CO-091

MODELS AND STANDARDS FOR ON-DEMAND MAPPING

BALLEY S., REGNAULD N.
Ordnance Survey, SOUTHAMPTON, UNITED KINGDOM

ABSTRACT

A large part of the current research activity in GIS looks towards on-demand mapping. The problem encompasses two main dimensions: understanding the user's needs and generating a product on-the-fly. Ordnance Survey research department is also investigating the delivery of on-demand products, with a high level of customisation and including user's data. Could available models be brought together into a complete, flexible on-demand mapping system? This paper proposes an overview on the research on on-demand mapping. It proposes a high-level architecture for an on-demand mapping system and analyses where and how it could be fed by existing models.

1. CONTEXT AND GOALS

The need for customised data products

The general public composes more and more maps on the Web, overlaying data from different sources, possibly their own data. Some National Mapping Agencies (NMAs) are now supporting the trend by providing their general purpose maps for free together with APIs enabling to integrate them into mash-ups. Such manipulation is tricky and the quality of resulting maps is a frequently raised issue. There is a need and a duty for NMA to deliver high-quality data adapted to the intended use, notably background data whose content and style would fit the user's data.

The need for customised data is not limited to the general public and map making. To set up analysis applications based on vector data, expert users may need to customise the content but also the structure of the data (e.g. to reorganise data into Feature Types with agreed names, attributes and relationships). Some applications may also have more advanced integration needs (e.g. to insert user's features as new nodes of an existing network and to reorganise this network accordingly). Even if more and more tools are available for users to perform such transformations by themselves, their use is not straightforward and may even damage data. An upstream customisation performed by data providers is a better option, indeed recommended by INSPIRE as "transformation web services to achieve interoperability". This alternative presents several other advantages, such as maintaining consistent metadata and tracing modifications for future updates.

Context at Ordnance Survey

Ordnance Survey has built its national, high resolution, seamless database. This database will be used to populate a multiple representation digital landscape model (MR-DLM) providing geographic components for current and future products. Initially, this new set up will improve the production lines of existing products, and ease the creation of new products. In the meantime, Ordnance Survey's Research department is investigating how this new set up could enable in the future a more flexible way of deriving products, where customers get products tailored to their needs (Regnauld, 2007).

The on-demand mapping process

Figure 1 represents the main steps of what we call an on-demand mapping system. These steps are necessary to derive, manually or automatically, a customised data product (either a map or a dataset).

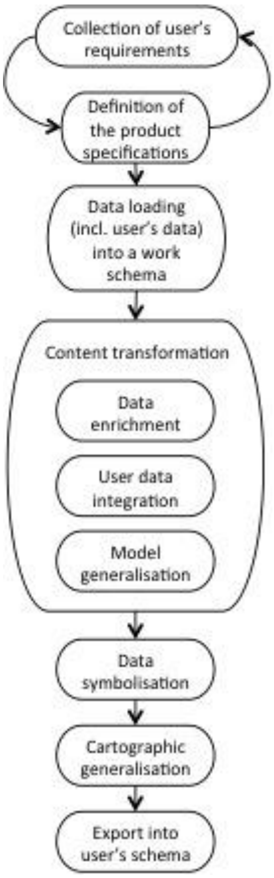

Figure 1: Main steps of the on-demand mapping process

The two first steps are iterative: the data provider collects the user's requirements, including their integration needs, and progressively builds the target product specifications. Then, relevant data components from the MR-DLM are loaded together with the user's data into some work schema. The content transformation step includes, in a variable order, operations performing the actual integration of user's data, data enrichment and model generalisation. If the target product is a map, the data is symbolised and a cartographic generalisation step improves the visual quality of the output. If the target product is a dataset, the data is restructured and exported according to the specification. Each step may require the user to precise their preferences or to validate the data provider's choice.

Our goals

The long-term goal of this project is to design a prototype for an online on-demand mapping system available to a wide range of users.

During a session the system would:

1.Help the user specify their requirements and describe their own data if they wish to integrate some,

2. Translate user's requirements into formal, machine-readable product specifications,

3. Design a derivation workflow.

4. Run the workflow by chaining available tools published as web services

As each step listed above constitutes a research field on its own, designing the overall application is not within the reach of an isolated team. Consequently, our **medium-term goal** is to set up the global modelling framework that will guide future contributions and enable them to interoperate.

