Towards a Prototype for Deriving Custom Maps from Multisource Data

S. Balley[1], K. Jaara[2], N. Regnauld[1]

[1]Ordnance Survey, Adanac Drive, Southampton, SO16 0AS, United Kingdom
Sandrine.Balley@ordnancesurvey.co.uk
Nicolas.Regnauld@ordnancesurvey.co.uk

[2]IGN, COGIT, 73 avenue de Paris, 94160 Saint-Mandé France ; Université Paris-Est
kusay.jaara@ign.fr

1. Introduction

Despite more and more flexible mapping services being freely available on the web[1], making a good quality map combining user's data with a referential background remains a real challenge for non cartographers. The research department of Ordnance Survey (OSGB) is currently investigating on-demand mapping, i.e. how to automatically derive custom maps driven by user's requirements. The goal is to select just the needed amount of information, to transform it into a custom map background and to carry out the task of user data integration. All the process must be planned, parameterised and executed without any human intervention. From this new way of deriving maps, users would benefit from more usable products and OSGB would benefit from exposing value-added data at a moderate cost.

During the first phase of the project, we analysed the requirements of an on-demand mapping system in terms of models and formalised knowledge (Balley and Regnauld 2011a). Aware of the complexity of the topic, we sketched a high-level, distributed architecture allowing collaboration with other research teams on its different parts (Balley and Regnauld 2011b). We notably decided to share a semantic referential (Kuhn 2003) composed of three ontologies: an ontology of geographic concepts (e.g., road network), an ontology of GIS concepts (e.g., line, stroke, feature) and an ontology of operations (e.g. filtering, computing spatial relation). These ontologies are growing progressively. We designed a product specifications model and planning mechanisms to automatically create abstract derivation plans.

The current phase of the project is a proof of concept. We are validating our models in a prototype. We are also trying to figure out what output quality can be achieved from such a system, using available derivation tools and a reasonable processing time (we are not trying to offer proper "on-the-fly").

The second section of this paper focuses on models. Section 3 describes current experimentations: the use case we are focusing on and the developments being carried out. Section 4 describes an attempt to interoperate with an external system proposed by IGN France.

2. The models

Section 2.1 presents the specifications model describing the target map. In section 2.2, specifications are set in the context of a complete on-demand mapping session. Section 2.3 explains the planning mechanisms and the underlying models of plans and tasks.

---

[1] http://geocommons.com/, http://www.google.com/mapmaker

2.1 The specifications model

The specifications model enables to formally describe how the final map should be. As shown in following sections, these formal map specifications guide every reasoning step of an on-demand mapping session. The translation of user's needs and preferences into map specifications is left for next stages of the project: specifications are currently manually generated by experts. The specifications do not mention any existing data sources, derivation tools or algorithms. This resource independence guaranties that the specifications can be handled by the system even if the resources (data and tools) evolve.

Several parts of our specifications model were inspired by the model used at COGIT laboratory to describe and integrate existing databases (Gesbert 2005). A simplified view of the model is displayed in Figure 1. Our central class, mapped concept, refers to a real world geographic concept that should be involved in the final product, either as represented features or as a support to represent another mapped concept. Real world geographic concepts are defined in an ontology. The model enables to specify the content (themes, selection of entities), modelling (geometric primitives, aggregation level, represented thematic properties, etc.), symbolisation and data schema of the target product. It also enables to describe the integration need, i.e. the user's data and how they should relate to the referential. These 5 categories of requirements are called *representation constraints* on the figure. Each one has its own model which is not detailed here. The model notably enables to express spatial relations between two mapped concepts. Spatial relations can appear in any representation constraint, as examples appearing in purple on figure 1 show.

