Coordination of geodata or digital data in general has always been and is still a great challenge, on which huge resources are spent. The large commercial players stick rigidly to their data formats, while open formats persistently try to gain sympathy and become implemented widely, so that minor players on the market can also come in from the cold. The challenge has only grown in step with GIS and the 3D systems gradually merging, as the CAD world is beginning to produce geodata in the form of virtual structures within both construction and infrastructure. Or rather, the 3D systems are having attribute data linked to the objects included. For instance, you could click on any object in figure 1 and find exhaustive information about the relevant object. Coordination for the production of parts lists, collision control and much more is quite common in advanced CAD systems today.

The model servers that are entering the market in the fields of construction and infrastructure are systems that can receive and deliver data in open formats, and they can control access to different parts of the model as regards updates etc. In this sense, they contrast significantly with the systems that can only load data in 2D and 3D and coordinate them to a limited extent. Most often with a loss of information, due to the data conversion from one format to another, as a consequence.

For instance, regular 3D formats coming from e.g. Autodesk and Bentley, such as the open formats IFC (Industrial Foundation Classes), must be able to function together with city models in CityGML and e.g. ESRI's GIS products. However, is this realistic? The format battle has one loser in particular – the user, or rather, the consumer.

The other, and equally as interesting part of the data fusion is to do with the coordination of data information. I.e. information that is gathered primarily by sensors. A good example is the traffic systems that collect data from different types of reels in the roads, from cameras that read the number plate, determining the immediate travelling speed on the stretch, from recordings along the road, e.g. road temperature or information from the actual vehicles, which know where they are and at what speed they are moving. All of this is information that can be used to inform road-users about any problems along the road. The great challenge here is the coordination of data, i.e. fusing data in such a way that it makes sense across the systems. This is an enormous challenge as the individual sub-elements are made by different manufacturers and therefore often communicate with different data protocols or 'idioms', if you like. Investments in roads are most often long-term, which can also be an issue, as an approved standard may be changed several times before the hardware in the road is replaced. Developments in data fusion therefore need to serve as a connecting link between the many technological islands that have emerged in the last couple of decades. But if we succeed, an ambulance responding to an emergency call-out will be able to warn travellers on the roads, adjust the signal system to a green wave, make sure a bridge is not raised at the wrong time, and generally, it can be directed around any obstacles on the road without any people needing to be involved, just to expand the example a bit. It goes without saying that if we are going to have 30 billion units on the Internet, we will need com-