The purpose of this paper is to provide an overview of on-demand mapping research (section 2) and to show how they could feed the high-level architecture we have in mind (section 3).

2. THE RESEARCH ON ON-DEMAND MAPPING

Many current research studies in the field of GIS relate more or less directly to on-demand mapping. The problem encompasses two main processes: understanding the user's need and generating a product on-the-fly. Each process can be decomposed along the dimensions of the target product: its thematic content, its spatial level of detail, its legend and its structure. Section 2.1 reviews research projects trying to accomplish the overall process. Section 2.2 reviews research domains providing more in-depth solutions to specific components of the process.

2.1 On-demand mapping systems

The central resource of these systems is the formal map specification created interactively with the support of the system knowledge.

A complete expert mapping system was specified in (Forrest, 1999). The targeted products are small scale thematic maps, for which the system must supply a topographic base layer. The initial inputs from the user are a map topic selected from a hierarchy (e.g. geology, industries, relief, etc.), a map purpose (detailed study or general overview), a user profile (naive or expert) and an output medium. Then, the map content (level of detail, themes and classes) is specified collaboratively by the user and the system which, using quantitative indexes and scores, guards against information overload. In this expert system, all rules and scores are defined by the expert cartographer.

The UE funded GiMoDig project (2001-2004) implemented the first on-demand mapping prototype (Sarjakoski and Sarjakoski, 2007). The general goal was to improve the accessibility and interoperability of national topographic databases in a mobile context. The key techniques were data integration and real-time generalisation. Custom map specifications are built from context parameters collected from the user (describing the device, map and user) and an internal knowledge base. These specifications enable the system to build requests to a generalisation service and web feature services.

The COGIT laboratory of IGN France specified a series of web services, partly implemented, to provide on-demand maps based on user's specifications (Bucher et al., 2007): a map specification service, a legend definition service and a legend evaluation service. The formal map model, mostly based on the OGC MapContext standard, allows to specify a communication level for each theme (background, to be read, first sight) and inter-theme relationships (order, association and differences). The first service helps the user defining some of the abstract properties of their map, independently from the available data and styles, and automatically instantiates other parts of the specifications. The two other services make use of the laboratory's large knowledge base about legends (Buard and Ruas, 2009) (Chesneau, 2007) (Christophe and Ruas, 2009) to propose adequate symbolisation and improve it if needed.

The DURP Ondergronden project (Poppe et al., 2006) encouraged Dutch local governments to publish thematic spatial data. It aims at generalising base maps as a context for these thematic data according to user requirements. The target map specifications are composed of:

- A selection of background and foreground map layers,
- Information for the symbolisation and level of detail of the map background adapted to the thematic data,
- Thematic features which need to keep topologically consistent with base map features
- Thematic features which need to keep a high level of detail,

These specifications were formalised in the distributed architecture for on-demand mapping proposed by (Foerster, 2010), through a Map Context document, a topological awareness list and a generalisation matrix linking base data and thematic data. The specifications are sent to a generalisation-enabled Web Map Service, which is the core element of the architecture.

These four global approaches demonstrate the feasibility of building map specifications interactively using a knowledge base, and of using these specifications in a mapping process. They also reflect the ever growing interest in standard models and modular applications. However, they only involve predefined data sources and predefined derivation tools whose parameters are hard-coded in the map specification. The focused approaches described in the next section enable to decouple the map specifications from the data and mapping process.

2.2 Research domains contributing to on-demand mapping

This section reviews other research areas contributing to on-demand mapping. Section 3 specifically identifies relevant models for each component of our planned architecture.

<u>Web services and service chaining</u>

The generalisation community has been interested for a while in publishing generalisation operators as web services, initially to promote collaboration between teams (Edwardes et al., 2003). The first implementation of generalisation web services was proposed in (Burghardt et al., 2005) on the WebGen platform. (Foerster and Stoter, 2006) proposed a WPS profile for such services. A working group of the ICA commission on generalisation and multiple representation has proposed a central registry of generalisation web services and released a dedicated client, WebGen2.0, based on the OpenJump platform (Foerster et al., 2008).

Like generalisation operators, schema transformations operators could be published as web services. (Foerster et al., 2010) have initiated a classification for unifying schema transformation and generalisation operators.