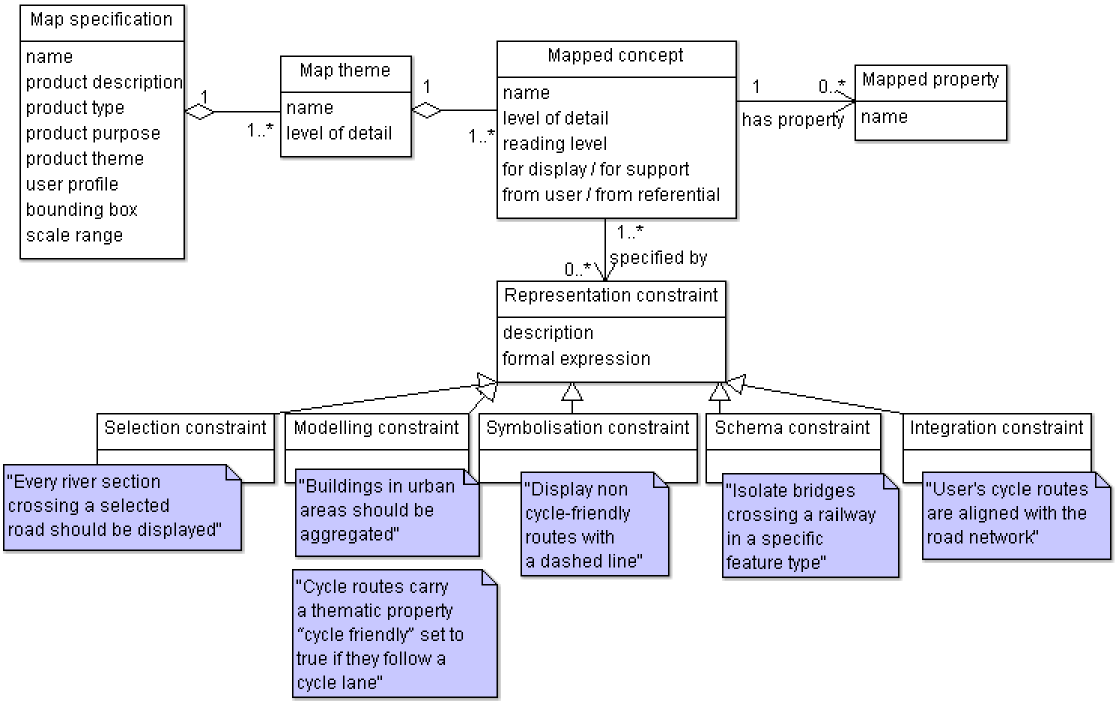


Figure 1: simplified view of the specifications model with examples of representation constraints.

2.2 Main steps of a custom mapping session

The sequence diagram on Figure 2 presents the components of the system and their interactions during a session.