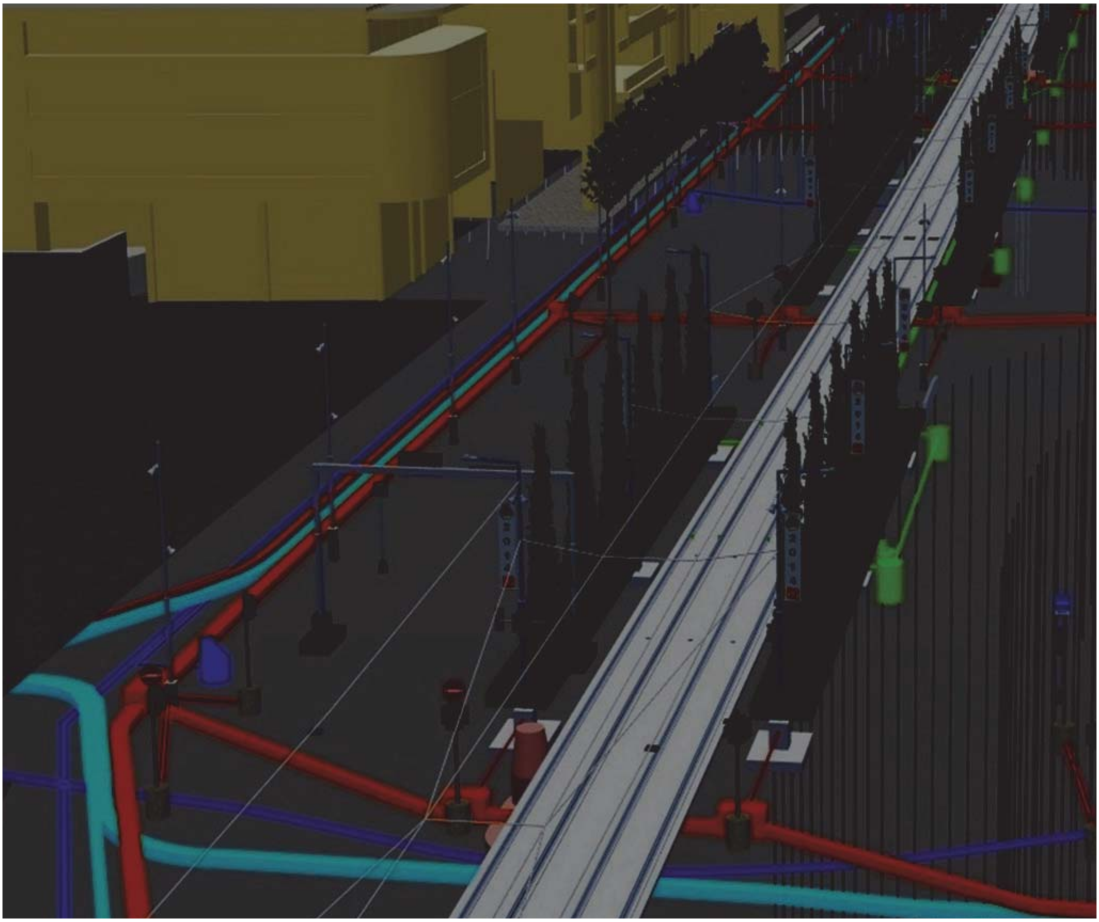

Figure 1. CAD model of the infrastructure facility in the underground by Bjørvika in Oslo, which contains information about included elements. (Vianova, Oslo)

pletely different systems to those that we know today in order to handle this. This will require the systems to be autonomous, not only as regards the registration of information, but also very much in terms of making decisions for us. If flooding of the sewer system is registered, the system must be able to redirect wastewater automatically in order to prevent a major disaster.

Geodata in tomorrow's real-time systems

It should be obvious to most people that it will be incredibly difficult to keep track of the many units in an abstract system consisting of countless tables, ID numbers etc. Especially when you imagine the 'Smart City' concept implemented in its most complex structure. The systems, as an isolated case, are relatively small and therefore still fairly manageable, and they are gathered within more or less standardised frameworks that facilitate a reasonable quality of communication between the units. However, the systems are beginning to grow significantly, both in geographical extension and in complexity. Furthermore, focus is now being directed at coordinating data. For instance, the trains' real-time system is to be linked to the buses' and the rest of the transport system. However, the coordination is currently hampered by lacking conformity between the systems.

This coordination presents many challenges, but all sensors and other information units have one common denominator, which is only being used to a very limited degree: the geographical reference. All sensors are placed somewhere in the real world and probably not randomly. It would seem that this information would be of great value because naturally, it must be quite unique. Unfortunately, a spatial coordinate will not be immediately understandable, even to an expert. This is where our traditional approach to geodata enters the picture, literally, as there is no reason why the interactive systems should not take their starting point in a basic map and 3D models in order to create an overview and show the information that is created out in real life. The interaction interface will immediately make sense to most people and will form an excellent basis for control and coordination of data. And the traditions we have for linking attribute data to geodata will also be of great value. It will also be possible in some contexts to use current GIS analysis models to produce useful information. In (Kjems, Kolar and Batty, 2005), we called them model maps.

I believe that it would be an obvious choice to expand our unified FOT data to include more than what we use them for today. I would even go so far as to say that unless we get a common FOT standard for geodata that describes the spatial structure, it will be very costly in the long term to keep track of the necessary coordination of geodata, although it is not possible to put an exact figure on this. The development of the digital map series, which started in earnest in the 80s, benefited from a driving force, i.e. the gas suppliers who were installing gas pipes across Denmark. The digital map series has been supplemented, renewed and refined many times since then, and it is now available in a common standard – 30 years on. 3D data are much more complicated and available in numerous mathematical descriptions and data representations. If we let things slide, many different systems and pseudo object standards will emerge, which will be expensive for the public authorities to coordinate, especially in the long term.

Large amounts of 3D data of cities and landscapes are being produced, and improved methods for increasing quality, including the geographical precision of the generated data, are continually being developed. But only very little is happening with the data representation. How are data to be included in our systems? And I am not thinking of the visualisation part here. There are plenty of options for handling that in an appropriate way. No, I am thinking of the geo object as an interactive carrier of information. Within cities and landscapes, CityGML is currently the best example of a description that on the one hand includes the possibility of carrying over semantic information, and on the other hand uses a scalable object structure that can handle geography in an elegant way. However, CityGML does not immediately solve the problem of handling objects as information carriers in the systems, but it will be a step in the right direction.