Although service chaining is often done manually, studies to automate it have already been conducted, for example by (Burghardt and Neun, 2006) for chaining generalisation services, or by (Letho, 2007) to automatically build scripts to perform schema transformations between already matched schemas.

In the meantime, research about service interoperability has progressed and provides helpful methodologies for web services chaining (Klien et al., 2006) (Lemmens, 2006). They rely on the semantic description of the functionality offered by the service and of its inputs/outputs. The required concepts are operations (e.g. simplification), real world concepts (e.g. road) and GIS constructs (e.g. a curve, a GM_Curve primitive). These concepts are stored in distinct ontologies (Lemmens, 2006).

Semantic data modelling and data integration

A central requirement for on-demand mapping is the possibility to dynamically select the most relevant data source for a user need. The first step is the semantic referencing (Kuhn, 2003) of data using an ontology of real world concepts. Several ontologies have been built at Ordnance Survey, which now enable a few datasets to be accessed on the web as LinkedData (Dolbear and Goodwin, 2007). However, to assess precisely the suitability of a data source, it is not only necessary to identify the represented concept (e.g. road), but also to describe how it is represented: which roads have been selected (e.g. only the public roads) what part of the road has been captured (e.g. its centreline), which properties have been observed (e.g. its classification and width) and how each properties have been encoded (which attribute name, which values domain?). This information constitutes the data specifications and is usually held in textual documents by data producers. (Gesbert, 2005) proposed a formal specifications model, which IGN-France has used to formalise the specifications of its topographic databases using OWL ontologies (Abadie et al., 2010).

Such an advanced semantic modeling is necessary not only to identify the most relevant dataset representing a given real world concept, but also to use derivation services correctly: a rich semantic modeling of data is required to match with the rich semantic modeling of the inputs/outputs of derivation operations. For example, if a network filtering operation requires the "width" of a linear sections, one must identify whether or not this property is represented in the data.

Semantic modeling plays a central role in data integration: ontologies (Fonseca et al., 2003) and formalised data specifications (Sheeren et al., 2009) (Abadie, 2009) have proved very useful to achieve data and schema matching. However, except from a few attempts (Grosso, 2009), there is no user-oriented service exploiting the above research results and enabling an easy integration of user data into a referential. This will probably change quickly as more and more spatial information is generated by users and could enrich data referentials (Brando and Bucher, 2010).

Product adaptation through user interfaces and dialogue

Some research studies aim to allow non expert users to customise one particular dimension of a product through advanced interactions. It reduces the impact of predefined user profiles, and brings much more flexibility to the user, still strongly guided by the system knowledge.

(Balley, 2007) focuses on the adaptation of the data structure. The user works at the conceptual schema level, by manipulating a class diagram on a graphical interface. They can apply atomic schema transformations (e.g. split a class), or more complex transformations suggested by the data producer to derive new meaningful classes.

(Hubert, 2003) proposes a dialogue application, based on map samples and natural language, to let non-expert users express their mapping requirements concerning the representation of buildings on a topographic map. The user reacts to map samples representing differently generalised buildings. His reactions ("too big", "too square") are interpreted in a space of generalisation parameters. Based on these reactions, the engine narrows his proposals until a satisfactory solution is found.

(Christophe and Ruas, 2009) proposes a dialogue application to design efficient and harmonious legends respecting user's tastes. The dialogue is based on inspiration choices: map samples or painter's palettes. The user can react on colours in general ("I don't like this red") or for a theme ("I like this red for the roads"). The system progressively adapts its proposals thanks to a database of predefined, characterised map samples. The end result is a legend that pleases the user, and still comply with semiology rules.

This section showed that many aspects of on-demand mapping have already been studied in separate projects. The challenge is to define a framework that allows these to work together. Section 3 presents a global architecture and links its components to models coming from these research domains and/or from standardisation initiatives.

3. PROPOSED ARCHITECTURE AND POSSIBLE INPUTS

Figure 2 represents our proposed high-level architecture for an on-demand mapping system. It is meant to be very modular and uses a semantic referential ensuring the interoperability of components.