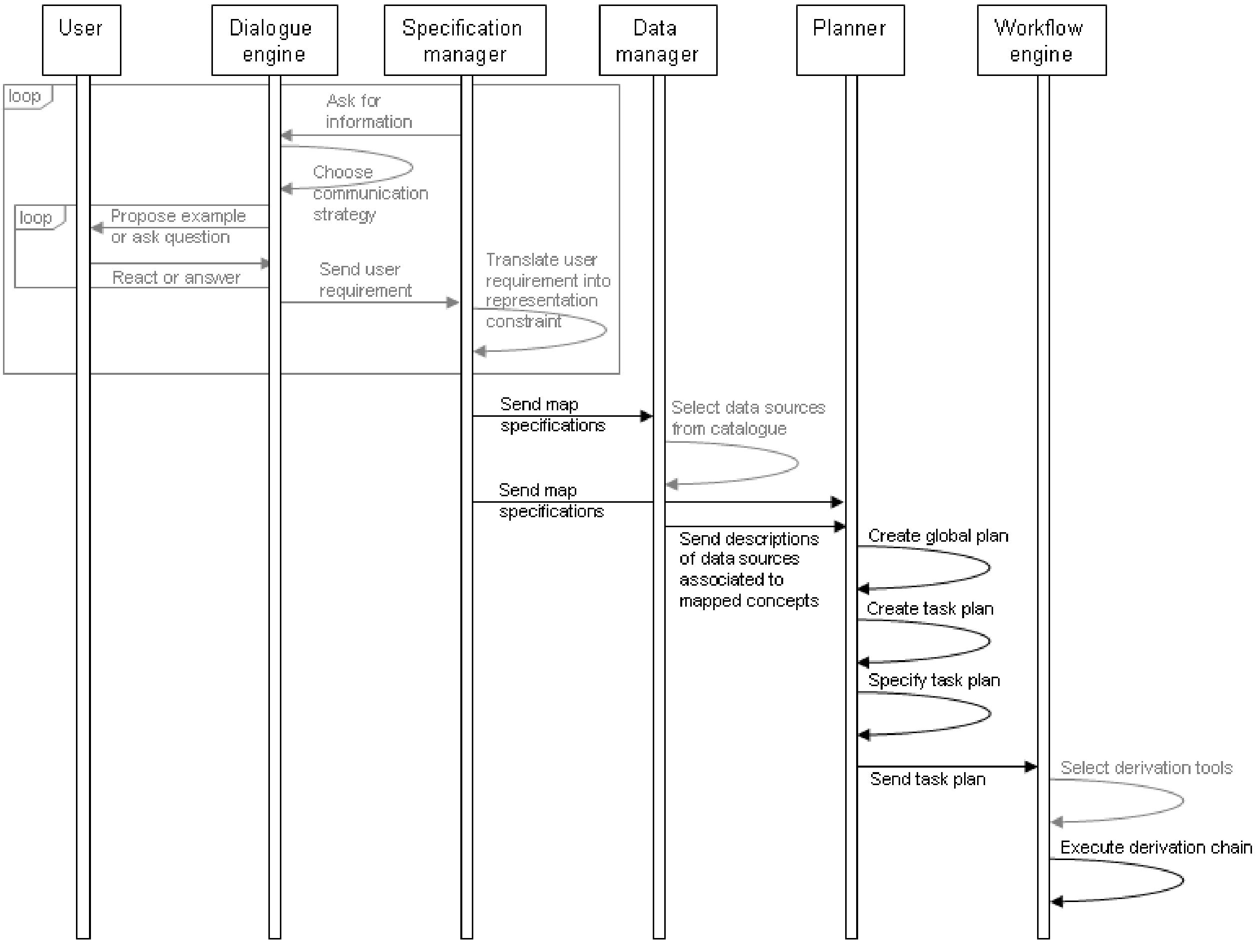


Figure 2: Sequence diagram of an on-demand mapping session. Grey parts are omitted or replaced by shortcuts in our current experiments.

The specifications manager reads and writes specifications. To write new specifications, it must be plugged to a dialogue engine collecting user's preferences, and translate preferences into data requirements.

The data manager has access to a catalogue of data sources. Its role is to select, for each mapped concept, the most relevant available data source. This selection is based on reasoning over geographic concepts from the semantic referential, over representation constraints from the specifications, and over representation rules stated in the data sources catalogue. An example of rule is given below.

Rule:
If a mapped concept is required at a low level of detail in the target map,
and this mapped concept has a specified selection constraint based on a relational constraint with an other mapped concept,
Then select a data source with a high level of detail.

Consequence in our use case (see section 3.1):
As every minor road followed by a user's cycle route must appear on the map, the data manager selects a reference road network with a higher resolution than required by the target map.

The planner generates a derivation plan using templates and rules, as explained in section 2.3. The generated plan is independent of any implemented tools.

The workflow engine executes the derivation plan by choosing the most relevant tools and invoking them with the right parameters.

2.3 The planning process

Designing a map derivation process has similarities with the composition of services, where services perform data selection, generalisation, symbolisation and restructuration (Lemmens 2006) (Lutz 2007) (Andrei et al. 2008) (Friis Christensen et al. 2009). Like in the semantic services chains of Lemmens (2006), we distinguish between abstract composition, where the steps of the process are designed, and concrete composition, where the actual services to invoke are identified. Our planning process is doing abstract composition only. We will be using shortcuts when it comes to identify and invoke available derivation tools. The studies of Gould et al. (2012), collaborating on the project, will propose more complete solutions for concrete composition.

