Thematic workshop on building an ontology of generalisation for on-demand mapping

William Mackaness, Nick Gould, Sean Bechhofer, Dirk Burghardt,
Cecile Duchene, Robert Stevens, Guillaume Touya

26-27th March, IGN, Paris.

Abstract
A workshop comprising experts in the field of map generalization came together with experts in ontological modelling in order to explore the role of ontologies in map generalization. A sequence of participatory activities identified the need for ontologies related to tasks, algorithms, data quality and truth, content selection, layout constraints, and geographical modelling (the set of relations among a set of map features). The workshop resulted in a deeper shared understanding of relationships between concepts that lie at the heart of map generalization. At the conclusion to the workshop attendees were tasked with developing case studies that will be shared via a wiki, and reported upon in a subsequent meeting devoted to ontological modelling for map generalization. This report is not exhaustive in its presentation of all the ideas and material presented at the workshop, but seeks to convey the over-arching methodology and its core findings.

Introduction

A one and a half day workshop was held in Paris in March at the offices of the IGN, the French national mapping agency. Organised by the ICA Commission on Map Generalization and Multiple Representation[1], the purpose of the workshop was to explore the relevance of ontologies to the map generalization process. The 15 attendees were a mix of practitioners, GIS vendors and researchers (Appendix A). The decision to host such a workshop followed from a previous workshop held in Vienna in September 2014 prior to the GIScience 2014 conference. Various ontology research activities have taken place over the past decade, notably work at IGN France and the Ordnance Survey of Great Britain (Gould et al., 2014). The workshop was seen as an effective way of pooling expertise and experiences, and further progressing collaborative efforts to develop ontologies relevant to map generalization research. Professor Robert Stevens and Dr. Sean Bechhofer (experts in methodologies for collaboratively building ontologies) led the workshop.

Motivation

Over the past 50 years, research has led to the design and implementation of countless algorithms for manipulating features represented in vector space. There now exists different taxonomies of operators and algorithms, and models that anticipate varying levels of human intervention and interaction. One measure of the success of this research is the high level of automation achieved in map generalization for specific map types and scale ranges. Some argue that insufficient emphasis has been given to understanding how the underlying geography affects the choice and application of generalization algorithms. Others have expressed a desire to share various software solutions, whilst others have observed the inconsistencies that exist in definitions of operators and evaluation criteria. There is also a desire to link the user (and their task) more closely with map generalization processes, in anticipation of web based services able to support visualisation of user generated content or other kinds of thematic data, more and more available with the web 2.0 and the open data movement, this data being itself integrated with other data. Can ontologies help us with any of these tasks? Ontologies are not a magic bullet to solving all map generalization research problems but if they can bring a shared clarity as to the underlying concepts, then this will greatly facilitate the sharing (and integration) of research outputs, such as generalisation algorithms.

[1] http://generalisation.icaci.org/

What are ontologies?
An ontology is a knowledge representation. Ontologies are similar in nature to taxonomies but whereas a taxonomy only describes the "type-of" or "is-a-kind-of" relationship, an ontology can describe any relationship, be it hierarchical or not, as well as properties of the concepts. In a taxonomy we might describe a lion as a type of big cat, which is a type of animal. We also might describe a zebra as a type of horse, which too is a type of animal. In an ontology we can additionally assert that the lion "prays-on" the zebra and that the zebra "eats" grass. Ontologies are frequently used to fix the knowledge of a domain such that the "domain experts" have a shared vocabulary; we both agree on what a lion is, its characteristics and behaviours. However, ontologies also support inference, to gain "new expressions from old". For instance, in the “lion and zebra” example, we can infer that the lion is dependent on grass. Ontologies have an Open World Assumption, which can be summarised as meaning that just because we have not asserted something it does not mean it is not true.

The Approach
After introductions, and working in turns and in pairs and small groups over the two days, the participants were given a sequence of tasks to undertake by the two ontology experts (Figure 2). Each task was performed in 20-30 minutes and then each pair (or small group) reported their findings back to the whole group for elaboration and further discussion (Figure 3).