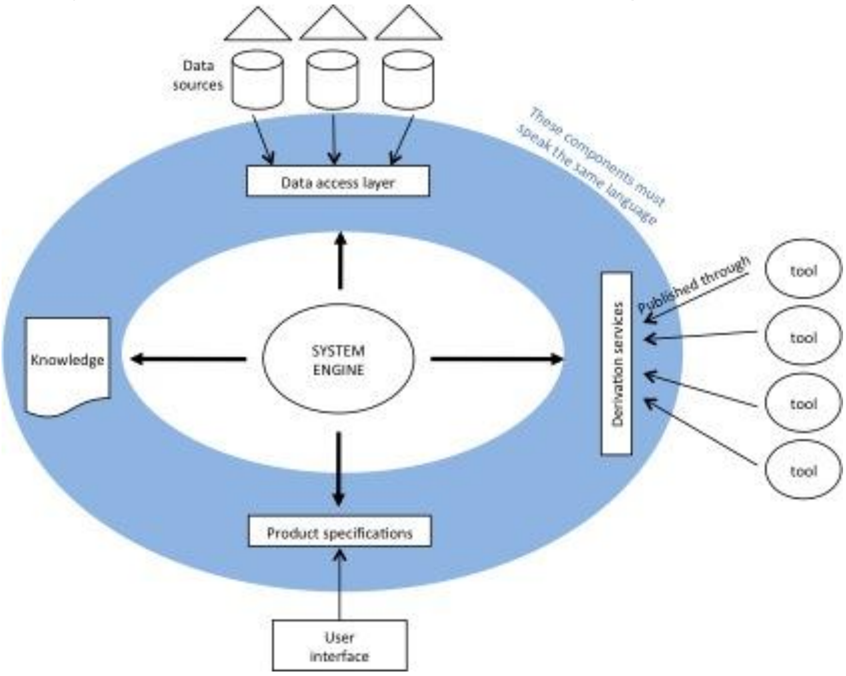

Figure 2: High-level architecture deriving on-demand products

The user interface collects user's requirements with an approach adapted to the user's level of expertise. It should spare the user the description of the needed product, but rather let them express somehow what they want to do, the data they want to integrate (if any), and their preferences. Different techniques will be explored, including user-system dialogue and reasoning on product examples.

The product specifications are a machine-readable translation of the user's requirements in terms of what the final product should be, including the integrated user's data. The main drawback of ad-hoc specification models of existing on-demand mapping projects is their dependence on data and processes. Standard models such as OGC MapContext and ISO 19131, dedicated to the description of existing products, are strongly coupled with data as well. Our specification model should not describe an existing but a target product. We propose to make it based on represented real world concepts and on representation constraints stating how these concepts should be represented in the product. These constraints differ from the cartographic constraints guiding the generalisation processes, considered as system knowledge. We have identified 4 types of constraints dealing with the selection of entities, the semantic and geometric modelling, the symbolisation and the data structure. For some of these constraints we can borrow elements from the specification models linking databases schemas to real world concepts (Gesbert, 2005).

The data access component enables the system to identify the most relevant data sources and to extract the data feeding the workflow. Data are accessed at a semantic level thanks to an ontology mapped to the data schemas through a model to be defined. It should be fully consistent with the model used for representing constraints, enabling target product specifications to be easily matched with the available data sources.

The derivation services make existing derivation tools available for the system to build the workflow. The intention is to build a library of generalisation tools, packages as OGC Web Processing Services. In order to be correctly chosen and chained, these tools will need to be described in terms of realised operations and manipulated concepts, for example with Semantic Annotated WSDL.

The system engine is responsible for building a derivation workflow, relying on the product specifications and knowledge, and running it. To stay independent from available tools, the system will at first reason on abstract operations. OWL-S is a candidate model to formalise these operations. The system will also need to perform dynamic service chaining, requiring an interoperability framework as described in section 2.2. What operations to chain will be mostly determined by the system knowledge, but interesting chaining could be found through machine learning and optimisation techniques.

The knowledge component represents all knowledge required by the system to do some reasoning. We have identified three types of knowledge.

- The product design knowledge guides the creation of the product specifications. It associates each user requirement (for instance a product type: map, an activity: hiking and a scale range) with some real world concepts and representation constraints. The Ordnance Survey's GeoUsers research group have acquired some of this knowledge through usability studies (Harding et al., 2009). Other knowledge sources are the Inspire data specifications produced by experts. Knowledge about the suitability of legends for different types of spaces or for different users state of minds have been formalised by expert cartographers (Christophe and Ruas, 2009).