Abstract composition is done in two steps. The first step consists in defining what to do at a general level, and in what order. It produces the *global plan*. The second step consists in decomposing and specifying each step of the global plan. It produces the *task plan*. These plans are presented in the following.

Global plan

A global plan is a sequence of goals. A goal is a high-level objective, like “filter road network based on a spatial relation”, “cartographically generalise road network” or “align user itineraries on the road network”. The system has a set of plan templates associated with well known geographic concepts: roads, buildings, itineraries, etc. Each template consists in a pre-defined list of goals. However, depending on the context, some goals are not relevant for the map targeted. Using templates enables to encapsulate the producer’s knowledge without writing too much rules.

Planning is carried out as follows. Firstly, the planner considers each mapped concept from the specifications and, reasoning on the ontology of geographic concepts, tries to associate it with a plan template. Then, the inference engine uses planning rules to validate only the relevant goals of each plan template. The facts analysed by these rules are representation constraints from the specifications and descriptions of data sources selected during the previous step by the data manager.

Rule:
If a mapped concept MC has a thematic modelling constraint regarding a property P,
and this modelling constraint is specified by a property constraint on P (not a relational constraint with an other mapped concept),
and MC is associated to a data source representing P,
then validate the goal “reclassify attribute values” in the global plan associated to MC

Consequence in our use case (see section 3.1):
As roads are characterised by 3 levels of importance in the target map, and importance is described by an attribute in the selected data source, the plan for roads includes a “reclassify attribute value” goal.

Every goal validated by the inference engine is kept in the final global plan. Goals must then be reorganised to be reached in a correct order. Sometimes, achieving a goal first is a prerequisite for some other goals to succeed. Sometimes, the order between two goals is only a matter of optimisation. Although we only use binary sequencing rules, it is possible at this stage to create “soft” and “hard” order relations. Examples of sequencing rules are provided below.

Rule:
If a mapped concept MC1 has to be filtered (guided or not by a spatial relation),
and MC1 is moreover involved in a spatial relation with another mapped concept MC2,
then the spatial relation should *absolutely* be assessed before MC1 is filtered.

Consequence in our use case (see section 3.1):
As a high resolution road network has been selected, it must be filtered (guided by the user’s cycle routes). However, the road network is involved in a relational constraint with rivers: every river section crossing a road must be represented. As a consequence, road/river intersections must be computed before roads are filtered, and this order relation has a high weight.

Rule:
If a mapped concept MC1 has to be skeletonised,
and MC1 is involved as a surface in a spatial relation with another mapped concept MC2,
then the spatial relation should absolutely be assessed before MC1 is skeletonised

(this rule has no consequence in our use case)

Rule:
If a mapped concept has to be filtered and cartographically generalised,
then it is more efficient to proceed in this order.

Consequence in our use case (see section 3.1):
Roads must be filtered before being cartographically generalised, but this order relation has a low weight.

At the end of the process, the global plan is a graph where vertices are goals and edges are weighted sequencing relations. If there are any loops, they must be removed by keeping only the highest weighted edges.

Tasks plan and operations

Each goal of the global plan can be reached through a more or less complex process relying on the model of tasks of (Bucher 2002). A complex task is recursively composed of tasks down to the level of atomic tasks. An atomic task corresponds to an operation defined in the semantic referential. Tasks have input and output variables, whose domains are Geo and Gis concepts from the semantic referential. Specifying a task means choosing its subtasks and setting its input variables. This is done depending on the context (specifications, selected data sources and other specified tasks), guided by the methodological knowledge associated with each task. Executing a task means executing its subtasks, which finally comes to performing operations from the semantic referential. Selecting and invoking the corresponding tools is called concrete composition. It is not the duty of the planner but of the workflow engine.

3. Prototype implementation

This section presents our implementation of the models. Our first prototype focuses on the core on-demand mapping functionalities only. The following limitations are introduced:

- Formal target map specifications already exist
- Only one data source per mapped concept meets the requirements. It is still identified in a catalogue using the semantic referential.

- Only one derivation tool is available to perform each atomic task. This tool is recorded in a registry using the semantic referential.

