AUTOMATED GENERALISATION WITHIN NMAs in 2016

Jantien Stoter\textsuperscript{a,b}, Vincent van Altena\textsuperscript{a}, Marc Post\textsuperscript{a}, Dirk Burghardt\textsuperscript{c}, Cecile Duchêné\textsuperscript{d}

\textsuperscript{a}Kadaster, The Netherlands, \texttt{[marc.post@vincent.altenavan@jantien.stoter@kadaster.nl]}  
\textsuperscript{b}Delft University of Technology, The Netherlands, j.e.stoter@tudelft.nl  
\textsuperscript{c}University of Dresden, Germany, Dirk.Burghardt@tu-dresden.de  
\textsuperscript{d} IGN, France, Cecile.Duchene@ign.fr

Special session – SpS 12 - EuroSDR: Innovative technologies and methodologies for NMCAs

KEY WORDS: Map Generalisation, Cartography, Scale, Multiple Resolution Databases, Automated Generalisation

ABSTRACT:
Producing maps and geo-data at different scales is traditionally one of the main tasks of National (and regional) Mapping Agencies (NMAs). The derivation of low-scale maps (i.e. with less detail) from large-scale maps (with more detail), i.e. generalisation, used to be a manual task of cartographers. With the need for more up-to-date data as well as the development of automated generalisation solutions in both research and industry, NMAs are implementing automated generalisation production lines. To exchange experiences and identify remaining issues, a workshop was organised end 2015 by the Commission on Generalisation and Multiple Representation of the International Cartographic Association and the Commission on Modelling and Processing of the European Spatial Data Research. This paper reports about the workshop outcomes. It shows that, most NMAs have implemented a certain form of automation in their workflows, varying from generalisation of certain features while still maintaining a manual workflow; semi-automated editing and generalisation to a fully automated procedure.

1. INTRODUCTION

The automatic derivation of small-scale maps with less detail from larger-scale maps with higher degree of detail (see Figure 1) has been intensively studied during the last twenty years. For a long time this process, called “automated generalisation”, has been considered as something to be never reachable (Anderson-Tarver et al. 2011) for National Mapping Agencies (NMAs). Mainly because of the complexity of automating a process that heavily depends on human interpretation. An automated generalisation solution would significantly reduce the costs and time required for producing multi-scale map series. In addition, for Spatial Data Infrastructures automated generalisation would be one of the keys to collect and maintain geographical information once and to use it many times by dynamically deriving a map with the required content and at the required level of detail when needed.

![Figure 1. Map series of Netherlands’ Kadaster (1:10k, 1:50k, 1:100k, 1:250k, 1:500k)](image)

1.1 Developments within NMAs

At the beginning of this century a few NMAs had implemented automated generalisation in specific parts of their production lines (Stoter 2005; Foerster et al, 2010). However, only recently NMAs succeeded to implement automated generalisation to derive a complete map with no or little human interaction (Duchêne et al. 2014). These achievements have several reasons First, successful research results have found their way into software. Secondly, because these software solutions do not provide off-the-shelf solutions, NMAs had to be willing to seriously invest in the development of automated generalisation workflows with the available toolboxes. Recently, NMAs were encouraged to do so by an increasing call from society to produce up-to-date maps. Update cycles of 4-6 years used to be common, but are not acceptable in our current information society in which people are able to capture and compare spatial information about the current environment, e.g. with help of smartphones and GPS devices. A final reason that made it possible to automate the production of multi-scale maps, related to the just mentioned increasing information demand of society, is that the output is no longer only driven by cartographic criteria, but also by the up-to-dateness. Consequently, resulting maps may be considered “good enough” by users even though they may not be cartographically perfect.

