GIS

GEOGRAPHICAL INFORMATION SYSTEMS SERIES

Innovative Software Development in GIS

Edited by

Bénédicte Bucher and Florence Le Ber

iSTE

WILEY

Innovative Software Development in GIS

Innovative Software Development in GIS

Edited by
Bénédicte Bucher
Florence Le Ber

ISTE

WILEY

First published 2012 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

Library of Congress Cataloging-in-Publication Data

Innovative software development in GIS / edited by Florence Le Ber [and] Benedicte Bucher.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-84821-364-7
1. Geographic information systems. 2. Geography--Data processing. 3. Geomatics. I. Le Ber, Florence. II. Bucher, Bénédicte.
G70.212.I556 2012
910.285--dc23

2012008578

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN: 978-1-84821-364-7

Printed and bound in Great Britain by CPI Group (UK) Ltd., Croydon, Surrey CR0 4YY

MIX
Paper from responsible sources
FSC® C013604
FSC www.fsc.org

Table of Contents

Chapter 1

Introduction

Research in geomatics must face major challenges to improve the management of the interaction of humankind with the planet at various levels. These challenges cover types of problems such as risk management (monitoring a volcano), sustainable development (the prevention of coastal erosion or the control of increasing urbanization in a given area), or even societal issues, such as the accompaniment and improvement of the integration of positioning techniques and their mobile applications in our everyday lives. To process these issues, we often need to turn to computers and develop software that can meet the requirements of the data handled. The goal of this book is to study the innovative software development activities carried out by geomatics research teams, and more specifically to analyze which of these development activities can be pooled, and whether it is relevant to do so, in the sense that it promotes research activities. We have chosen to focus on one aspect of geomatics research: the design of models and analysis methods to utilize geographical data.

Chapter written by Bénédicte BUCHER and Florence LE BER.

The rest of Chapter 1 clarifies the contextual elements that are essential to the study of geomatics, and more specifically the definitions of the terms used. We successively clarify the notions of geomatics software and pooling in our context before presenting the goals and structure of the book.

1.1. Geomatics software

Geomatics is a technical and scientific field derived from geography and computer science. It develops methods to represent, analyze, and simulate geographical space. Its goal is to improve the understanding of this space and the management of human activities and human interventions on the planet. Thus, the core activities of geomatics is made up of techniques of Earth observation as well as techniques of model design – mainly maps – useful for analysis and reasoning. The traditional spatial representations are printed maps, gazetteers, or lists of triangulation points. For the past 20 years, geographical data have become digital and geomatics has been characterized by the intensive use of computer science. This development is highlighted by two phenomena. The first is the increase in data, specifically satellite data, and this increase requires the development of automatic processing. The second phenomenon is the increasing role of geographical information in information infrastructures (use of maps on the Web, localized services, etc.).

1.1.1. *Digital geographical data*

A core specificity of geomatics is its data.

A primary aspect is the distance between the data and the information represented through them. This is partly due to the fact that space observation often happens through the measurement of physical signals that must then be interpreted into meaning. This distance between the data and the information is also due to the difficulty in representing the notion of position in space so as to carry out operations on the shapes of the objects and the spatial relations they represent. More specifically, a digital model of geographical space must render two important notions: positioning in space and the nature of the phenomena. Positioning in space is shown through projections, which relate the different parts of the Earth's surface to an ellipsoid linked to coordinates in a stable mathematical referential versus the Earth. Geographical projection is usually followed by a cartographic projection to view the data on a plane screen. Thus, part of the Earth's surface or its subsurface is positioned by a geometry provided with coordinates – eventually reduced to a point. From there, two major positioning methods exist: the *vector* and the lattice [COU 92]. For example, a road is generally represented by an object of linear geometry (corresponding to the axis of the road on the ground) with attributes taking its nature into account (identification number, classification, and type of surface). This is a *vector* model. However, in three-dimensional (3D) virtual worlds, roads are often not represented in the data as *vector* objects, but the human user can see them in the terrain image (due to texture). Other phenomena, such as air pressure, must be represented as fields which have a given value in any point of space. More specifically, discretized versions of these fields are used. These are lattice models. The continuous/discrete duality that exists at the level of the observed reality and in both models of representation can also be found in the principles of software development and sometimes leads researchers to adopt different approaches to study one phenomenon. When we study a city, for example, we use ORBISGIS with a preference for lattice representation manipulation and GEOXYGENE with a preference for the manipulation of *vector* objects. Overall, the choice of a representation often frames a domain of expertise and the joint manipulation of two types of representations remains complex even though there exist proposals to integrate them [LAU 00].

