Harmonized access to heterogeneous content: towards a European SDI

Kap, AP; van Loenen, B; de Vries, ME

Publication date
2004

Document Version
Final published version

Published in
AGILE 2004 7th conference on geographic information science

Harmonized Access to Heterogeneous Content: Towards a European SDI

Aegidius Kap[1], Bastiaan van Loenen[1] and Marian de Vries[2]
Sections [1] Geo Information and Land Development and [2] GIS Technology
OTB Research Institute for Housing, Urban and Mobility Studies,
Delft University of Technology
Delft, The Netherlands
a.kap@otb.tudelft.nl

SUMMARY

The development of a SDI, in Europe and elsewhere, has many aspects. Some are more technological in nature, and others have to do with organizational, financial and juridical issues. In this paper we will focus on an issue that is identified in the INSPIRE report of September 2003 as having a high priority: harmonization. We will look at two aspects: data (model) integration and data access policy.
As far as data integration is concerned we propose a 'light' approach: for the shared use of geo-information in a SDI strict harmonization of data models is not always necessary.
In the field of data access policy two EU Directives came into effect, one on the reuse of public sector information, and one on data protection. Both make a first step towards harmonization of data access policy, but further steps will be necessary.

KEYWORDS: *SDI, data access policy, data model integration, harmonization*

INTRODUCTION

Many steps have already been taken towards the creation of a European Spatial Data Infrastructure (ESDI). Some of these steps are in the field of information technology (setting up online metadata registries or creating interoperable Web services in pilot projects). Other issues have to do with data access policy, with pricing and with privacy legislation. In all these fields there is the question how to 'merge' the different national and regional existing infrastructures into an integrated Europe-wide Spatial Data Infrastructure.

In the INSPIRE report of September 2003 'harmonization' is mentioned as one of the issues having the highest priority. The term 'harmonization' can easily lead to different interpretations and consequently to misunderstanding of view points. In his report about the EuroSpec project Luzet (2003) therefore proposes to talk about 'interoperability' in stead.

In this paper we will try to use the term 'harmonization' in a 'implementation-neutral' way, because there are many levels and forms of harmonization. As also pointed out in the INSPIRE report (2003), harmonization of e.g. data models does not necessarily mean that the data structure of existing data sets has to be altered. What has to be done is to create common (and in that sense 'harmonized') views on existing data, in as far as this is necessary for combining information from different data sources.

The goal of regional or cross-border SDI initiatives is clear: to improve access to geo-information regardless of its 'native' format or national jurisdiction on pricing or data access.

A number of European 'cross-border' pilot projects have already been carried out. There have been reports on success, in setting up the technical infrastructure, or in creating common metadata registries (e.g. GiMoDig 2003; Riecken 2003). But the pilot projects have also shown what the pitfalls and difficulties

are in 'seamlessly' combining geo-information from data sources from the different participating countries or regions. Differences in coordinate system can in principle be solved by choosing a common spatial reference system, e.g. the European ETRS89. Other issues are not so easy to solve, e.g. differences in Level of Detail (scale) and in precision and data quality (e.g. Luzet 2003). Riecken (2003) points at the necessity of a multi-lingual user interface and metadata in the respective languages. Apart from these aspects there is the challenge of differences in data structure and conceptual (information) model.

In section 2 we will look at the data integration aspect of SDI's. We will point at some of the options around real-time integration of information from heterogeneous data sources. Is harmonization of data models necessary and if so, what is the right 'time and place' for it?

In section 3 and 4 the focus is on data access policy and privacy regulation. As pointed out by Craglia (2003) data policy is an area of research that has not been given that much attention so far. In the creation of a European SDI the co-existence of many diverse regulations can prove to be a major bottleneck however. When the more technological issues are solved, but the more political ones are neglected, there will not be a European SDI.

In section 5 we will look at the EULIS project: an example of European cooperation by seven national land registries (cadastres) with the aim of creating a European Land Information Service that will be accessible via the Web.

HARMONIZATION AND DATA MODELS

Geo-data (like other data) is stored according to a certain conceptual view of that part of (geographic) reality that is considered relevant to the business processes within an organization. The purpose of collecting or creating (digitizing) that specific set of geo-data will influence modeling decisions. As a consequence also the actual data models (the database schemes) will differ from one application to another: names of tables and attributes, granularity (many object types with few attributes, or few object types with many attributes), domain values, etc. (e.g. Hart 2003).