1.2 Sharing experiences

Although automated generalisation has resulted in successful implementations, open issues remain. To identify the state-of-the-art and to discuss remaining issues a workshop was organised by the Commission on Generalisation and Multiple Representation of the International Cartographic Association and the Commission on Data Modelling and Processing of EuroSDR (European Spatial Data Research) on 3\textsuperscript{rd} and 4\textsuperscript{th} December 2015 at Kadaster in Amsterdam. Over 60 people from 18 NMAs exchanged experiences on this topic and discussed issues for further research. Questions that were addressed during the workshop are:

- How to deal with heterogeneous source data, which are more and more frequent since NMAs act more and more as integrators of data produced by other administrations?
- How can successful implementations of some NMAs be transferred to others while all NMAs have their specific context? A key for success is for example a clean and semantically rich source data set.
- What does “good enough” mean in terms of cartographic generalisation? Some NMAs chose to go for fully automated processes while accepting a lower cartographic quality or a less rich content of the resulting maps, while...
others prefer to keep some manual edits to assure the best cartographic quality.

- If generalisation of a complete map is feasible, is there still a need to maintain object identifiers for the derived products? Maintaining these identifiers also implies to support incremental updates as part of the automated generalisation process. This is yet an unresolved problem.

- How can maps be generalised on-the-fly as required when disseminating these via the web within SDIs? Since current automated generalisation solutions do not fulfil this requirement, maps at intermediate zoom levels are currently being pre-processed.

- Is it necessary to distinguish and maintain Digital Landscape Models and Digital Cartographic Models at several levels of detail?

- How can automated generalisation solutions be used to derive on-demand maps for different purposes (hiking, cycling, water navigation etc.)?

1.3 Structure of the paper

The workshop consisted of presentations from the participating NMAs about their multi-scale workflows as well as of break out sessions on common issues. This paper gives an overview of the main workshop outcomes. It starts with an introduction of attendees (section 2) and continues with the state-of-the-art in automated generalisation as concluded from the workshop materials (section 3). Section 4 elaborates on one of the open issues in automated generalisation related to updates of multi-scale maps with either support of incremental updates or automatically generalise a complete map in case of updates. The paper closes with concluding remarks in section 5.1

2. BACKGROUND OF ATTENDEES

The interest in the themes of automated generalisation and multiple resolution databases is increasing, as is evidenced by the number and background of participants. At the first NMA workshop, held in Barcelona in 2013 (Duchene et al. 2013), there was a group of approximately 25 participants from 12 countries/regions, all from a regional or national mapping agency. This second NMA workshop was attended by a larger audience with diverse backgrounds: besides the 48 attendees from 17 NMAs, nine representatives from two software vendors were present, as well as four academics from different universities (see Figure 2).

Also the variation in lands of origin increased: attendants were from Belgium, Czech Republic, Denmark, Finland, France, Germany, Ireland, Israel, Latvia, the Netherlands, Norway, Poland, Spain, Sweden, Switzerland, Turkey and United Kingdom (see Figure 3).

Figure 3. States of origin of participants in 2013 (small dot) and 2015 (large dot)

Given the diverse backgrounds, also the experiences with automated generalisation and map production is wide-ranging. Some NMAs should be considered as veterans, having spent more than twenty years on studying generalisation, while others are relatively new in the domain, having started one or two years ago. And there are many NMAs in between. As could be expected, most veteran NMAs have implemented semi-automatic procedures years ago, while the novice NMAs are targeting on fully automated procedures. Another evident difference between veterans and novices is the scale of the base data: nowadays NMAs start with the largest available scale, while in earlier years mid- and smaller scale base data were generalised by semi-automated procedures. Figure 4 illustrates the shift in focus to large-scale data as input data.

Figure 4. Source map scales of generalisation processes at NMAs

The focus of target data is also shifting to larger map scales as figure 5 illustrates.

Figure 5. Target map scales of generalisation processes at NMAs.
Most NMAs use similar software in their production workflows (most of the time highly customised): 1Spatial – 1Generalise (Regnauld 2015), Esri ArcGIS (Hardy 2015), Clarity, SCAN Express, Safe FME, Geomedia, Lamps2, Astrapd and Peikka. Databases in use are Oracle, Gothic and filegeodatabases.

