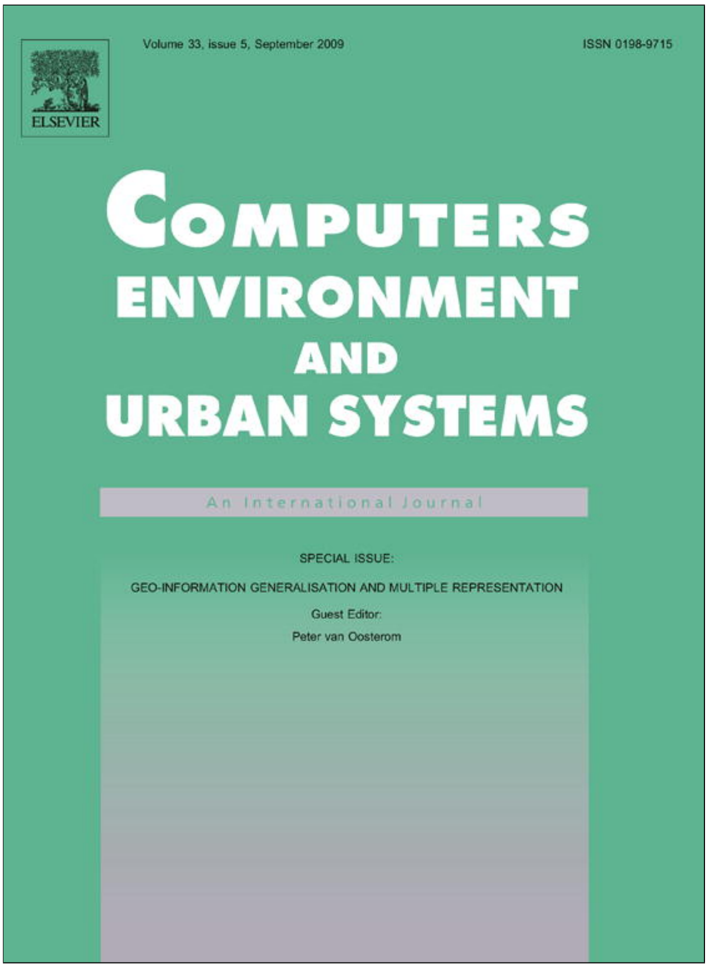
Volume 33, issue 5, September 2009
ISSN 0198-9715
ELSEVIER
COMPUTERS
ENVIRONMENT
AND
URBAN SYSTEMS
An International Journal
SPECIAL ISSUE:
GEO-INFORMATION GENERALISATION AND MULTIPLE REPRESENTATION
Guest Editor:
Peter van Oosterom

Computers, Environment and Urban Systems 33 (2009) 303–310

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Research and development in geo-information generalisation and multiple representation

Peter van Oosterom *

Delft University of Technology Section GIS-Technology, Jaffalaan 9, P.O. Box 5030, 2628 Delft, The Netherlands

ARTICLE INFO

Keywords:
Geo-information
Generalisation
Multiple representation

ABSTRACT

This paper analyses the difficulty in fulfilling the user requirements related to geo-information generalisation. Despite the fact that this is a long-standing research topic, the results are not satisfactory and therefore there is a very active research community trying to better meet the expectations of the users, both at the side of the geo-information producers and at the side of the geo-information users.

It is argued that part of the difficulties are due to the fact that the generalization problem is not specified formally enough. Therefore, currently the most important benchmark for the generalization software is the work of human cartographers doing manual generalization, supported by automated tools, and includes subjective aspects such as taste, resulting into artistic solutions. So, a very important, intermediate, research goal is formalizing the generalization problem.

In addition, the expectations of the users are growing over the past years and will continue to do so in the future: faster updates propagated between different scales, ever growing size of geo-information, support for vario-scale (instead of just multiple fixed scales), integration of formal semantics and computational geometry techniques, support for 3D representations, and so on. This paper identifies the current state of the art and provides descriptions of further research and development directions in generalisation.

1. Introduction

This special issue of Computers, Environment and Urban Systems (CEUS) is devoted to the long-standing, but still very active, research field of generalisation. Since the early days of using computers to handle geo-information, it has been the desire of both the producers and users of geo-information to fully automate the process of map generalisation so that the geo-information can be reused for a range of details or scales. Presentations have been made and publications distributed on this topic at every geo-information congress, such as International Cartographic Association (ICA), International Society for Photogrammetry and Remote Sensing (ISPRS), and International Federation of Surveyors (FIG), as well as at every geo-information conference, such as Auto-Carto, Spatial Data Handling (SDH), Geo-Information Science (GI-Science), Conference On Spatial Information Theory (COSIT), and Urban Data Management Symposium (UDMS). Many journal publications can be found on generalisation and multi-representation, for example in the past volume of the journal Computers, Environment and Urban Systems there were at least four papers directly related (Chaudhry & Mackaness, 2008; Kiester & Sahr, 2008; Nöllenburg, Merrick, Wolff, & Benkert, 2008; Tomko, Winter, & Claramunt, 2008) and many more papers indirectly related, that is applying generalization and multi-representation techniques in a specific application domain. Quite a few textbooks have been published, and many PhD theses have been devoted to generalisation (see lists of these two categories in the References section). In addition, many workshop series and seminars are specifically devoted to this topic, and solutions for dedicated problems are becoming available in commercial software as well.