Section 3.1 presents our use case. Section 3.2 reports our experimentations and difficulties.

3.1 The cycle routes use case

Although the system is generic, we are basing our experimentations on a use case: a user needs a map dedicated to route planning and navigation at a regional level (scale range 1:200K to 1:100K) and including cycle routes provided by his sport club. The specifications of this map are as follows.

The map foreground is composed of the road network and landmarks. Roads are symbolised with 4 importance levels. Every road followed by a cycle route must be displayed, even if it would normally be omitted at the considered scale range. Selected landmarks are special buildings (castles and churches), standard buildings if they are isolated from built areas, and river crossings. Buildings are represented by surfaces, except for special buildings situated inside built areas, which are represented by punctual symbols. Some touristic information and the cycling routes constitute the second level of reading of the map. Where they follow a road, these cycle routes must share geometry with the referential. Cycle route sections supported by a cycle lane must be displayed with a specific style, although cycle lanes themselves are not displayed on the map. The built areas, forests, river network and relief constitute the map background.

The test area is located near Manchester. Referential data are from the OSGB internal high-resolution database. Cycle lanes and equipped road sections were graciously provided by Transport for Greater Manchester[2]. Some user's advised cycle routes were found on the web[3] and some were created by hand.

3.2 Developments

The prototype is being developed in Java. We chose Geotools for the manipulation of spatial data, JBoss Drools for rules and inference, OWL API and HermiT to reason over the semantic referential. Models of target product specifications and data sources are encoded as java packages and as XML schemas with a JAXB binding.

A catalogue of data sources and some product specifications were created in XML for the use case. Figure 3 displays an extract of specifications.

---

[2] http://www.cyclegm.org

[3] http://www.gps-routes.co.uk/routes/home.nsf/cyclemap/Greater%20Manchester

```
- <odm:Theme id="idvalue1" title="communications">
  + <odm:MappedConcept id="idvalue2" name="roads" odm:forDisplay="true"
    odm:fromUser="false" odm:levelOfDetail="high" odm:readingLevel="first sight">
    </odm:MappedConcept>
  + <odm:MappedConcept id="idvalue14" name="cycle lanes" odm:forDisplay="false"
    odm:fromUser="false" odm:levelOfDetail="high"></odm:MappedConcept>
  - <odm:MappedConcept id="idvalue12" name="cycle routes" odm:forDisplay="true"
    odm:fromUser="true" odm:levelOfDetail="low" odm:readingLevel="first sight">
    - <odm:GeoConcept>
        http://www.owl-ontologies.com/2011/OnDemandMapping
        /GeoConcepts.owl#cycle_route
      </odm:GeoConcept>
    - <odm:ModellingConstraint id="idvalue6">
      + <odm:SpatialModellingConstraint id="idvalue61"
        naturalLanguageDescription="cycle routes are linear">
        </odm:SpatialModellingConstraint>
      + <odm:ThematicModellingConstraint id="idvalue62"
        naturalLanguageDescription="condition for cycle-friendly routes">
        </odm:ThematicModellingConstraint>
      </odm:ModellingConstraint>
    - <odm:IntegrationConstraint formalExpression="" id="idvalue10"
      naturalLanguageDescription="cycle routes aligned with roads">
        <odm:PreferenceRank>1</odm:PreferenceRank>
      - <odm:GeometricAdjustment topologicIntegration="false">
        - <odm:GuidedByRelation>
            <odm:AlignedWith tolerance="20"/>
            <odm:RelationToMappedConcept>idvalue2</odm:RelationToMappedConcept>
          </odm:GuidedByRelation>
        </odm:GeometricAdjustment>
      </odm:IntegrationConstraint>
    </odm:MappedConcept>
  </odm:Theme>
```