3. STATE OF ART WITHIN NMA’S

At the workshop, each NMA provided an abstract and presentation on the state of the art of “Automated Generalisation” within its own organisation. Details can be found in the corresponding abstracts and presentations: (Augustyn 2015; Baella et al. 2015; Celik & Simav 2015; Curtinot 2015; Færch-Jensen et al. 2015; Fehir 2015; Frick & Johansson 2015; Goldner 2015; Haug 2015; Howland & Walters 2015; Käuberle et al. 2015; Kettunen et al. 2015; Lebiecki 2015; Madden et al. 2015; Reimerink 2015; Urbanke 2015; Wehrhan 2015).

The paragraphs below provide a summary of the most significant items: (1) What are the drivers for studying and implementing automated generalisation? (2) How is the issue of generalisation approached? (3) How are the individual NMAs progressing in the field of automated generalisation? (4) Which level of automation is achieved in the generalisation process? (5) How much tailoring of out-of-the-box software is applied and at what expenses? (6) What are the open issues currently being met and/or addressed?

1. Drivers for automated generalisation: NMAs in the twenty-first century are challenged by several issues. First, society is demanding data with a higher update cycle. The speed of traditional update cycles, between five to ten years, does not meet the need of society nor legal demands. Second the economical crisis caused austerity measures and accompanying budgets cuts, which made costs-effectiveness an even more urgent issue. Actual staff sizes and traditional workflows were not sufficient to comply to legal demands or to the pressure from society to obtain derived products for visualisation in internet and mobile devices. A third driver for automated generalisation, open data policy, is faced by some countries. Automation of generalisation contributes to solving this issue in two ways: some NMAs seek to (fully) automate their workflow to reduce costs and opening up their data to society, while others use automation to provide open data-sets as a ‘light’- alternative to their premium data.

Fourth, Volunteered Geographical Information - initiatives such as Open Street Maps and commercial solutions like Google Maps, Bing Maps challenge the original tasks of governmental topographical datasets in the 21st century.

Fifth, more and more NMAs have to cope with the incorporation of external datasets. Some of the NMAs do not have a full coverage in their source database, or policies are to combine data from multiple providers to one data set. This urges NMAs to look into efficient procedures for quality control and automation of manual procedures.

2. Approaches for generalisation: Although most NMAs are facing the same challenges, the approaches to study and implement automated procedures vary. This variation is explained by the history and culture within an NMA (Duchêne et al. 2014). Some NMAs have a vivid and lengthy history in studying automation of mapping procedures, which has also contributed to the current implementation of algorithms in commercial software. Other NMAs started recently, driven by external factors as explained in the previous paragraph.

The veteran NMAs usually approached the study of automated generalisation in a classical form, starting with specifications, technical design and implementation. New coming NMAs tend to approach the issue from a more pragmatic point of view looking at the lessons learnt from other NMAs, interacting with customers and working together with industry partners.

A common approach of automating the generalisation process of multiple scales from a single source cannot be recognised: both star (where all small scale maps are derived from the same most detailed map scale) and ladder approaches (where maps are derived from the next-higher scale map) are applied, as well as a hybrid form which mixes star and ladder approaches.

3. Progress: The progress in the implementation of automated procedures varies from NMA to NMA. Some NMAs have just started preparations and still have to face the challenge to create multi-scale data models and implement them in appropriate databases, e.g. relational or object oriented databases (Augustyn 2015). Others have small-scale-generalisation processes implemented for years (Kettunen et al. 2015; Lebiecki 2015), and are now developing towards the design of a large scale architecture for generalisation. Novices in the field of automated generalisation, who have taken the hurdle of designing and implementing data models for their map data tend to start with automation of large scale databases (10K to medium-scale databases (50k). Three NMAs have implemented fully automatic procedures (Stoter et al. 2014; Regnauld 2014; Lafay et al. 2015) and other NMAs aim on following within the near future (Madden et al. 2015; Frick & Johansson 2015).

Many others have automated parts of the generalisation workflow.