The results of all of these on-going research and development activities indicate a paradigm shift towards native multi/vario-scale support by re-engineering geographic data and providing tools for data providers and end users to apply these data at any desired level of detail. This requires that spatial objects be managed over a range of resolutions (see Fig. 1), allowing for seamless transition when zooming through data. No matter what happens at each resolution (for instance, buildings may be mutated into a settlement limit), the user will be able to reference information across those resolutions in a consistent way. The goal of multi/vario-scale support is to deliver the right amount of information at the chosen level of detail without any contradiction between representations at different scales. This will enable decision-makers to quickly make correct decisions based on reliable, easy-to-understand location information.

Finding automated solutions for generalising and deriving consistent multiple representations requires a multi-disciplinary

* Tel.: +31 15 2786950; fax: +31 15 2782745.
E-mail address: oosterom@geo.tudelft.nl.

0198-9715/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compenvurbsys.2009.07.001

Fig. 1. A range of map scales of the same area (© Kadaster).

approach, because those based on a single discipline have been inadequate. Contributions are needed from digital cartography, knowledge engineering, computational geometry, computer graphics, and cognitive science. This multi-disciplinary approach was the starting point for a series of seminars organised at the Schloss Dagstuhl–Leibniz Center for Informatics in Wadern, Germany (see http://www.dagstuhl.de), including the following:

- Seminar 09161 Generalization of Spatial Information, 13–17th April 2009.
- Seminar 06101 Spatial Data: Mining, Processing and Communicating, 5–10th March 2006.
- Seminar 03401 Computational Cartography and Spatial Modelling, 28th September–3rd October 2003.
- Seminar 01191 Computational Cartography and Spatial Modelling, 6–11th May 2001.
- Seminar 99381 Computational Cartography, 19–24th September 1999.
- Seminar 9645 Computational Cartography, 4–8th November 1996.

This multi-disciplinary approach was also applied in a series of workshops conducted under the umbrella of the ICA Commission on Generalisation and Multiple Representation (see http://ica.ign.fr), including the following:

- Montpellier, France, 20–21st June 2008.
- Moscow, Russia, 2–3rd August 2007.
- Portland, Oregon, United States, 25th June 2006.
- La Coruña, Spain, 7–8th July 2005.
- Leicester, United Kingdom, 20–21st August 2004.
- Paris, France, 28–30th April 2003.
- Ottawa, Canada, 7–8th July 2002.
- Beijing, China, 2–4th August 2001.
- Barcelona, Spain, 21–23rd September 2000.
- Ottawa, Canada, 12–14th August 1999.
- Gävle, Sweden, 19–21st June 1997.

This special issue contains a selection of contributions that were presented in their preliminary form at the last ICA workshop in Montpellier, France, 20–21st June 2008. Selected authors were invited for possible publication in the CEUS special issue and asked to submit a more-developed follow-up article that included changes based on discussions at the workshop and/or new developments related to the topic of their paper. The papers were all reviewed by at least three independent reviewers. After the first round of reviews, the conclusion was that all the papers to be included needed at least some major revision. The revised papers were again reviewed by three independent reviewers, and this resulted in a final minor revision for each of the seven accepted papers, which are included in this special issue. It is expected that readers will appreciate this strict quality control and enjoy the variety of generalisation research topics covered in this collection.

First, Section 2 is a re-examination of why consistent representations at multiple scales, or even vario-scales, are important from the user's perspective and presents a number of application areas (i.e., use cases) that appear to require this approach. Next, Section 3 describes a growing problem that occurs as the most detailed representations (that is, the datasets that are largest in scale) increase in size, while users are simultaneously expecting quicker, more up-to-date data delivery. This is a complex puzzle that challenges even the major Information and Communication Technology (ICT) players, whose attempts to provide solutions include products such as Microsoft Live Maps (www.maps.live.com) and Google Maps (www.maps.google.com). However, as indicated in Section 4, these solutions have serious drawbacks, and more research and development are still needed. Since a fully automated solution for generalisation is not yet available, an interim solution might be to maintain a set of multiple representations. However, this will bring its own challenges, because semi-redundant information needs to be consistent. When performing updates, care has to be taken that this consistency is maintained by propagating relevant parts of changes from one scale to another (and again, if possible, with a minimum amount of human interaction). Section 5 elaborates on the various aspects of multiple representation databases. Another challenge is to provide solutions for continuous or gradual generalisation (described in Section 6)—that is, by not sticking to the well-known fixed map scales. Applications of the techniques might include support for smooth zooming and progressive transfer in a client–server setting. Perhaps the complexity of generalisation can be best explained by comparing this challenge to machine translation of one natural language (e.g., Dutch) to another natural language (e.g., English). In the early days of information technology, it was expected that with digital dictionaries and a few grammar rules, this problem would be relatively easy to solve. History has proven otherwise, and we are still struggling to automate the translation of natural-language text. The key problem is that the machines do not 'understand' the content of the text and therefore cannot always make the right decision. The same is true for generalisation: to make the right decisions, computers must better understand the world behind the representations. We have to make the knowledge explicit and create links between formal semantics and computational geometry approaches (see Section 7). Finally, Section 8 of this paper provides an overview of the papers included in this special issue.