So, maybe we should start by getting the entire country presented in a CityGML version and then see how data for this can be refined over time. GML has a hierarchical structure, and in principle, there are no limits to the details that you can choose to include. At the moment, the limitations are in the 'City' part and the definitions that have been approved here. Today, there is an interface between the IFC definition and CityGML, and more of this type will emerge, but all things being equal, FOT data should be a part of this and form an expanded core data set containing a spatial description that can be used for interactive systems as well as for visualisation and many other things.

The alternative

Just before concluding, I would like to mention an alternative that would, however, require another article if I were to describe it in detail. At the Centre for 3D GeoInformation at Aalborg University, we participated in a research project called InfraWorld during the period 2008-2012, which was funded by Norway's Research Council, and which took its starting point in the data model that has been under development at the Centre since 2006. The end product was a user interface consisting of a spatial city model, which could be linked to real-time information with data from the real world. The pivotal point was the system's object description, which provides information about all of the issues that have been mentioned above, and then some – no less. Imagine receiving a pile of virtual boxes with unknown content. You upload the boxes to the system, e.g. a web browser, whereby all the boxes are activated and start unpacking themselves and presenting themselves as e.g. a house. All parts that come out of the boxes are geographically referenced and placed in the correct places. Once the house is in place, it starts to communicate with the real house that it represents and fetches e.g. current consumption data. There are almost endless application options, and a few of these were tested during the project.

Conclusion

In geodata contexts, we often think quite traditionally, using GIS terminology. However, geodata is so much more, and we should not be afraid of moving on and mixing with others from the world of construction, building, operation and maintenance etc. Geodata are used everywhere, and we should provide quality and cohesion in these data. It is disgraceful that so often you come across projects that are handled in local coordinate systems and then they do not really get any further. Geography should play a much more significant role in order for us to avoid redundant data and to create better cohesion between subject domains. As described above, large amounts of data that can be georeferenced are under way, and the industry should be ready to handle these data and utilise their existence. The first step is therefore to really get a foot inside the 3D world with a unified description of that which is above ground.

REFERENCES

Atzori, L., Iera, A. and Morabito, G. (2010). *The Internet of things: A survey*. Computer Networks, 54(15), 2787–2805.

Batty, M., Chapman, D., Evans, S., Haklay, M., Kueppers, S., Shiode, N., Klosterman, R. E. (2001). *Visualizing the City: Communicating Urban Design to Planners and Decision Makers*. Redlands: ESRI Press.

Big Data Definition. (n.d.). *MIKE2.0, the open source methodology for Information Development*. Retrieved 14 March, 2014, from http://mike2.openmethodology.org/wiki/Big_Data_Definition

Flemming, L., Schack Madsen, P., Sørensen, M., Lindeneg Johansen, R., Nielsen, T., Bodum, L., Hjorth, F. (2011). *Geoforums vejledning i 3D-bymodeller (Geoforum's guidelines for 3D urban models)* (p 105).

IBM. (n.d.). *IBM Intelligent Operations Center*. Retrieved from http://www-03.ibm.com/software/products/en/intelligent-operations-center/

Kjems, E., Kolar, J. and Batty, S. E. (2005). *From Mapping to Virtual Geography*. In Proceedings of CUPUM 2005. London.

Mashey, J. R. (1997). *Big Data and the Next Wave of InfraStress*. In Computer Science Division Seminar, University of California, Berkeley.

Middleton, P., Kjeldsen, P. and Tully, J. (2013). *Forecast: The Internet of Things*, Worldwide, 2013.

Morton, P. J., Horne, M., Dalton, R. and Thompson, E. M. (2012). *Virtual City Models: Avoidance of Obsolescence. Education and Research in Computer Aided Architectural Design in Europe-eCAADe*, Prague, Czech Republic.

Sarma, S., Brock, D. L. and Ashton, K. (2000). *The networked physical world*. Auto-ID Center White Paper MIT-AUTO-ID-WH-001.

Schielder, R. (2014). *Hvad nu FOTdanmark? (What Now FOT Denmark?)* Geoforum Retrieved 28 February, 2014 from http://www.geoforum.dk/GEOFORUM-151-10612.aspx

Tufte, E. R. (1990). *Escaping Flatland*. Envisioning Information, pp 12-36.

Waltz, E., Llinas, J. and others. (1990). *Multisensor data fusion* (Vol. 685). Artech house Boston.

Weiser, M. (1991). *The computer for the 21st century*. Scientific American, 265(3), pp 94-104.