Alternative approaches are applied by Danish Geodata Agency (Færch-Jensen et al. 2015) and Norwegian Mapping Authority (Haug 2015) since they have to cope with external source data provided by manifold municipalities. Their main focus is on quality assurance and control as well as on leveraging customizable products via web portals.

4. Level of implementation of automated processes: The extent of implementation and thoroughness in implementation within NMAs varies. Most NMAs have implemented a certain form of automation in their workflows. The level of implementation varies from generalisation of certain features while still maintaining a manual workflow; semi-automated editing and generalisation to a fully automated procedure (like the aforementioned three examples). Most of the NMAs that have implemented semi-automated workflows planned to substitute these by fully automated workflows within the next two years (2016-2018).

5. Tailoring: While some NMAs use out-of-the-box tools only, most of them develop specific customisations of the out-of-the-box tools, often outsourced to the software provider. The main instruments to develop additional custom-designed workflows are geoprocessing tools and python scripts. Incidentally, more advanced programming languages (such as ArcObjects or C++) are used to tailor generalisation processes.

6. Issues: Several NNMA’s are facing obstacles to explore, develop or implement automated generalisation. For instance, NMAs lack the required amount of staff and budget resources. Also the role of several NMAs has changed from data collector into data distributor. This causes several NMAs to sincerely reflect upon their existing workflows, upon the use of external source data, quality demands, update cycles and processing units. It is not said these reflections necessarily cause the replacement of existing workflows (sometimes a considerable amount of time and budget has been spent and redesign does not improve the end results per sè), but it opens up doors to other solutions. In addition, it causes the discontinuation of production flows and the replacement by automatic procedures.
4. UNIVERSAL IDENTIFIER AND THEIR PROPAGATION THROUGH SCALES

With the possibility to automate the generalisation process, an important question arises about the design of future production workflows of multi-scale maps. If generalisation of a complete map is feasible, will updates consist of the generalisation of complete maps or the generalisation of only the updates? The latter requires the support of incremental updates and maintenance of object identifiers of derived products. Besides that object identifier might be used by the NMA themselves for incremental updates, they may also be used by customers of NMA data, who combine topographic with thematic data and may have to update NMA (background) reference data independent or in combination with own thematic data. The maintenance of incremental updates is yet an unresolved problem.

The issue of incremental updates fits within the developments of Life-cycle management and traceability of objects by a Universal Identifier (UID) through several scales. And the question is whether NMAs or customers really need those. From one side, it can be argued that both Lifecycle and Unique Identifier are needed. The arguments are: 1) it is a user-requested feature (but the discussions during the workshop showed that we can mention no or little real examples that shows this and also not why users would need it), 2) for certain applications you do not want to use your source data, 3) NMAs would like to be able to deliver a subset of changes only; 4) UID’s enable implementation of Linked Data concepts also for small scale data, and 5) Software vendors do provide the option. Other arguments advocate why we do not need these lifecycle information or Unique Identifiers for small-scale data and therefore we can suffice with generalisation of complete maps. Objects in the most detailed map should always be identifiable. However UID’s should not be created for derived small scale data, because: 1) given the current hardware and software environments, it is much easier to reprocess a derived dataset completely instead of managing the complex lifecycle information; 2) some argued that the analysis with generalised data should be discouraged, because generalisation decreases the quality of the data which takes away one of the requirements for UID’s; 3) traceability is difficult to maintain and implement at a conceptual level specifically within multi-scale production workflows where objects are aggregated, typified, selected and deleted and in some cases also enlarged and displaced.

To meet both points of view, it was suggested to use the URI for objects and to consider geometrical representations at several scales as attributes. This sounded as an interesting suggestion, but then the question was: What is an object? Do we mean the object in real life or do we mean an object stored in the database? If the latter, how should an object be defined? And how to handle relationships between objects in the case of n-1 or m-n transformations?