All tasks were guided by the following idea (simple, but important) : the ontology (or ontologies) workshop attendees were trying to define what will be used by one or more systems, dedicated to generalisation and most likely in the context of on-demand mapping. Each of these systems will be used by users (of various profiles), in order to accomplish their respective tasks (Figure 1).

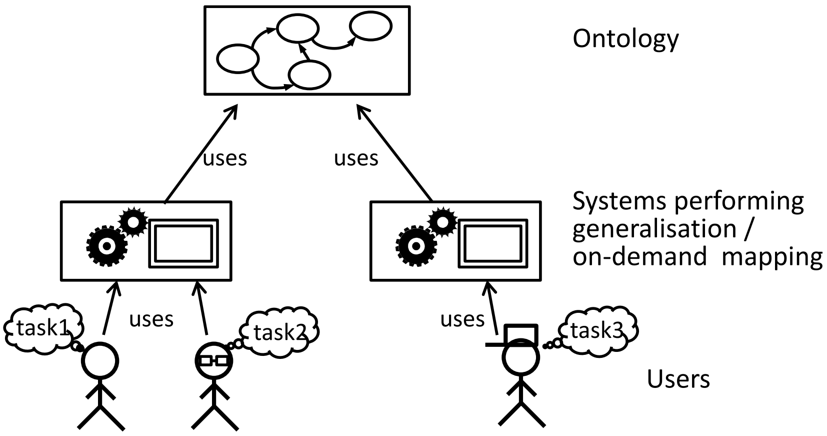

Figure 1: Basis of reflections conducted during the workshop: the ontologies open the way to development of systems able to support production of various thematic maps

Figure 2: Robert Stevens and Sean Bechhofer listen as Guillaume Touya avidly draws.

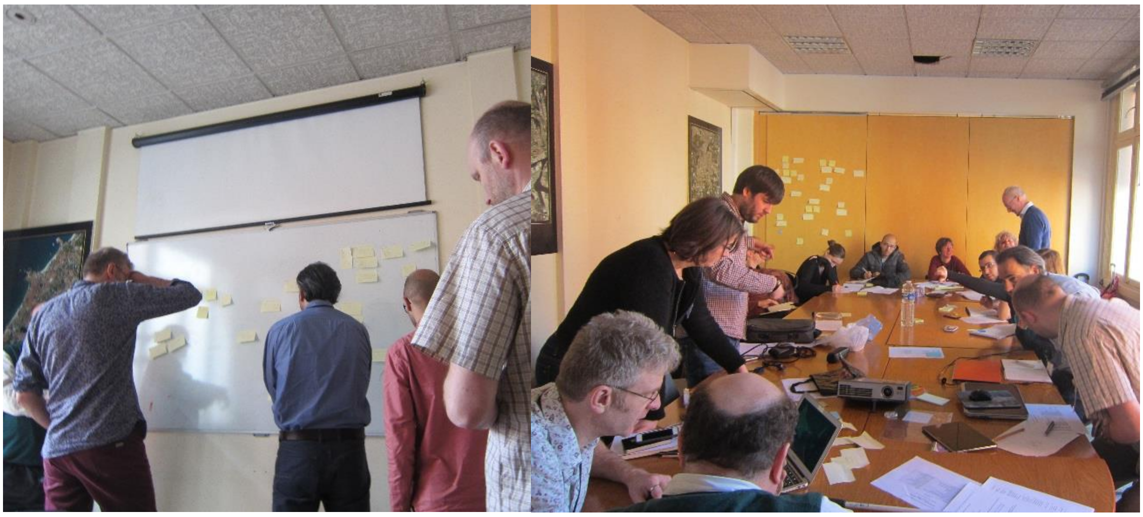

Figure 3: Small and large group discussions

The first task given to the four groups by the ontology experts, in accordance with a common way of exploring the scope and requirements of ontologies, was to imagine a set of personae that would use the generalisation system. The idea being that the ontologies would be used by the system to support the delivery of thematic maps according to the tasks of the users. Tasks usually performed by these personae were also identified. Table 1 shows the personae imagined by the participants.