2. Multiple or vario-scale use cases

Everything happens somewhere, and the ever-increasing information provided in maps helps us understand more about the nature of location. We can plan better by knowing more about where events happen, manage risk better by evaluating the impacts on

Fig. 2. Mobile devices requiring vario-scale support.

people and assets at a location, and use our resources better by knowing more about the environment we are living in. While maps have historically been produced independently at different scales, the tasks carried out by today's users require support at multiple levels of detail (scales) or even vario-scale support. Some of the issues maps should address today are the following:

- *spatial/physical planning*—national, provincial, and municipal plans must all fit together;
- *crisis management*—incidents have defined procedures at different scale levels that must be coordinated among a variety of individuals, from workers in the field to the mayor or the minister. Also, different incidents have different information needs, ranging from a local incident requiring large-scale information (e.g., a fireworks factory) to global incidents requiring small-scale information as well (e.g., Mexican flu);
- *land cover classification*—detailed provincial and national data must be harmonised to European Union (EU) levels;
- *massive community-generated content*—together, citizens voluntarily collect massive quantities of spatial data, which need to be aggregated, ranging from counting species of flora and fauna to following tracks and traces by mobile devices;
- *mobile map use*—more and more geo-information is used in mobile applications (see Fig. 2), and depending on the specific query, users may want either an overview map or a detailed map, and they may want to smoothly switch between the two. Smooth zooming (continuously changing scale in an animation style of visualisation) is also urgently needed, because the limited size of the screens can cause users to become easily disoriented when jumping between scales.

It is obvious from these examples that the different levels of detail must be consistent, otherwise the losses to society could be huge, ranging from expensive legal conflicts to serious casualties, and perhaps even fatalities in the case of crisis management. National Mapping and Cadastral Agencies (NMCAs) are endeavouring to deal with their legacy databases, trying to maintain consistency across the different scales. But it is far from obvious how to achieve this under conditions of ever-growing datasets (see Section 3) and increasing frequency of updates. Furthermore, the problem is even less well resolved for massive community-generated content or application data within Web 2.0.

3. Large geographic datasets

Detailed geographic datasets are among the largest highly structured datasets maintained. For a single nation these datasets already contain hundreds of millions of identifiable objects at a single moment in time. For many applications this high level of detail is necessary, but, as indicated above, users also need consistent medium and lower levels of detail in representations of the same area. These are called generalised, or aggregated, representations.

NMCAs maintain these levels of detail separately due to the complexity involved in automatically generating a lower level of detail from a higher level. Currently, this is often implemented in completely separate 'production' lines and update cycles based on differences between original purposes. Furthermore, at a given scale different visualisations of the data may be needed for different applications. These huge geographic datasets are continually growing due to growing demands for additional data, such as the following:

- More data providers now also include history, resulting in several versions of the same object.
- Instead of 4- to 6-year update cycles, pressure is growing to provide more frequent updates (databases with daily updates are already operational in some countries).
- More and more types/classes of geographic objects are systematically collected; for example, the Infrastructure for Spatial Information in the European Community (INSPIRE) directive requires 34 different themes (INSPIRE, 2007).
- Due to the ever-increasing pressure on space, there is a need to describe themes at an increasingly higher level of detail and accuracy.

It is clear that we are moving from hundreds of millions to hundreds of billions of objects that have to be managed, maintained, kept consistent, and served to the end-users, and all under growing cost pressure. The emergence of the EU's INSPIRE Directive has introduced a new dimension to the problem: the need to view spatial objects at ranges starting from a continental level, through a national and provincial level, to the local level, while seamlessly zooming into the largest scale of the object available—for example, from a view of the entire River Rhine down to a detailed view of the river at any point along its course. While progress is being made on the ability to fulfil such a requirement, a consistent framework for consistent handling of the continuum of levels of detail is not being defined. The EU directive INSPIRE, articles 8 and 9, which requires data to be consistent between scales and themes, and also between neighbouring regions/countries, provides a non-trivial challenge. The only way an end-user can interact with these huge datasets is via multi-scale or vario-scale representations and related tools in which end-users are barely aware of the transitions between the different scales.

Furthermore, the tendency exists not only to publish the resulting map representations (geographic datasets), but also to hold links from the map objects to the original observations (fieldwork/survey plans, GPS traces, ortho-photos, remote sensing data, and other types of imagery, which can amount to a volume of data of 100s of Terabytes). Finally, there is also a trend towards including large volumes of user-generated (shared) content for which the multi-scale requirements are equally important. In summary, this is a BIG DATA-management challenge. Introduction of multi-scale or vario-scale solutions (levels of detail/aggregation levels) enables the management and use of these huge datasets for a range of applications, including those mentioned in Section 2.

4. Commercial solutions and current limitations

From the early beginnings of handling geo-information in digital environments, software developers have attempted to automate the process of generalisation of geographic information. Traditionally for the production of different map scale series, but more and more in other contexts as well, such as the desktop/web/mobile use of geo-information, the need to process, handle, and understand possibly huge masses of data has existed. A reasonable set of commercial software tools (see the References section for a list) is available to solve parts of the generalisation process. However, the overall generalisation process has not yet been fully automated, including topographic base data and additional thematic datasets in a large range of scales/level of details. In using multi-scale data, some existing products such as Microsoft Live Maps (www.bing.com/maps) and Google Maps (maps.google.com) do seem to offer a satisfactory user experience, especially in very smooth navigation between different scales (and regions). However, this is limited in a number of different aspects:

- Specific themes only—navigation, orientation, and no consistency across the EU.
- Static data—no real-time updates as needed in, for example, crisis management.
- Absence of temporal/historical data, current data only.
- Brute-force redundancy (15–20 explicit levels of details).
- User-added content not automatically adapted at different scales.
- Sophisticated data schemas (i.e., with rich semantics) not supported.