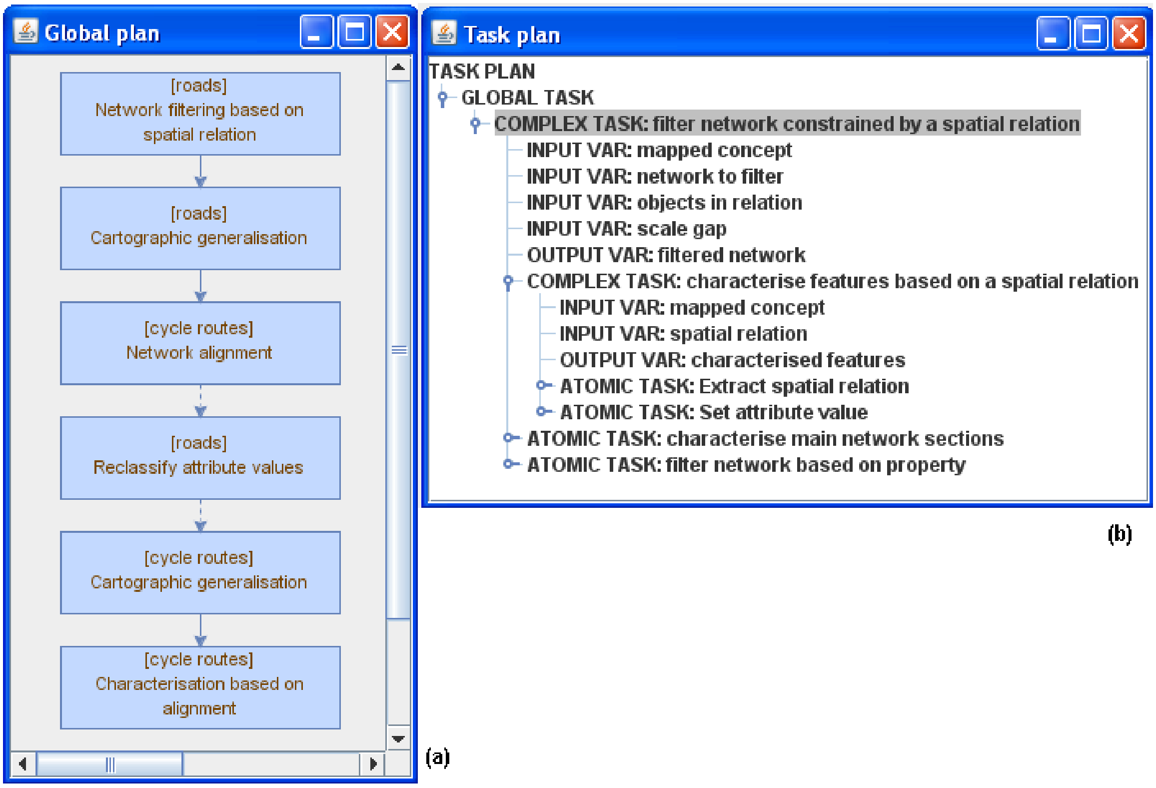
Figure 3: Specification extract: the "communications" theme with 3 mapped concepts (road, cycle lane, cycle route). The integration constraint for cycle routes is developed.

Plan templates were created for Roads and User Itineraries, as well as a default plan template which is used when no ontological match is found between mapped concepts and available plans. These plan templates and the goals composing them are simply encoded as Java classes.

Rules were written in the Drools dialect. With less than 10 planning rules and sequencing rules (more sequencing rules are implicitly represented by the order in which goals are defined in the plan templates), the planner produced the first part of the global plan as shown on Figure 4(a).

The task model was implemented by an interface (Task) and two abstract classes (ComplexTask and AtomicTask). They define the generic mechanism of cascading specification and execution of tasks, and leave space for context-dependent methodological knowledge (i.e. specification of variables driven by map specifications). A few complex task templates were created and proved the generic mechanism. The complex task “Network filtering constrained by a spatial relation” is represented as a tree on Figure 4(b).

At the lower end of the task plan, atomic tasks are the only tasks having a non-generic execution. A registry class performs a 1 to 1 mapping between atomic tasks and available tools, using operation names as a pivot. For now, the only tools invoked by the system are java methods from external projects. We notably invoked methods of the network matching API from Geoxygene[4]. These methods perform the *conflation* and calculate the *follow* relation between network sections. As reported in section 4, we also tried to interact with a relation-based system to extract a spatial relation. Accessing web services is the next step. To prepare to it, we wrote a Geotools client for WebGen[5].