Whether it is for setting up a national or supra-national SDI or for short-term projects where different agencies cooperate and use each others spatial data, the question of how to combine data from different database schemas (based on different conceptual models) is raised at one time or another.
There are basically three phases in the process of combining and 'merging' spatial data from different data providers, and in all these phases it is necessary to have knowledge of the data structure of the source, be it that the risk of misuse (because of misunderstanding) increases from phase 1 to phase 3: discovery, retrieval and query and analysis.
Harmonized access to geo-information in a Spatial Data Infrastructure is therefore important for a number of reasons:

a. To make 'integrated' querying and data analysis over these separate data sets possible (selection and analysis);
b. For consistent visualization it is necessary to have the same cartographic representation (colors, line width, symbology) for 'things' (objects) on the map that are conceptually similar, although the terminology used in the source data sets is different (cartographic);
c. To provide unambiguous metadata for catalogue services meant for discovery of the data sets (discovery).

One could argue that harmonization is not always necessary. In the case of digital libraries (or clearinghouses) of geo-information for example, it could be left to the user to decide whether or not the data to be downloaded can be combined with other (spatial and non-spatial) data used in the project.

In the case of Internet applications where data from different sources has to be combined real-time, the need for data model harmonization is much greater.

This is even more so for mission-critical applications, where information needs to be up-to-date and

unambiguous. Both the cartographic quality and the quality of query results (the query results should be correct and complete) is then essential.

When time is also a critical factor, as in disaster response applications, combination of data from different sources should be fast and without errors. Any time-loss between retrieving the data and producing an integrated digital map has to be avoided.

On the other hand there is also a variety of possible (technical) implementations. The 'time and place' of harmonization can vary. There is a broad spectrum here, from adaptation of the actual data storage itself, to on-the-fly harmonization carried out by mediator software at the business layer.

Table 1 gives an overview of possible solutions, depending on the type of application.

Type of use	**Quality of result**	**Creation of harmonized 'model'**	**Use of harmonized model**
Exploratory, ad hoc	Best guess, fuzzy	-	-
Projects or structural cooperation	Medium or high	On-the-fly	By application software, during retrieval (Styled Layer Descriptor, XQuery filters, semantic mediators, ontology matching, …)
Time-critical applications	High	Pre-defined	At data source (views on tables, automated replication)

Table 1: Data model harmonization: when, where and how

In other words: shared use of geo-information (between organizations or in a national or cross-border SDI) can take a number of forms, and all of these pose their own requirements and call for other solutions.

Basic principle should be that the internal data structure and content of existing geo-data will be kept intact as much as possible. Only for time- and mission-critical applications a more strict harmonization approach is the best option: e.g. in the case of disaster response systems (in border regions, but also in one country between e.g. police, fire department, and hospitals). In these types of situations the availability of semantically unambiguous (geo-)information is essential. And, secondly, fast replies to selection queries are very important in these circumstances, which leaves less 'time' for semantic mediators or intelligent agents to harmonize 'on-the-fly'.

HARMONIZATION OF DATA ACCESS POLICIES

Also in the field of data policy, harmonization does not necessarily mean that countries or organizations loose their autonomy. The EC Directive on the 'reuse of public sector information' (Directive 2003/98/EC), for example, can be seen as a set of minimum requirements for data providers in the public sector. The directive obliges all public sector bodies to publish the conditions for reuse. This obligation will allow European operators to obtain information on the data sets available for reuse in the various countries and the conditions for reuse. In this way the European operators can assess the feasibility of a cross-border product or service based on public sector information. The directive also requires clear procedures with fixed deadlines, and promotes the use of online licenses. This should result in reasonable waiting times and a more favourable environment for reuse (see Volman 2004).

The directive further enhances fair competition by a non-discrimination principle stating that any applicable conditions for commercial reuse shall be non-discriminatory. The directive prohibits exclusive agreements based on public sector information. The directive intends to create a level playing field in which access to public sector information is provided on equal terms for those requiring access, including public sector bodies. If a public sector body uses information as input for its commercial activities, and these fall outside the scope of its public tasks, then it must be subjected to the same charges and conditions that apply to other users (see Volman 2004).