Using these existing technologies in a time-critical crisis-management scenario, such as the coordination of a large rescue operation, would be impossible. All current solutions for supporting different levels of detail are based on (static) copies that are (redundantly) stored at these levels. This makes it impossible to dynamically adapt the map to new information and to changing context of the user. In addition to the classic geo-information visualisation requirement (i.e., supporting different scales), which has been only partially solved, new requirements also exist for generalisation, making it even more difficult. For instance, generalisation should be dynamic and suitable for progressive transfer. Furthermore, objects to be visualised have expanded in dimension, as in the emerging fields of 3D and temporal data.

To make further progress in automated machine generalisation, both the semantics of the spatial information and the user needs should be further formalised. Methods and techniques from the semantic web might be useful in this formalisation, and tools from knowledge engineering could be applied in the actual generalisation based on reasoning (see Section 7). Interpretation of spatial

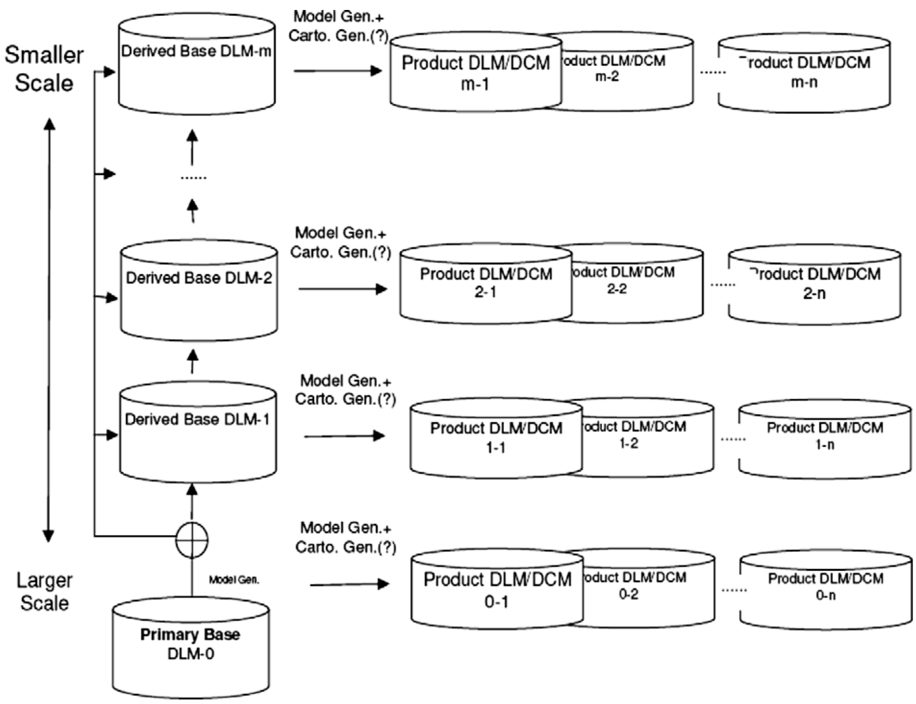

Fig. 3. Multi-representation databases (Zhou, Regnauld, & Roensdorf, 2008).

constellations or situations is a process that is closely linked to human capabilities and can be formalised using formal semantics; e.g., Web Ontology Language (OWL), Ontology Definition Metamodel (ODM). Making implicit information explicit is necessary, not only for many spatial analysis problems, but also for aspects of information communication.

Spatial data also pose exciting questions for the algorithms-building and data-structuring communities. It is vital that specialists in the field of computational geometry meet with the spatial data community to exchange ideas, pose problems, and offer solutions. Most algorithmic problems arising in that field are indeed geometric. In this context it must be noticed that the focus is more and more on 3D (and 3D plus time) geometric computations (see Section 7). In this respect, generalisation operations and the resulting data have to be understood as processes, which will allow a broader and more flexible usage and re-generalisation when changes in reality have occurred. For a mass market (e.g., consumers of mobile maps), the human factor is very important. The currently available solutions (often mobile maps) still have inadequate user interfaces. The issue of context is extremely important, because users can get 'lost' very easily on small mobile displays when zooming and panning (see Section 6). Based on a selection of use cases, such as navigation and tourist support, user-centred design techniques should be applied to evaluate the interaction and the quality of the maps.

5. Multiple representation databases

As indicated in the introduction, a multiple representation database (MRDB) may be seen as a data producer's issue. However, it could also be considered as a data user's issue when the user starts to develop applications exploiting the MRDB, for example, to smoothly navigate from one representation to another representation. As shown in Fig. 3, this results in a large number of related datasets for which consistency should be maintained—a task that has already gone beyond the capabilities of NMCAs in Europe and around the world. It should be noted that multiple representations do not always have to be related to different scales and the links between the objects at these different scales in the MRDB. An example of multiple representations at a single scale is two geometries that represent the same feature, for example, a polygon representing the road area needed for one set of applications and a polyline representing the road centre-line needed for another set of applications. This principle is applied within the Netherlands Cadastre in the Topographic data for scale 1:10,000 (TOP10NL), which indeed has two different representations of road networks for different applications: road centre-lines for car navigation and areas for road maintenance. Whatever type of multiple representations is intended, it is tempting to perform the maintenance and data collection only at the level of the base datasets and propagate this via the links in the MRDB to the related representations.