[roads]
Network filtering based on spatial relation

[roads]
Cartographic generalisation

[cycle routes]
Network alignment

[roads]
Reclassify attribute values

[cycle routes]
Cartographic generalisation

[cycle routes]
Characterisation based on alignment

TASK PLAN
- GLOBAL TASK
  - COMPLEX TASK: filter network constrained by a spatial relation
    - INPUT VAR: mapped concept
    - INPUT VAR: network to filter
    - INPUT VAR: objects in relation
    - INPUT VAR: scale gap
    - OUTPUT VAR: filtered network
    - COMPLEX TASK: characterise features based on a spatial relation
      - INPUT VAR: mapped concept
      - INPUT VAR: spatial relation
      - OUTPUT VAR: characterised features
      - ATOMIC TASK: Extract spatial relation
      - ATOMIC TASK: Set attribute value
    - ATOMIC TASK: characterise main network sections
    - ATOMIC TASK: filter network based on property

Figure 4: (a) Graph representing the global plan. (b) Tree representing the task plan.

Difficulties

So far, writing rules tends to be the most difficult part of the system implementation. The rules we are writing are as general as possible, even if they are determined by our use case. When the system grows, the risk of writing general rules is to fire too many rules, or not to take the good decision because some rules are contradicting each other. How to handle these problems has not been studied yet.

Unsurprisingly, binding atomic tasks with external tools is time-consuming and sometimes problematic. Several adapters have been implemented. Some external tools have proved to drop feature attributes or modify feature identifiers, which can impair the rest of the workflow. In the following section, an atomic task is bound with a process developed by a collaborator from IGN France.

4. An experiment to collaborate with an external system based on spatial relations

This section presents a collaboration with IGN France, where Kusay Jaara, researcher of COGIT laboratory, has been working on the extraction and the usage of relations

---

[4] http://oxygene-project.sourceforge.net/

[5] http://kartographie.geo.tu-dresden.de/webgen_core/index.html

between thematic user data and reference topographic data (Jaara et al. 2012). As shown by examples provided in section 2.1, spatial relations can play an important role in the specifications of an on-demand map. The extraction of these relations may be planned in different tasks and used in subsequent tasks (e.g. the *follow* relation between cycle routes and roads may control the filtering of roads, ensuring every road supporting a cycle route is kept).

Jaara is developing a *relation-based system.* This is a system extracting relations, storing them and making them available. The relation-based system can perform tasks for other systems, like extracting some relations for an on-demand mapping system, or assessing the satisfaction of some relational constraints for a generalisation system. The relation-based system can also perform complex built-in processes based on spatial relations, such as the data migration process currently being developed by Jaara. This process is used when user's thematic data, that are initially combined and consistent with a topographic database, must be transferred to another topographic database (e.g. a generalised version or an updated version of the initial one). The migration process relocates the thematic data by preserving the meaningful spatial relations existing between them and the initial topographic data (e.g., the fact that an accident is at the turning point of a bend).

The principle of the present collaboration is centred on spatial relations: the relation-based system would provide relation occurrences for the on-demand mapping system (Figure 5).

The main challenge to make both systems interoperate was to determine a shared formalism to express relations. Each spatial relation being a special case, we proposed some modelling facets and used them to model relations of our use-case. This work, participating to a more global initiative to enrich our semantic referential with spatial relations, is described in (Touya et al. 2012).