In addition, the directive limits the income generated on the basis of public sector information. The directive recognizes that several public sector bodies have to finance part of their activities by selling information. The full recovery of the costs for producing, reproducing, and disseminating these documents are therefore allowed, together with a reasonable return on investment. The upper limit for charging in no way prevents a Member State from applying lower charges, or no charges at all. The extension to the directive suggests that Member States should encourage public sector bodies to make information available at marginal costs. Moreover, the extension invites Member States to exercise their intellectual property rights in a way that facilitates reuse (see Volman 2004).

The directive could have gone much further, for example, by accepting an amendment to promote the use of open formats/ standards. The pricing of public sector information is still subject to discussion and a point of critique has been that the directive only rules about the reuse of public sector information and not on the access to public sector information. However, for this moment the directive is already a step forward to harmonization of data policies within the European Union. For some countries it may not be enough, but for others it is clearly a step forward. Some countries are advancing policies that promote the use of open standards and open source software (Tweede Kamer 2003), or promote a pricing regime of the cost of dissemination for public sector information, while others are negotiating access to source codes in order to guarantee sustainable access throughout time (Denmark XML-schema's of Microsoft Office 2003-application-suite, AutomatiseringGids 2003).

At least, the directive sets a framework that is acceptable for all the Member States and is a sufficient basis for further harmonization of data policies throughout the EU.

The Directive was published in the Official Journal of the European Union on 31 December 2003. Member States will have until 1 July 2005 to incorporate the directive into national law. The directive will be reviewed within three years of its entry into force. This review may be an opportunity for a new step necessary for the further enhancement of the European Spatial Data Infrastructure

HARMONIZATION OF PRIVACY LEGISLATION

The EU members are supposed to have implemented the EU data protection directive (EU 1995), which entered into effect on 25 October 1998. This directive establishes a regulatory framework to ensure both a high level of protection for the privacy of individuals in all Member States and the free movement of personal data within the European Union. The directive minimizes differences between Member States' data protection rules, setting a relatively high level of privacy protection within the EU.

Due to the minimization of the differences between Member States, the Directive did not result in full harmonization of privacy law in Member States; it only provides a certain bandwidth within which Member States may operate (Tweede Kamer 1998, 5). Therefore, different interpretations of the explanation of, for example, the term personal data may exist among Member States, and different rules may apply to the use of personal data. In his study for the EC on the implementation of data protection directive, Korff (2003, 12) identified "mostly only minor variations in the definitions in the laws of the Member States of the terms "personal data", "processing", "filing system", "processor", "third party", "recipient" and "consent". However, he recommended "As far as matters of ambiguity are concerned, it must be made clear whether (or when) not-fully (or not-immediately) identifiable data - such as encoded or pseudonymous data, should always be regarded as "relating to an identifiable person", or whether this

should only be the case if the person processing the data can link the data to such a person (typically, by means of a decoding “key” or number)” (Korff 2003, 13). It should similarly be clarified when the use of geo-demographical or statistical data and so forth is such as to turn these data into “personal data” (Korff 2003, 13). Van Loenen (2002), for example, found that most countries participating in the EULIS project, implemented the privacy directive in national legislation, but give different interpretation to the application of the legislation to geographic data like an address, parcel numbers and similar not-immediately identifying data. This has resulted in different policies in the Member States of the European Union with regard to access to and possible use of these data (see also Laarakker and Gustafsson, 2004).

Although data heterogeneity may in many cases be resolved with technological solutions, the issue of heterogeneity within data protection policies cannot. Therefore, the interpretation of the terminology in the current directive needs further harmonization with respect to the applicability of this directive to geographic information.

EXAMPLE: European Land Information Service (EULIS)

The European Land Information Service (EULIS) project is developing a demonstrator of a future European Land Information Service. The overall objective is to provide access to land information of seven European countries via the Internet, thereby creating better conditions for professional actors in the real property market and eventually for private citizens. The project will illustrate the positive effects of having land information available across borders and can be seen as a measure to improve the single market for financial services. Also the possibilities for private sector companies to exploit public sector information may be increased (http://www.eulis.org).

The project does not aim to harmonize the land administration systems of the participating registries. The project explores obstacles towards a unified way of delivering relevant information to fixed categories of users, and proposes ways to overcome these barriers, including changes to legislation (E-Content, 2001, 9). The project with its demonstrator will link the individual land registries and cadastres through a single portal. The provision of the land administration data will be accompanied with sufficient metadata. Metadata may include the terms of use, the price for the information, but also semantic information like the meaning of the concept of real property, or a building.