Uniformity or a general definition of objects appeared to be an illusion, since an object definition is dependent on the use case, user context and related target scale. Since a “one size fits all” solution was not an option, it was proposed to consider groups of users who can agree upon object definitions. These objects could be provided an URI or UID for the given context. This raised a new question: If we are to do this, are the members of the group willing to pay for these efforts of NMAs, since it will take a considerable amount of time and effort to first agree on these definitions and secondly to implement generalisation processes accordingly. Underneath this question about the URI’s lies another question: “why do users want a URI or UID?” The reason usually mentioned is to carry out updates on reference data, which might be combined with additional thematic information. However, as mentioned before, the workshop participants provided arguments in favour and against but were not able to fully answer this question from the users’ perspective.

5. CONCLUDING REMARKS ON REMAINING ISSUES

From the presentations, abstracts and discussions of the workshop several open issues were identified that need further attention by either academics or industry.

At first, operators for automated generalisation provided by industry are often implemented as black boxes. Since successful generalisation requires adjustments, the generalisers at NMAs mentioned a need for more transparency and better possibilities to experiment with the underlying implemented algorithms.

Another remaining issue in generalisation is the lack of appropriate personnel. Implementing automated generalisation within NMAs requires high-qualified people with knowledge on information technology and skills on data-generalisation. Both are characterised by steep learning curves and the lack of such personnel may hinder the implementation of automated solutions within NMAs.

Also automated generalisation at NMAs requires an improved scalability of processes. One of the challenges for the full automation of the generalisation process lies in the possibility to process a complete country. Besides computer power, this requires a smart way for partitioning to be able to apply area- and context dependent algorithms and parameter values as well as a tool to handle feature morphology (Altena 2014) for morphology-tailored generalisation processes. A good solution for partitioning also includes distribution of the computation as well as the management of dependencies between partitions.

Finally the integration of 2D, 3D and 4D was mentioned as open issue by most of the participants. Many NMAs are making the step from 2D mapping to 3D mapping with maintenance of temporal information (4D). The 3D maps are increasingly considered within the context of multi-scale products and some NMAs even maintain 3D data as source data from which 2D data is derived, like implemented by swisstopo. This brings another challenge for generalisation, i.e. deriving small-scale products via 3D generalisation. While automated generalisation research in 2D has a rich history, research on generalisation of 3D urban models is rather new. Several researchers have studied the generalisation of individual buildings and groups of buildings. However, they often focus on a single generalisation problem while we have learned from the 2D cartographic domain that for successful generalisation solutions it is essential to generalise urban objects with respect to their surroundings. This context dependent generalisation is hard to implement and not yet well understood in 3D (Figure 6).

Figure 6: Context dependent generalisation solutions in 2D extended into 3D. Simplification and amalgamation (above) and simplification and displacement (below)
ACKNOWLEDGEMENTS

We could never have written this paper without the active contribution of all participants of the workshop. Therefore, we would like to thank representatives from the following organisations:

- Bundesamt für Kartographie und Geodäsie / European Location Framework
- Danish Geodetic Agency
- Esri
- Finnish Geospatial Research Institute / National Land Survey of Finland
- General Command of Mapping, Turkey
- Head Office of Geodesy and Cartography, Poland
- IGN Belgium
- IGN France
- Institut Cartogràfic i Geològic de Catalunya
- Kadaster, The Netherlands
- Land Survey Office, Czech Republic
- Landesamt für Geoinformation und Landenentwicklung Baden-Württemberg
- Lantmäteriet, Sweden
- Latvian Geospatial Information Agency
- Norwegian Mapping Authority
- Ordnance Survey United Kingdom
- Ordnance Survey Ireland
- Santiago de Compostela University
- Survey of Israel
- Swisstopo
- T-MAPY spol. s r.o.m, Czech Republic
- Delft University of Technology
- University of West Bohemia
- VUGTK vvi: Czech Republic
- 1Spatial

REFERENCES


Goldner, Shirley. 2015. ‘The Survey of Israel: Topographic 1:25,000, 1:50,000 and 1:100,000 Scale MRDB and Maps’. In NMA Symposium 2015 - Program and Presentations, 2.