Once the various personae had been shared, groups were then asked to identify, for each personae, what types of questions the ontology should be able to answer – so called 'competency questions'. Three competency questions associated with each personae were identified. These competency questions have the effect of scoping the types of concepts that need to be modelled within the ontology as well as the relationships that link those concepts (Table 2). These were written on post it notes, and collectively the teams attempted to group questions according to their similarity (Figure 3). It is tempting to want to create a holistic solution but this is a daunting task. By considering particular personae and their competency questions it was possible to make meaningful and pragmatic progress, as well as draw consensus among the participants.

Map Generalization Expert, with specialist knowledge in using map generalization software
Average Joe - Non expert in cartography, wants to create maps on demand of a thematic nature
Tony the Tourist – wants a map of places to visit on his stay in Sydney
Roland the road safety officer, QGIS user, but with little expertise in map making, wanting to view accident hot-spots across the city
The Old Cartographer - traditional manual cartographer working within NMA, no knowledge of automatic map generalization methodologies
A GIS analyst, with limited knowledge of cartographic design, but familiar with postGIS, web geoportals
Penny the GIS facilitator – making different thematic maps for various departments within her organization
Sally the PhD student, expert in spatial analysis, some knowledge of mapping software, developing an interface to interacting with atlas data
Harry the Hydrologist – GIS technician wanting to make a spatio temporal map showing a pollution event
Laura - runs an amateur cyling group and wants a web map for her group
Carla the cook - wants a map of places to find suppliers of particular ingredients

Table 1: Some of the imagined people (personae) who would utilise the generalization system.

Expert cartographer	Which algorithms would best simplify this hydrographic network? Which is the best line simplification algorithm to use?
Sally Student	I visualise my data and see a lot of streets - how do I simplify these networks? What are the most appropriate parameter settings for this building generalization algorithm?
The Old cartographer	Does the solution meet the requirements of the user task? What algorithm do I use to simplify my hydrographic network?
Tony	Can I complete this task in realtime?
Penny	What is the quality of this data – is it fit for purpose?
Roland	How does the map content change with scale?
Tony	How do I get to shark island? What are the nice places to visit around Sydney Harbour?
Penny	What background data are required given the salient data associated with this task? What data are relevant to show flood analysis?
Expert cartographer	How do I convey the constraints associated with this algorithm?
Roland	Can I represent this choice of features (landmarks, accidents, cameras, traffic lights etc) all together on the same map? How do I represent this choice of features at different LoD?
Penny	How can I visualise with clarity whilst ensuring truth (minimum distortion)?
Harry	What are the data relevant to flooding analysis?

Table 2: Example competency questions

The development of these personae and the clusters of competency questions enabled the group to loosely identify a number of ontologies - any of which are in need of development. Not intended to be discrete, there is considerable overlap between the ontologies, and they may be related one to another; the list is illustrative not exhaustive (Table 3).

Tasks	An ontology describing the appropriate map content (foreground, background) according to task
Algorithms	An ontology describing the choice of algorithms relevant for a given set of cartographic constraints
Relationships	An ontology describing the metric and topological properties that exist within and between a set of features on a given map
Metadata/Quality & Truth	An ontology of the links between data quality, truth and clarity – and various mixes of map generalization algorithms
cartographic constraints	An ontology of the constraints that determine the representation of features on a map. Features cannot overlap, minimum building size etc.
Geographical Features	An ontology describing the semantics of the geography that we wish to represent. A river flows into a lake, a road "hosts" a road accident etc.

Table 3: Six possible ontologies that need to be developed

Given the time constraints, it was decided, by voting, to focus on two ontologies for modelling geographical relationships and cartographic constraints, respectively (Table 4 and 5).