Also the EULIS project can be seen as a step-by-step approach. Its first aim is to better satisfy the needs of the professional users; in a second version individual citizens may be involved. Further by only exploring the potential (legal) barriers for cross boundary access, participants are not forced to change their current technical systems or the policies that they have. They only have to add meta-information and link their system to the portal. In this way, EULIS improves access to land information of seven European countries with positive support from all participants.

CONCLUSION

The creation of a European SDI is a step-by-step process, where 'bottom-up' and 'top-down' initiatives reinforce each other.

As can be seen in many of the cross-border projects mentioned in this paper an important step is to make the existing data 'visible', in pilots, web portals or 'just' metadata registries in the form of clearinghouses. Showing the richness of existing geo-information that is available with local, regional and national agencies and organizations is the best way towards the next step: creating the possibility for seamless and high-quality combination of these heterogeneous data sources (see also Riecken 2003; Luzet 2003).

Data integration by information model harmonization can take different forms. Basic principle should be that the internal data structure and content of existing geo-data will be kept intact as much as possible. Only for time- and mission-critical applications a more strict harmonization approach is the best option.

In the field of data access policies, privacy regulation and pricing, it is harder to combine 'bottom-up' and 'top-down' initiatives, because this involves directives or other kinds of legislation. But also in this field a step-by-step approach should be taken.

The EU Directive sets a framework that is acceptable for all the Member States and is a sufficient basis for further harmonization of data policies throughout the EU. A further step would be the harmonization of *access* to public sector information, where the directive at present only rules about the *reuse* of public sector information.

Secondly, the interpretation of the terminology in the current directive on data protection needs further harmonization with respect to the applicability of this directive to spatial information. This would be another major step forwards.

ACKNOWLEDGEMENTS

This publication is the result of the research programme Sustainable Urban Areas (SUA) carried out by Delft University of Technology.

REFERENCES

AutomatiseringGids (2003): Office XML-schema's gratis beschikbaar, AutomatiseringGids 21-11-2003, 37e jaargang nr. 47, p. 1

Craglia, M. (2003) and A. Annoni: Where to find GI in European Policy, in: Proceedings of the 6th AGILE conference, April, Lyon, France.

E-Content (2001): EULIS European Land Information Service, 11004Y1C1DMAL1, Draft Annex 1 – "Description of Work", October.

GiMoDig (2003): http://gimodig.fgi.fi/news.php

Hart, G. (2003) and J. Greenwood: A component based approach to geo-ontologies and geodata modelling to enable data sharing, in: Proceedings of the 6th AGILE conference, April, Lyon, France.

INSPIRE Framework definition support (FDS) working group (2003): Contribution to the extended impact assessment of INSPIRE, September 2003, http://inspire.jrc.it/reports/fds_report_sept2003.pdf

Korff, D. (2003): EC Study On Implementation of Data Protection Directive Comparative Summary of National Laws. Study Contract ETD/2001/B5-3001/A/49) Cambridge (UK) September.

Laarakker, Peter, and Stefan Gustafsson (2004): European Land Information Service; Ambitions, bottlenecks and policy solutions. In: B. van Loenen and B.C. Kok, Spatial data infrastructure and policy development. Delft, the Netherlands: DUP Science.

Loenen, B van, (2002): Analysis report on the publication of land administration data in an EULIS setting. Project acronym: EDC-11004 EULIS/27575 11-11-2002, 26 p.

Luzet (2003): EuroSpec – Providing the foundations to maximize the use of GI, in: 9th EC GI & GIS Workshop, ESDI Serving the User. Coruna, Spain, June.

Riecken, J., Bernard, L., Portele, C. and Remke, A. (2003): North-Rhine Westphalia: Building a Regional SDI in a Cross-Border Environment, in: 9th EC-GI & GIS Workshop ESDI

Tweede Kamer (1998): Wet bescherming persoonsgegevens, Memorie van Toelichting, Kamerstuk 1997-1998, 25892, nr. 3, p.5.

Tweede Kamer (2003): Programma Open Standaarden en Open Source Software voor de overheid, bzk0300075, Minister of the Interior and the state secretary of Industry and Economic Development.

Volman, Y. (2004): The reuse of public sector information in the EU: increasing the scope for cross-border exploitation. In: B. van Loenen and B.C. Kok, Spatial data infrastructure and policy development. Delft, the Netherlands: DUP Science.

Legislation

Directive 2003/98/EC of the European Parliament and the Council of 17 November 2003 on the reuse