Although all National Mapping Agencies (NMAs) currently have different representations for different scales (i.e., an MRDB), a short poll of the NMA participants at the ICA 2008 workshop made it clear that the explicit links between the various representations are almost never maintained in practice today. So, one could indeed pose the question of whether these are true MRDB solutions, where one does expect some explicit support for multiple representations, for example, links. Within IGN France[1], one situation was mentioned where explicit links within an MRDB are maintained, but this was related to yet another type of multiple representation: the link between objects is their base Digital Landscape Model (DLM) to the Digital Cartographic Model (DCM) at the same scale; the horizontal line in Fig. 3. The current situation in France is that it is still difficult to build, and especially maintain, these links. At the moment, the UK Ordnance Survey has links only at the conceptual level in their data modelling framework. Another form of multiple representations was mentioned by users of the Ordnance Survey data: the 'plan, build, live, extend, tear-down' cycle for handling geographic objects. Users talk about real-world objects that are not always also represented on a map, but they are interested in a particular phase of the geographic object (in which objects might not even exist; e.g., only as a plan). Updates generate history and versions of representations. Do we need temporal links for those kinds of multiple representations? A fundamental question here is whether or not an object and its identity change. How one sees the concepts versus the reflection in the physical model (e.g., using the same identifier, or a link to different identifiers) is important in this instance. In summary, several types of multiple representations can be identified due to:

1. differences in scale;
2. DLM-DCM separation;
3. temporal/life-cycle (plan, realise, modify, remove); and
4. one-object multiple geometric representations (e.g., road area/centre-line, or building a 2D footprint/3D detailed model).

All of these types of representations are related and certainly not mutually exclusive. For example, updating will produce history (temporal versions), but must also be applied to both DLM and DCM at the different scales for each of the geometric representations of an object. In some countries the DCM is occasionally even more up to date than the DLM. This is caused by the fact that manual work is needed when going from DLM to DCM, when at the same time higher requirements exist in actuality (e.g., as was the case in Spain at the ICC). Besides the technical aspects of how to design and use an ideal multiple-representation model, another important aspect is whether it is worthwhile for the data producers to implement such a model, which includes smart links. This requires a huge effort by the NMAs, with clear costs, to create all of the required links. But what are the benefits, and are these only for the data producer or for the data user as well? NMA itself can then benefit from update propagation. And users can in the end benefit from higher update rates. It is further necessary, when the links are initially created/added, that the edit tools support maintaining the links. An interesting development is that not a single organisation may be responsible for the content of an MRDB, but that the data are increasingly maintained by a 'collective network'. Each actor can have different update rates in such a network. Are NMAs in this network the big integrator in which they maintain one very large geographic database? What is their future role? Regarding authentic registrations, for example, should NMAs always be able to re-create a geographic situation at a specific moment in time?

6. Continuous/gradual generalisation

Continuous generalisation is a term that implies smooth transformations when switching between scales, that is, when performing generalisation for a continuous scale range. However, in reality these scale changes are often small steps, much smaller than the fixed number of scales now typically produced by NMAs. Continuous (or smooth) zooming may use these small steps in combination with 'display' techniques such as morphing, shrinking and fading. A number of different motivations can be listed for generalisation resulting in gradual perception:

[1] In other countries there are related developments; in addition to France (Lecordix, Gallic, Gondol, & Braun, 2007), also in Belgium (Féchir & De Waele, 2005) and in some of the German Bundeslander (Stoter, 2005).

1. user interface aspect: smooth zooming (continuous);
2. data management aspect: non-redundant data storage (compared to multiple representations); and
3. data communication: progressive transfer of refinements.

When trying to put into practice smooth or gradual generalisation, one should realise that different users might have different preferences for different themes. In a user interface this could be specified via theme/object class sliders indicating the relevance of the different themes. The resulting continuous scale representation should then be created based on user preferences, and for all scale levels representation should be consistent. An example in which two different approaches (models) supporting both gradual generalisation, even reflecting the same theme, should not result in inconsistent representations as in the context of terrain/elevation modelling. One approach is to have a raw point/triangle-based representation that can be queried at different scales to select the appropriate number of points/triangles. Another approach is to model the critical points—peaks, pits, and saddle points—and base the generalisation structure on the merging of the neighbouring critical points. For example, two peaks and a saddle point can be merged to form one peak, and resulting structure can be stored in a graph. Furthermore, a third approach is in use at the NMAs and that is generalisation of the terrain elevation based on contour lines. Even if this might be considered the same theme (elevation), keeping these different multi- or gradual-scale representations consistent is an issue.

One option to enable non-redundant representations when implementing smooth zooming is to apply more procedural techniques. This should in principle work in both directions: from a detailed to a coarse procedure (e.g., a procedure to collapse two neighbouring buildings) and from coarse to detailed (in this case the procedure does need some 'delta' information as input parameters). Especially in the case of re-occurring patterns, a procedural approach may be efficient (as in the PostScript printer language).