Once the competency questions had been reviewed, it was decided to focus on what needs to be modelled geographically speaking in order to support the process of map generalization. For example we know that rivers flow downhill, or that natural features tend to have soft boundaries, and built objects rigid ones. These are geographical properties that we wish to retain during generalization since they govern how things can be generalized, and what qualities we wish to retain. We can distort the shape of a lake, but the boundaries of built structures need to remain rigid and angular. Table 4 is a small set illustrating these ideas, with examples on the right. Simultaneously, another group explored qualities of the map we wish to model in order to govern and assess the quality of map generalization solutions. A subset of such constraints is illustrated in Table 5. The tables are not intended to be mutually exclusive – all ontologies have shared elements. Neither are they intended to be exhaustive; ontologies evolve over time, through discussion and exploration of ideas and concepts.

Constraint & relationship	Example
The generalization of one feature will impact on the generalization of another	If the river is removed then perhaps the bridge does not need to be shown
That shape & distribution affects the choice/ degree of algorithm(s)	A block of houses might be amalgamated, their shape conforming to the shape of the road
Choice of feature classes, their role governed by task and scale	A map of the 10 longest rivers in the world requires continents & labels, seas, river labels
Instantiated features have internal and external properties that depend on existence of other features	A road has curvature, connectivity, network density and angularity that varies between the rural and the urban
A feature can be classified as being a surface, network, or isolated/discrete object	A landscape, river network, buildings
A feature hosts activities	Road accidents happen within the confines of a road network.
A map is an expression of a subset of relationships defined by the presence/ absence of a set of features	Relationships of urban connectivity conveyed in a road atlas. The link between descent (relief) and river energy (its shape)
A feature performs a role in conveying relationships	The continents provide a context to the map showing the longest rivers; building density correlates with road network density
Relationships between features are altered through the generalization of features	Remove railways and stations and you no longer see how cities are connected by rail
The emphasis in conveying relationships is governed by how features are symbolized & generalized	Bertin's variables, iconic/pictorial symbology, model and cartographic generalization all affect the 'strength' of how something is conveyed

Table 4: A mix of map generalization concepts and their relationships (incomplete, illustrative)

Need for map generalization process to operate in real time	Model density changes to retain patterns during changes in LoD
Model positional accuracy to ensure quality control	Model relief for quality control during changes in LoD
Model scale and LoD to ensure continuity in zoom	Model colour choice to ensure comprehension of colour
Model distribution patterns to ensure quality control (that patterns remain)	Model of sequencing of algorithms to ensure efficiency in map generalization
Need to maintain legibility (min. sep., min. size min. width, min. overlap, etc)	Model internal/external properties to reduce loss of information (eg Bld alignment, rigidity)
Model topology to ensure non intersection/ minimize chances in topology & ensure quality control	Model balance between foreground/background to ensure 'balance'
Model network to ensure continuity during pruning & ensure quality control	Model of positioning of labels to minimise overlap
Content models to enable consistency across scales	

Table 5: A non-exhaustive, mix of un-prioritized constraints

After grouping of constraint concepts and relationship concepts, the next step was to develop a set of blob and line relationships, focussing on four personae: *Tony the Tourist*, *Roland the Road Safety Officer*, *The Thematicist* (a conflation of personae interested in thematic mapping) and the *Old Cartographer.* Four groups of 3 to 4 participants were formed, each of them in charge of one persona.