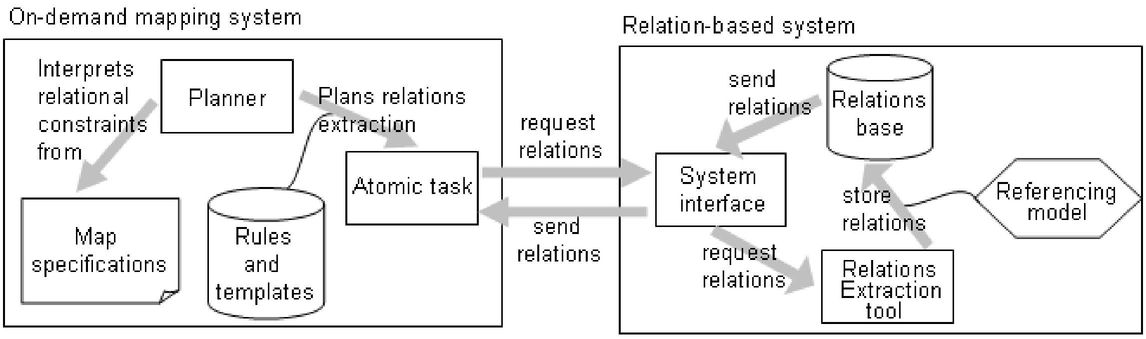


Figure 5: Interaction between the on-demand mapping system and the relation-based system

As shown on Figure 5, the relation-based system receives relation requests from the on-demand mapping tasks, using the shared formalism. If these relations are not already available, the extraction tool extracts and stores them in the relations base, using the internal referencing model of (Jaara et al. 2012). The system interface then delivers occurrences of relations to the on-demand mapping system using the shared formalism.

Other collaboration directions could be explored. An experiment would consist in using the planning mechanism described in section 2.3 to help guiding the data migration process described above, considering the spatial relations to preserve as specifications for the on-demand mapping system.

5. Conclusions

We still have a long way to go before delivering a custom map taking all requirements of our use case into account. Our core models are being tested progressively: selection, modelling and integration constraints are currently taken into account by the planner, but not symbolisation and schema constraints. Even when we have done, we suspect our system will be strongly coupled to the use-case. A large amount of knowledge will still have to be inserted as rules and templates to make it more flexible.

Future steps will focus on current limitations: the collection of user requirements, the semantic-driven choice of data sources and the identification of the best available derivation tools. Progressing through collaboration has been rewarding so far and we expect to follow the same way to explore these challenging topics.

References

Andrei M, Berre A, Costa L, Fitzner D, Hoffmann J, Klien E, Langlois J, Limyr A, Maue P, Schade S, Steinmetz N, Tertre F, Vasiliu L and Zastavni N, 2008 “SWING: A Geospatial semantic web environment.” *Workshop: Semantic Web meets Geospatial Applications, in conjunction with the 11th AGILE Conference on Geographic Information Science*. Girona, Spain.

Balley S and Regnauld N, 2011a, Models and Standards for On-Demand Mapping. In: *Proceedings of 25th International Cartographic Conference (ICC'11)*, Paris, France.

Balley S, and Regnauld N, 2011b, Collaborating for better on-demand mapping. In: *Proceedings of 14th ICA Workshop on Generalisation and Multiple Representation*, Paris, France.

Bucher B, 2002. L'aide à l'accès à l'information géographique : un environnement de conception coopérative d'utilisations de données géographiques. PhD Thesis, Université Paris 6.

Friis Christensen A, Lucchi R, Lutz M, and Ostländer N. 2009, Service chaining architectures for applications implementing distributed geographic information processing. *International Journal of Geographical Information Science* 23(5): 561-580.

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.

Gould N, 2012, Semantic description of generalisation web services for on-demand mapping. In: Pundt H and Bernard L (eds), *Proceedings 1st AGILE PhD School*, 38–48. Shaker Verlag.

Jaara K, Duchêne C, Ruas A, 2012, A model for preserving the consistency between topographic and thematic layers throughout data migration. In: *Proceedings of 15th Symposium on Spatial Data Handling (SDH)*, Bonn, Germany.

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.

Lutz M, 2007, Ontology-based descriptions for semantic discovery and composition of geoprocessing services. *GeoInformatica* 11(1).

Touya G, Balley S, Duchêne C, Jaara K, Regnauld N and Gould N, 2012, Towards an Ontology of Generalisation Constraints and Spatial Relations. *15th ICA Workshop on Generalisation and Multiple Representation*, Istanbul, Turkey.