Does it matter for gradual generalisation whether the theme is related to discrete (e.g., man-made objects, roads, buildings) or to more continuous themes (often natural; e.g., elevation and vegetation)? The answer to this question is not so clear. However, it is obvious that there are differences. For example, currently a set of 10 building polygons may be generalised to three building polygons (on the next smaller scale), then to one built-up-area polygon (next smaller scale), and finally represented by a point (on the smallest scale). How does one deal with this in the context of continuous generalisation? In summary, there are many open questions in the field of continuous/gradual generalisation.

7. Semantics and computational geometry

Members of the 'generalisation research community' feel strongly that both formal semantics (theories and tools) and computational geometry (theories and tools) are needed to somehow solve the map generalisation challenge (see the success of the Dagtsuhl seminar and recent publications). However, the two theories and tools are currently quite far apart: computational geometry needs well-defined (geometric) problems as input, and a small re-phrasing of the problem statement may lead to a completely different problem and solution. On the other hand, formal semantics might try to function in an environment that is less well defined (and contains at times even contradictory information).

Still, it is quite plausible that with the help of formal semantics it will be possible to develop more flexible solutions, that is, to avoid hard-coding everything (e.g., which computational geometry solution to apply to which objects/situations), and to use some generic intelligence to characterise this and apply the correct computational geometry solution. Taking generalisation as an example, it may be beneficial to characterise the problem area (say, rural or urban environment) and impose some hierarchical classification of object types (e.g., on the top level: natural versus man-made, then further refinement). Then a decision can be reached, based on formalised knowledge, as to which solution to apply. Also, the users' wishes and requirements could be formalised, for example, using semantic web technologies (and therefore being applied in a more flexible solution). Once these characteristics have been defined, computational-geometry tools could be used in two scenarios, namely: (i) to analyse a specific dataset or collection of instances and attach to this the proper characterisation (including more complicated patterns) and (ii) after having characterised the situation and understood the user requirements, the reasoning process (based on formal semantics) could determine which computational geometry tool to apply in which situation and in which order.

Despite the above-described complementary value of the two approaches, it is not yet clear how the benefits could be obtained in practice. Perhaps one could try to realise some kind of hybrid system that would manage the collected generalisation knowledge. The system could be considered as an 'automated designer' for the development of a generalised process in a specific situation, given datasets and user preferences. Or perhaps more modestly stated, the system could orchestrate the workflow within the generalisation process. In any case, it is fair to state that more research is needed in this area.

8. Overview of accepted papers

It is easy to say that generalisation is a difficult problem and that current commercial software solutions do not yet provide satisfactory results. But for a fair evaluation of the various solutions in the form of software products, it must be explicitly known what is required in the generalisation process. This is the first result presented in the paper by Stoter et al. (2009). The paper proceeds by describing the evaluation of the generalisation results of four cases tested by different generalisation software products, taking the NMA requirements into account. As both human and machine evaluations are applied (with possible inconsistencies), the two have to be combined into a final evaluation. The paper concludes with reflections on the developed methodology and identifies areas for further research.

The paper by Stanislawski (2009) focuses on a specific generalisation problem: the pruning of a hydrographic network and related drainage areas. As is often the case, the existing dataset did not explicitly contain all the information as used in the pruning/generalisation rules, so first a data-enrichment phase was applied assigning relative prominence to the network features. Using the enriched dataset, it was then possible to perform good-quality pruning; this was applied to a test area of 48 sub-basins at the 1:24,000-scale, which was pruned to the 1:100,000-scale. The results were then compared to the benchmark dataset: the standard (and independently produced) 1:100,000-scale hydrographic data. Finally, an indicator, the coefficient of linear correspondence (CLC), was developed to estimate the matching between the two datasets.

Accepting the fact that fully automatic on-the-fly generalisation is not feasible and that datasets for several scales have to be maintained, the challenge is then to create an environment in which this is most possible. In the paper by Zhou, Regnauld, and Roensdorf (2009), a solution is proposed that supports update propagation between different scales. The results of the interactive generalisation process, which were a mix of manual and automated actions, are reflected in a directed a-cyclic graph. The graph includes the features with different scales—the generalisation operators and the applied parameter values—between them. In addition to representing the full generalisation history, the model also includes the feature versioning mechanism, which is of course closely related to the update process. The proposed solution has been implemented in the Gothic object-oriented database management system.

Also, the paper by Chaudhry, Mackaness, and Regnauld (2009) accepts the multi-scale representation approach. In the paper the authors argue that, for topographic base maps, the 'functional perspective' has advantages over the 'geometric perspective' (as, for example, used in the hydrographic network generalisation in the paper by Stanislawski). Functional compositions, such as hospitals, airports, or cities, are the key elements. They are modelled using various relationships such as partonomic, taxonomic, and topological and lend themselves directly to analysis in general and generalisation in specific. However, first the functional units have to be detected (again, through data enrichment, as this information is not available), and the authors present an approach that borrows ideas from robotic vision. The paper then shows how the dataset enriched with explicit functional units can be used for multi-scale representation and generalisation.

For machines to perform high-quality generalisation in a fully automatic process in the future, somehow the 'implicit' knowledge, as used by humans when reasoning about geographic concepts and when performing generalisation based on these concepts, must be made explicit and available for machines. The paper by Lüscher, Weibel, and Burghardt (2009), tries to formalise some of these concepts. The paper uses the example of buildings in a city and develops a (partial) ontology for this. A method is presented to classify the different (building) objects in a map according to the ontology, given the fact that there is a certain amount of 'vagueness' in the definition of the concepts. Supervised Bayesian inference is used for inferring complex concepts, such as the example used of the English 'terraced house' concept. Further, classification tests were conducted on datasets representing buildings in four different cities. The research supports integrating vague, but important, knowledge about conceptualisations in cartography and therefore enables better generalisation.