A second specificity of geographical data is the multiplicity of models built to represent geographical space in the data [BIS 97]. As [WOR 96] mentions it, geographical space isn't a *table top space*, which is a space observable from outside, similar to objects placed on a table. It is a space in which each person acts, and builds, a representation of the space in the context of his/her own action. For example, the information obtained from a geographical landscape isn't the same depending on whether the user is interested in road transport, risk management, or development. Differences appear at the level of the types of relevant objects: the watering places and pools are remembered by the fireman but not by the hauler. Differences also appear at the semantic and geometrical levels of detail: a building can be represented by its footprint and access points or in a simplified manner. Beyond the real-world ontology that is used – the categories of objects of the world observed and the logical diagram – the data also sometimes depend on specific rules of representation, such as a building of less than 20 m^2 is represented by an object of the `IsolatedConstruction` class if it is highly isolated (over 100 m from another building). Finally, the coding of the data and the required geometry discretization leads to other choices that can vary from one producer to the other.

All in all, the manipulation and interpretation of geographical data requires dedicated software and expertise. Moreover, the heterogeneities in the data stand in the way of pooling.

1.1.2. *GIS-tools*

A very popular type of software in geomatics is the geographical information systems tool (GIS-tool), which allows the manipulation of geographical data. The term "tool" allows us to distinguish the piece of software from the complete system made of data, software, and users. The term GIS generally refers to the entire system. From now on in this book, we will use the term GIS to refer to a GIS-tool. A GIS is characterized by many functionalities that are essential in geographical information and detailed as follows. Up until the 1990s, GIS software fulfilled all these functionalities. Monolithic architectures then became architectures made up of modules dedicated to various functionalities, which are required to use the geographical data. This evolution was helped by interface specifications between GIS components produced by International Organization for Standardization (ISO) and Open Geospatial Consortium (OGC)[1]. These specifications were deliberately made abstract at first so they wouldn't restrict the market. Implementations were quickly suggested and included into the standard ones: XML implementations for the interoperable Web service components and JAVA (GEOAPI) implementations for interoperable libraries. Today, the notion of GIS thus refers to an information system made up of data and functional modules. It holds definite interest for pooling since it encourages researchers to focus on their core interest and reuse functional modules for the supporting functions they need.

The GIS functionalities were referred to in France by the acronym "5A": "Acquire", "Afficher" ("Display"), "Archive", "Abstract", and "Analyze" [DEN 96]. A sixth "A", for

1 The glossary presented at the end of the chapters gives an inventory of the organizations, tools, and formats quoted in this book.

"Anticipate", appeared along with the concern about sustainable development and simulation software.

The acquisition of geographical data in a GIS essentially consists of importing existing data. The software must thus be capable of reading the more common formats, which is greatly aided by the generalized adoption of standard formats such as ESRI's *shapefile* format or the GML format proposed by ISO/OGC [ISO 07]. The software must also allow the interpretation of models with imported data that is still problematic in spite of the many schema transformation tools such as the FME Workbench of the Safe Software company. Schema transformation is still an active research field today [BAL 07]. The software should also allow the direct creation or editing of geographical data, for example the description of a new piece of road by creating an object and drawing its geometry on a referential map. The function of integration and fusion mentioned by [STE 09] is also important at this stage. It is made difficult by the differences between the geographical space representations mentioned earlier. Indeed, a new list, which goes into more detail, of nine functionalities was recently suggested by [STE 09] to define a *GIS software* in a geographical encyclopedia: visualization, creation, editing, storing, integration/merger, transformation, query, analysis, and map writing. This list does not have acquisition but details the integration functionalities that are the key functions to build the database of a geographical information system. Finally, due to the rise of distributed architectures, the acquisition function is now doubled up with a function to discover existing data and existing functionalities. The MDWEB software presented in this book is a solution to this need provided by research teams (IRD and the University of Montpellier). The software was designed as a specific component of a GIS architecture, and turned out to be the most able to simply complete existing structures since it does not offer redundant structures and its interface is clearly identified.

The display is available in various functions: visualizing the data geometry, visualizing their attributes, and writing and visualizing a map from these data. The last function requires the association of geographical data and cartographic styles, and then to draw the corresponding figure, which means having graphical objects linked to geographical objects. The cartographic representation is specifically studied in the GENGHIS proposition described in this book. A cartographic style is the association between a piece of information and a graphical symbol. The styles are defined for object classes such as roads and avalanches and eventually refined within a class according to the attributes of the said class: roads, for example, are represented differently depending on the value of the "classification" attribute given to the road. It was for a long time impossible to transfer a legend (from the cartographic style definition) from one type of software to another, due to the lack of a standardized format. The current proposition of the OGC consortium, entitled *Styled Layer Descriptor*, aims to become just such a standard. Besides, within the context of pooling, display processing is not simply about being able to transfer a display specification from one type of GIS software to another. It is also about knowing how to adapt the display of data to the context. This issue has been studied in the field of collaborative GIS architectures, which aim to allow multiple actors (such as researchers) to work on the same set of data.