The groups where asked to revisit their competency questions. This was because, having identified concepts and their relationships, it was not so clear as to precisely how these concepts and relationships (Table 4 and 5) would enable the competency questions to be answered. The blob and line diagrams describe the concepts of the ontology and the relationships between them and participants were asked to touch on all aspects of the ontologies (Table 3). It is important to stress the iterative nature of this process and the importance of group discussion in modifying (sometimes radically) the various outputs that came from these group discussions (Figure 4 & 5). Each blob and line diagram reflected the different approaches of each group; the left of Figure 5, for example, has the "algorithm" concept at its centre whereas the right of Figure 5 is centred on the concept of "Accidents". Relationships were either "is a type of" relationship or every other kind of relationship. For example "administrative unit *is a type of* statistical unit", "proportional symbols *is a type of* representation, "Algorithm *has a* property", "scale *governs* representation". Eventually the blobs will map to objects in the ontology and the lines will map to relationships.

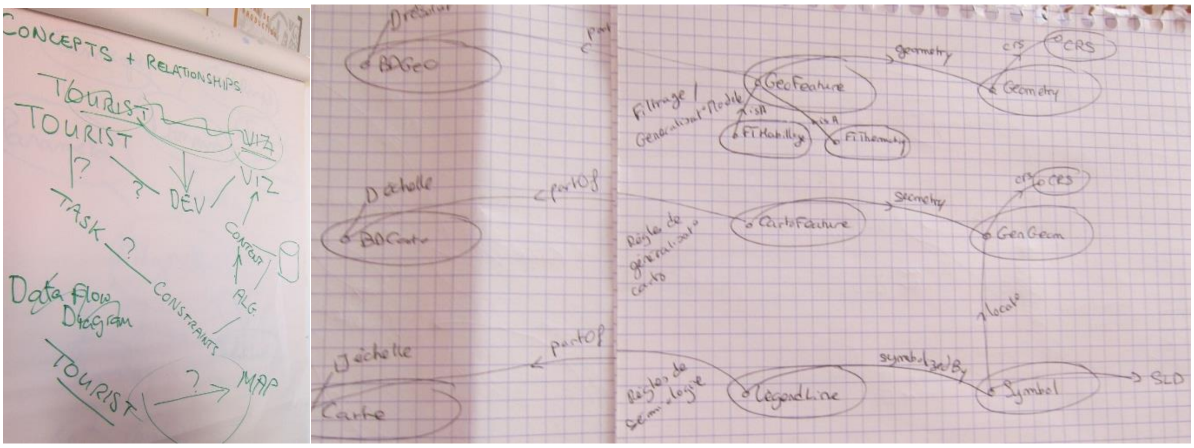

Figure 4: Examples of blob and line diagrams

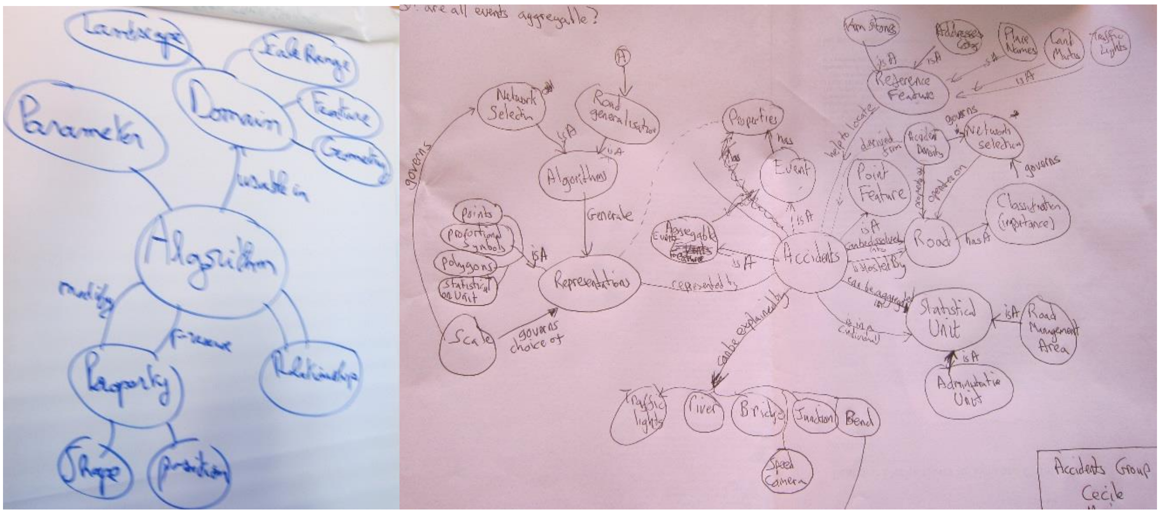

Figure 5: Examples of blob and line diagrams

The group discussions were steered by a need to move from a vertical slice to a horizontal slicing of the problem. In other words, rather than more detailed development of the ontologies, to consider the end to end process – from the personae and their task, through to the selection of data and its generalization – in other words, the life cycle by which a thematic map might be produced in response to a particular request. Some of the discussions were inconclusive or stopped as it proved hard to reach consensus. The discussions continued to focus around contexts of use, and some discussions led to a realization that other topics were more central to ideas in map generalization. For example some of the discussion moved towards ideas of the role of 'scale' and 'algorithms'. In terms of further discussion, it was felt useful to see an example of existing ontologies, so part of the second day was devoted towards a demonstration of a sushi ontology that had been developed by the facilitators in OWL, and how various queries can be posed. At the conclusion of the morning, the ontology experts summarised how they perceived the defining cores of map generalization (Figure 6).

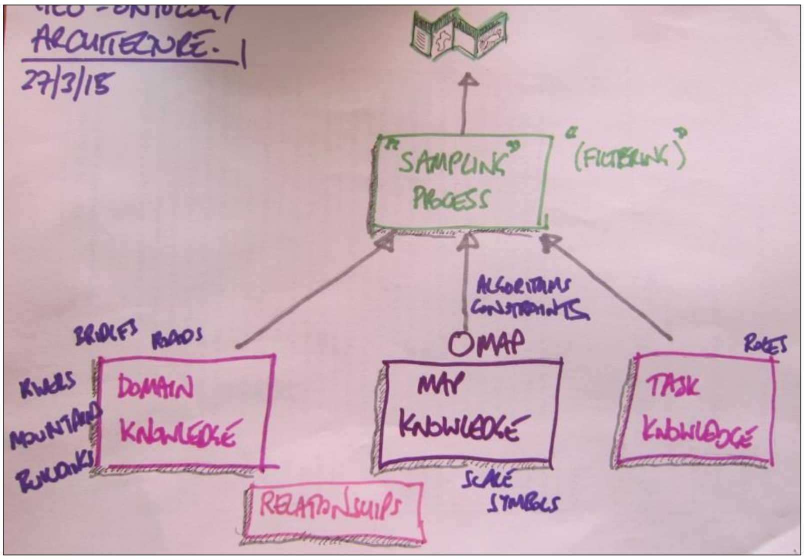

Figure 6: A synthesis of our thinking

In Figure 6, the domain knowledge refers to the real world geographic features we wish to represent on a map. The task knowledge represents the requirements of the end user and the roles the map may be expected to perform: to aid the tourist, to identify accident hotspots, to depict polluted rivers. These two blocks of knowledge go through a sampling or filtering process, guided by the map knowledge, which determines how the features are represented. A useful analogy might be with that of a relational database; the map is equivalent to a particular view of the database, which has been produced by a process of filtering, which has been guided by database queries and the relationships between the tables. This synthesis is close to the seminal proposition of Regnauld (2007).

What next?