It is quite realistic to assume that the future will bring more and more 3D geo-information, whereas today 2D geo-information still dominates. It is therefore likely that the need for generalisation will grow for various applications in 3D, among which is, of course, efficient 3D visualisation. In their paper, Glander and Döllner (2009) describe their initial attempts to use 3D generalisation techniques in the context of visualisation of large city models. Their emphasis is on satisfactory interaction with the users, and the system must therefore present the 3D information at the most relevant abstraction level, that is, distance- and importance-based. The authors used the infrastructure network to create the several levels of aggregated representations, for example, groups of buildings are replaced by a more abstract 3D 'cell block'. Also, a landmark hierarchy is computed and related to more abstract representations. In a number of examples, the authors show the use of their techniques and demonstrate smooth visualisation of transitions among pre-computed representations including dynamic landmark highlighting.

The authors of the last paper in this special issue attempt to eliminate the fixed scales that are so well known. A modest attempt is made to set up a vario-scale environment, with the condition that when 'arriving' at the well-known (fixed) scale, the vario-scale based representation should be equal to its fixed-scale counterpart. van Oosterom, Dilo, and Hofman (2009) explain their approach to providing good-quality vario-scale representations. The presented structure is a single non-redundant representation, which can be queried at any arbitrary scale between the source/input data scales. The input consists of the largest scale and (one or more) smaller scale datasets, which act as constraints in the generalisation process. The generalisation process results are captured in a structure that contains accurate geometry of the large-scale objects, enriched with the generalisation knowledge of the medium-scale data. Real topographic data in the large (1:1000) and medium scales (1:10,000) range from two Dutch cities, Almere and Rotterdam, where they were used for validating the proposed method.

Acknowledgment

Parts of this paper are based on the inspiring discussion (or break-out) sessions at the most recent ICA workshop on 'Generalisation and Multiple Representation', Montpellier, France, 20–21th June 2008 and the seminar 'Generalization of spatial information' at Schloss Dagstuhl – Leibniz Center for Informatics, Wadern, Germany. I would like to thank the participants for their ideas and focused generalisation discussions. Special thanks to Jean-Claude Thill, Elfriede Fendel and Jantien Stoter for proofreading this paper and for suggesting many useful corrections and additions.

References

Chaudhry, O. Z., Mackaness, W. A., & Regnauld, N. (2009). A functional perspective on map generalisation. *Computers, Environment and Urban Systems, 33*(5).
Chaudhry, O., & Mackaness, W. A. (2008). Automatic identification of urban settlement boundaries for multiple representation databases. *Computers, Environment and Urban Systems, 32*(2), 95–109.
Féchir, A., De Waele, J. (2005). The future production of generalised maps at IGN Belgium. In 8th ICA (International Cartographic Association) workshop on generalisation and multiple representation, A Coruña, Spain, 7–8 July 2005.
Glander, T., & Döllner, J. (2009). Abstract representations for interactive visualisation of virtual 3D city models. *Computers, Environment and Urban Systems, 33*(5).
INSPIRE. (2007). Infrastructure for spatial information in the European community; see *Official Journal of the European Union, L, 108*(50), 25 April 2007 (<http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2007:108:SOM:EN:HTML>).
Kiester, A. R., & Sahr, K. (2008). Planar and spherical hierarchical, multi-resolution cellular automata. *Computers, Environment and Urban Systems, 32*(3), 204–213.
Lecordix, F., Gallic, J.L., Gondol, L., Braun, A. (2007). Development of a new generalization flowline for topographic maps. In 10th ICA workshop on generalisation and multiple representation. Moscow, Russia, 2–3 August 2007.
Lüscher, P., Weibel, R., & Burghardt, D. (2009). Integrating ontological modelling and Bayesian inference for urban concept classification in cartographic vector data. *Computers, Environment and Urban Systems, 33*(5).
Nöllenburg, M., Merrick, D., Wolff, A., & Benkert, M. (2008). Morphing polylines: A step towards continuous generalization. *Computers, Environment and Urban Systems, 32*(4), 248–260.
Stanislawski, L. (2009). Feature pruning by upstream drainage area to support automated generalization of the United States national hydrography dataset. *Computers, Environment and Urban Systems, 33*(5).
Stoter, J. (2005). Generalisation: The gap between research and practice. In 8th ICA (International Cartographic Association) workshop on generalisation and multiple representation, A Coruña, Spain, 7–8 July 2005.
Stoter, J., Burghardt, D., Duchene, C., Baella, B., Bakker, N., Blok, C., et al. (2009). Methodology for evaluating automated map generalisation in commercial software. *Computers, Environment and Urban Systems, 33*(5).
Tomko, M., Winter, S., & Claramunt, Ch. (2008). Experiential hierarchies of streets. *Computers, Environment and Urban Systems, 32*(1), 41–52.
van Oosterom, P., Dilo, A., & Hofman, A. (2009). Constrained tGAP for generalization between scales: The case of Dutch topographic data. *Computers, Environment and Urban Systems, 33*(5).
Zhou, S., Regnauld, N., Roensdorf, C. (2008). Towards a data model for update propagation in MR-DLM. In Paper presented at 11th ICA workshop on generalisation and multiple representation, Montpellier 2008.
Zhou, S., Regnauld, N., & Roensdorf, C. (2009). Generalisation log for managing and utilising multi-representation spatial database in map production. *Computers, Environment and Urban Systems, 33*(5).