- The cartographic knowledge provides guidelines for the derived product. They typically include cartographic constraints guiding the generalisation process,. The generalisation community has widely studied the content and formalisation of cartographic constraints (Burghardt et al., 2007).

- The procedural knowledge helps the system engine to build the workflow. It can have the form of rules, of a global derivation plan, or be enclosed in the description of complex processes composed of atomic operations.

We intend to model and encode this knowledge in a platform-independent way. In their attempt to orchestrate several generalisation processes, (Touya et al., 2010) propose a platform-independent model for both cartographic constraints and procedural rules. In terms of the format of the knowledge base, we are exploring the XML-based Rule Interchange Format (RIF), a W3C standard which could simplify the reuse of knowledge between components.

The semantic referential (blue circle): As resources such as data, derivation operations, constraints and rules will need to be acknowledged by different components of the architecture, a key requirement of our global model is to support the semantic description of these resources. Every geographic or GIS concept used by a product specifications, an operation or a piece of knowledge must be defined in an ontology. Several proposals have been described in section 2.2. (Touya et al., 2010) are building such a referential specifically for the generalisation process. A key point in our approach will be to build our referential in collaboration with the teams likely to contribute to one or the other component of the architecture.

4. FUTURE WORKS

This paper has presented our on-demand mapping objectives and situated our approach in the related research domains. A high level architecture has been proposed and will be specified progressively, either during targeted research projects initiated at Ordnance Survey, or taking input from research partners pursuing research on some components of the system.

Meanwhile, we have defined two use cases aiming at validating it globally. In the first use case, the target product is a route planning map around the scale of 1:100K. User's data about cycling itineraries must be integrated and made geometrically consistent with the referential road network. The map must display landmarks selected by the users: special and isolated buildings and river crossings. This use case involves sequencing issues: the matching between user's itineraries and referential road network must be carried out at the beginning of the workflow as it enables to filter the road network; the road network must be processed first as its final state is required to filter rivers (as river crossing are seen as landmarks).

In the second use case, the target product is a vector dataset dedicated to route processing. The user wishes each type of functional site (service areas, train stations, airports, etc.) to be represented in its own feature types, which requires a schema transformation operation. User's data about car park areas have to be fully integrated: their access nodes have to be identified or created and inserted into the road network.

In the next year, we intend to build a demonstrator illustrating the complete process on the first use case. Another task starting shortly is the progressive creation of our semantic referential involving partners.

REFERENCES

Abadie N. (2009) "Schema Matching Based on Attribute Values and Background Ontology". 12th AGILE Conference on Geographic Information Science, Hannover.

Abadie, N., A. Mechouche, and S. Mustière. (2010) "OWL-based formalisation of geographic databases specifications". 17th International Conference on Knowledge Engineering and Knowledge Management, Lisbon.

Balley, S. (2007) "Aide a la restructuration de donnees geographiques sur le Web - Vers la diffusion a la carte d'information geographique." PhD Thesis, Université Paris Est - Marne la Vallee.

Brando C., and B. Bucher. (2010) "Quality in User Generated Spatial Content: A matter of specifications". 13th AGILE Conference on Geographic Information Science, Guimarães.

Buard, E., and A. Ruas. (2009) "Processes for improving the colours of topographic maps in the context of map-on-demand." 11th International Cartographic Conference. Santiago.

Bucher, B., E. Buard, L. Jolivet and A.Ruas. (2007) “The need for Web legend services.”. 7th International Symposium on Web and Wireless GIS, Cardiff.

Burghardt, D., M. Neun, and R. Weibel. (2005) “Generalization Services on the Web—Classification and an Initial Prototype Implementation.” Cartography and Geographic Information Science 32(4). 257- 268.

Burghardt, D., and M. Neun. (2006) “Automated Sequencing of Generalisation Services Based on Collaborative Filtering.” 4th GIScience Conference on Geographic Information Science, Munster.

Burghardt, D., S. Schmid, and J. Stoter. (2007) “Investigations on cartographic constraint formalisation”. Workshop of the ICA Commission on Generalisation and Multiple Representation. Moscow.