A concluding session identified action items; these were to publish a summary of the workshop on the Commission website (the present report), and the creation of a wiki where participants could share their efforts. There was interest in developing ontologies based on a number of different scenarios such as: linking task to content selection, using 'Tony the tourist' as an example, development of an ontology making explicit the link between algorithms and constraints, thematic mapping focused around river pollution and another related to road accidents. There was also interest in creating a topographic map linking OSM data, and a statistical map. For each of these case studies, the ambition is to create a class schema in WebProtégé, an online tool dedicated to collaborative edition of ontologies, such that a number of queries can be posed to the ontology. In combination with database technologies and the OWL interface, can we answer some of the competency questions posed at the outset of the workshop? It will also be important to involve those who expressed an interest in the workshop but were unable to attend. In addition, funding is required for a planned meeting in about six months' time in Paris, with interim meetings planned via Skype. It may be possible to disseminate via a workshop that should be organised together by the ICA Gene & MR Commission and the working group WGII/2 of ISPRS (Multi n-dimensional Spatial Data Representations, Data Structures and Algorithms) in Prague, prior to the ISPRS conference, in early 2016.

Conclusion

There is considerable value in having non domain experts facilitating the meeting. Their presence makes it necessary to make explicit what is often assumed among domain experts. Of particular value was the "Agile" approach adopted by the facilitators; there was considerable, and necessary, deviation from the planned sequence of activities and the facilitators frequently reviewed progress and discussed the next step with the participants.

There is considerable value in examining map generalization from a semantic perspective, and from a user perspective. The construction of an ontology will help clarify the core concepts associated with map generalization and the links between those concepts. Whilst RDMS are good at regular and complete data, ontologies are good at modelling irregular and incomplete data. An ontology is not a panacea. A further challenge will be its integration in the map generalization process. The continued collective effort and involvement of others will be essential to continuation of this project.

Acknowledgement

The organisers are very grateful for a contribution to funding from the ICA and to the IGN for hosting the event.

Appendix A

Participants list

Last name	First name	Affiliation	Email
Abadie	Nathalie	IGN-France, COGIT	nathalie-f.abadie@ign.fr
Augustýn	Radek	Geodesy, Topography and Cartography Research Institute	raugustyn@post.cz
Baella	Blanca	Institut Cartographic i Geologic de Catalunya	blanca.baella@icc.cat
Bechhofer	Sean	University of Manchester	sean.bechhofer@manchester.ac.uk
Burghardt	Dirk	Technische Universität Dresden	dirk.burghardt@tu-dresden.de
Duchêne	Cecile	IGN-France, COGIT	cecile.duchene@ign.fr
Dumont	Marion	IGN-France, COGIT	marion.dumont@ign.fr
Feliachi	Abdelfettah	IGN-France, COGIT	abdelfettah.feliachi@ign.fr
Gould	Nick	Manchester Metropolitan University	n.gould@mmu.ac.uk
Mackaness	William	University of Edinburgh	william.mackaness@ed.ac.uk
Pla	Maria	Institut Cartographic i Geologic de Catalunya	maria.pla@icgc.cat
Regnauld	Nicolas	1Spatial / Star Apic	nicolas.regnauld@1spatial.com
Stevens	Robert	University of Manchester	robert.stevens@manchester.ac.uk
Touya	Guillaume	IGN-France, COGIT	guillaume.touya@ign.fr
Solano	Sergio	Int Geo Inf Lab, Computer research center, Mexico	sergio.solano.p@gmail.com

Footnote

An overview of this paper was presented at a workshop on Ontologies in Leeds as part of the GISRUK conference. Mackaness, W.A. Gould, N, Bechhofer, S. Burghardt, D. Duchene, C., Stevens, and R. Touya, G 2015 Thematic workshop on building an ontology of generalisation for on-demand mapping presented at Workshop of UKON2015 14 April 2015 University of Leeds. www.ukontology.org

References

Gould, N., Mackaness, W., Touya, G. and Hart, G. (2014) Collaboration on an Ontology for Generalisation. In: 17th ICA Workshop on Generalisation and Multiple Representation, Vienna, Austria.

Regnauld, N., 2007. Evolving from automating existing map production systems to producing maps on demand automatically. In: 10th ICA Workshop on Generalisation and Multiple Representation, Moscow, Russia.

Gould, N. and Mackaness, W.A. (in press). From taxonomies to ontologies: formalizing generalization knowledge for on-demand mapping. Cartography and GIScience.