Annex: Overview of Commercial software, books and Ph-D theses on generalisation

Commercial software

Axpand and SysDab (Axes Systems), for map generalization; see <http://www.axes-systems.com>.
Change, Push, Typify for resp. simplification of buildings, displacements of objects and typification of buildings; see <http://www.ikg.uni-hannover.de/de/dienstleistungen/change>, push, typify.

ESRI's Research prototype "Optimizer Engine", presented in J. Monnot, P. Hardy, D. Lee (Eds.) (2007). An optimization approach to constraint-based generalisation in a commodity GIS framework. International Cartographic Conference, July 2007, Moscow, Russia.

Radius Clarity, 1Spatial, a rule based environment for automated generalization; see <http://www.1spatial.com/products/radius_clarity/index.php>.

Books

Buttenfield, B. P., & McMaster, R. B. (Eds.). (1991). *Map generalization: Making rules for knowledge, representation*. London: Longman.

Joao, E.M. (1998). Causes and consequences of map generalization. Taylor and Francis.

Zhilin, Li. (2006). *Algorithmic foundation of multi-scale spatial representation*. CRC Press.

Mackaness, W., Ruas, A., Sarjakoski, L.T. (Eds.) (2007). Generalisation of geographic information: Cartographic modelling and applications. A volume in the International Cartographic Association series. Elsevier.

Molenaar, M. (1996). Methods for the generalization of geo – databases. Netherlands Geodetic Commission, 1996.

Mueller, J.C., Lagrange, J.P., Weibel, R. (Eds.) (1995). GIS and generalization, methodology and practice. Taylor and Francis, London.

van Oosterom, P. (1994). *Reactive data structures for geographic information systems*. UK: Oxford University Press.

PhD-Theses

Badard Thierry. (2000). Propagation des mises à jour dans les bases de données géographiques multi-représentations par analyse des changements géographiques. Thèse de doctorat soutenue à l'Université de Marne la Vallée.

Beard, K. (1988). Multiple representations from a detailed database: A scheme for automated generalization. Ph.D. Thesis, University of Wisconsin–Madison.

Boffet, A. (2001). Méthode de création d'informations multi-niveaux pour la généralisation cartographique de l'urbain. Thèse de Doctorat soutenue en 2001. Université de Marne la Vallée.

Burghardt, D. (2000). Automatisierung der kartographischen Verdrängung mittels Energieminimierung. Ph.D. Thesis, Deutsche Geodätische Kommission.

Duchêne, C. (2004). Généralisation par agents communicants : Le modèle CARTACOM. Application aux données topographiques en zone rurale. PhD thesis, Paris VI – Pierre et Marie Curie.

Cecconi, A. (2003). Integration of cartographic generalization and multi-scale databases for enhanced web mapping. Ph.D. Thesis, Department of Geography, University of Zurich.

Galanda, M. (2003). Automated polygon generalization in a multi agent system. Ph.D. Thesis, Department of Geography, University of Zurich.

Harrie, L. (2001). An optimisation approach to cartographic generalisation. Ph.D. Thesis, Department of Technology and Society, Lund University.

Haunert, J.-H. (2008). Aggregation in map generalization by combinatorial optimization. Ph.D. Thesis, Leibniz Universität Hannover.

Jaakkola, O. (1997). Quality and automatic generalization of land cover data. Ph.D. Thesis, University of Helsinki, Finland.

Kilpelainen, T. (1997). Multiple representation and generalisation of geo-databases for topographic maps. Ph.D. Thesis, Finnish Geodetic Institute.

Mustière, S. (2001). Supervised learning for cartographic generalization. Ph.D. Thesis (in French), Université Pierre et Marie Curie, Paris IV.

Neun, M. (2007). Data enrichment for adaptive map generalisation using web services. Ph.D. Thesis, Department of Geography, University of Zurich.

Peng, W. (1997). Automated generalization in GIS. Ph.D. Thesis, Wageningen University.

Regnauld, N. (1998). Généralization du bâti: Strucuture spatiale de type graphe et répresentation cartographique, Ph.D. Thesis, Université de Provence, Aix-Marseille 1.

Richardson, D.E. (1993). Automated spatial and thematic generalization using a context transformation model. Ph.D. Thesis, Wageningen University.

Ruas, A. (1999). Modèle de généralisation de données géographiques à base de contraintes et d'autonomie. Doctoral Thesis, Université de Marne-la-Vallée.

van Oosterom, P. (1990). Reactive data structures for geographic information systems. Ph.D. Thesis Department of Computer Science, Leiden University.

van Smaalen, J. (2003). Automated aggregation of geographic objects. Ph.D. Thesis, Wageningen University.

Timpf, S. (1998). Hierarchical structures in map series. Ph.D. Thesis, Technical University of Vienna.

Weibel, R. (1989). Konzepte and experimente zur automatisierung der reliefgeneralisierung. Ph.D. Thesis, Department of Geography, University of Zurich.

Yaolin, Liu. (2002). Categorical database generalization in GIS. Wageningen/ITC.