Chesneau, E. (2007) "Improvement of Colour Contrasts in Maps : Application to Risk Maps." 10th AGILE Conference on Geographic Information Science. Aalborg.

Christophe, S., and A. Ruas. (2009) "A process to design creative legend on-demand." 11th International Cartographic Conference. Santiago.

Dolbear, C., and J. Goodwin. (2007) “Expressing Relational Data as RDF.” W3C Workshop on RDF Access to Relational Databases.

Edwardes, A., Burghardt, D., Bobzien, M., Harrie, L., Reichenbacher, T., Sester, M., et al. (2003) "Map generalization technology: Addressing the need for a common research platform." 21st International Cartographic Conference. Durban.

Foerster, T. and J.E. Stoter. (2006) Establishing an OGC web processing service for generalization processes. ICA workshop on generalization and multiple representation, Portland.

Foerster, T., D. Burghardt, M. Neun, N. Regnauld, J. Swan, and R. Weibel. (2008) “Towards an Interoperable Web Generalisation Services Framework – Current Work in Progress.” 11th Workshop of the ICA Commission on Map Generalisation and Multiple Representation. Montpellier.

Foerster, T. (2010) “Web-based architecture for on-demand maps - Integrating meaningful generalisation processing.” PhD Thesis, University of Twente.

Foerster, T. , Lehto, L., Sarjakoski, T., Sarjakoski, L.T. and Stoter, J. (2010) “Map generalization and schema transformation of geospatial data combined in a Web Service context”. Computers Environment and Urban Systems 34: 79–88

Fonseca, F., D. Clodoveu and G. Camara. (2003) "Bridging Ontologies and Conceptual Schemas in Geographic Information Integration." GeoInformatica 7(4): 355-378.

Forrest, D. (1999) "Developing Rules for Map Design: A Functional Specificationfor a Cartographic-Design Expert System." Cartographica 36(3): 31-52.

Gesbert, N. (2005) "Étude de la formalisation des spécifications de bases de données géographiques en vue de leur intégration." PhD Thesis, Université de Marne la Vallée.

Grosso, E. (2009) “Data Integration GeoService: A First Proposed Approach Using Historical Geographic Data.” 9th International Symposium on Web and Wireless GIS, Maynooth.

Harding, J., S. Sharples, M. Haklay, G. Burnett, Y. Dadashi, D. Forrest, M. Maguire, C.J. Parker, and L. Ratcliffe. (2009) "Usable geographic information – what does it mean to users?" AGI Geocommunity.

Hubert, F. (2003) “Modèle de traduction des besoins d'un utilisateur pour la dérivation de données géographiques et leur symbolisation par le Web.” PhD Thesis, Université de Caen.

Klien, E., M. Lutz, and W. Kuhn. (2006) “Ontology-based Discovery of Geographic Information Services – An Application in Disaster Management.” Computers, Environment and Urban Systems 30(1): 102-123.

Kuhn, W. (2003) “Semantic Reference Systems”. International Journal of Geographical Information Science, Guest Editorial, 17(3): 405-409

Lemmens, R. (2006) "Semantic interoperability of distributed geo-services." PhD Thesis, University of Twente.

Lehto, L. (2007) “Schema Translations in a Web Service Based SDI.” 10th AGILE Conference on Geographic Information Science, Aalborg.

Poppe, E., C.P.J.M. van Elzakker, and J. Stoter. (2006) “User-centred design for application-driven generalisation of base maps.”. 25th Urban Data Management Symposium. Aalborg.

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

Sarjakoski, T., and L.T. Sarjakoski. (2007) "A Real-Time Generalisation and Map Adaptation Approach for Location-Based Services." in Generalisation of Geographic Information: Cartographic Modelling and Applications, edited by William Mackannes, Anne Ruas and L. Tiina Sarjakoski, 137-159. Elsevier - In behalf of the International Cartographic Association.

Sheeren, D., S. Mustière, and J.-D Zucker. (2009) “A data-mining approach for assessing consistency between multiple representations in spatial databases.” International Journal of Geographical Information Science 23(8): 961–992.

Touya, G., C. Duchêne, and A. Ruas. (2010) "Collaborative Generalisation: Formalisation of Generalisation Knowledge to Orchestrate Different Cartographic Generalisation Processes." 6th GIScience Conference on Geographic Information Science, Zurich.