Abstraction corresponds to the possibility of creating and manipulating a more or less sophisticated model of geographical space. For example, if a user uploads a set of points from sensors, describing temperature and humidity data, a first level of abstraction would be to create zones in which these values are described as average and a second level of abstraction would be to create a classification of these zones. As we have mentioned it previously, there is no universal model to represent space. Within a GIS, abstraction also corresponds to the information formatting before its processing. There is also here a great diversity of abstraction models, which complementarity isn't always simple to explore, such as the abstractions based on agents or the abstractions based on cellular automata, such as [BAT 05] does for cities. The analysis carried out in a GIS corresponds to complex operations or reasoning on spatial properties or relations of the phenomena represented, as for example, the choice of the buildings surrounding an airport, or the calculation of an itinerary. In geographical information, the query is specifically complex since it often uses various criteria: the position in space, the nature, and the position in time. Moreover, the spatial criterion is multidimensional. Owing to their volume, it is usually necessary to index geographical data to allow these requirements. The construction of spatial indexes is made complex by the multidimensional nature of localization [KAM 08]. Moreover, the indexed objects can evolve, for example a fleet of taxis or planes [WOL 99]. Or the query itself can evolve, for example the query, made by a user on the move, for the closest Vélib bicycle docking stations in Paris, which is also called a continuous query [TER 92]. All this requires the organization of indexes so that they allow complex spatiotemporal queries, are not penalized by updates, and allow for a swift answer to a changing query. In this book, the GEOLIS software presents a different abstraction from the classical entity-relationship model to organize geographical data so that we can carry out exploration queries on them. Finally, the rise of the Web, and the first Web document, increased the importance of unstructured information searches. In this field, it is important to take into account the geographical dimension, since a major part of the queries made over the Web have a geographical dimension. Providing software that manages the spatial component in the indexation and the classification of answers improves search engine performance [PAL 10, PUR 07].

Analysis carried out in a GIS corresponds to the possibility of automatically carrying out complex operations or reasoning on the properties and spatial relations of the objects represented, such as the buildings around an airport, or the calculation of an itinerary. Among the functionalities defined by [STE 09], we have the query function. The query is specifically important and complex in geographical information for it requires the indexation of information under various crossed criteria: the position in space, the nature, and the position in time. In this book, the GEOLIS software offers a different abstraction from the classical entity-relationship model to organize these elements of geographical data aiming to make exploration queries on this data. The manipulation of spatiotemporal data has increased in importance, whether to manage moving objects or dynamic objects. The GENGHIS software presented in this book is dedicated to the implementation of spatiotemporal information systems (STIS).

1.1.3. *Software innovation and geomatics research*

Geomatics research aims to improve the knowledge and tools of geomatics, as well as promote the use of this knowledge and these tools and their integration into the information society. It is a multidisciplinary field, essentially made up of human and social science researchers and of computer science researchers, but also of researchers from other scientific fields such as law and signal processing. The research group MAGIS, "Méthodes et applications pour la géomatique et l'information spatial" (Methods and applications for geomatics and spatial information), covers 42 research laboratories and institutions. The research carried out in these laboratories focuses on localized services, new map types, models and applications for sustainable development, geographical information integration, spatial analysis, simulation, and geographical information science epistemology, among others.

Geomatics research is often inseparable from software usage to manipulate geographical data, whether they are complete GIS systems or specific modules. Researchers can be users. For example, geography researchers rely on GIS software to improve the knowledge of certain phenomena. Many models developed to study spatial phenomena, such as the erosion of agricultural land [DER 96], runoff and flooding [LAN 02], urban development [PIO 07, SIR 06], rely on sets of data stored in GIS that produce new data.

Researchers can also be developers, either to develop an *ad hoc* tool or suggest software innovations, which are developments whose scope is not restricted to solving a specific case. Some researchers work by developing extensions to existing software where these offer a programming interface, whether to offer new processing procedures or enrich a data model. These are typically works based on the ARCINFO software, widely used in American universities, or on the GRASS software, one of the first free pieces of GIS software. The ESRI international user conference thus welcomes some communications from researchers, the proof of which is the publication every year of a special issue of the scientific journal *Transactions in GIS* [WIL 10]. Other researchers ascribe to the development of a new tool. For example, this was the case for the graphical query interfaces CIGALES [MAI 90] or LVIS [BON 99], as well as for projects presented in this book.

Innovation can lie in the development of new analysis methods based on theories from mathematics or knowledge engineering fields. It can also be by suggesting a new interface to disseminate existing functionalities on a broader level.

Or yet, the innovation can be in the architecture itself. The range of corresponding software solutions is wide: 3D view reconstruction from pictures, multiagent architectures for distributed processing, a mobile data management system, robot cartographer, geographical search engine, etc. Innovation can also pertain to the development of tools specific to certain research programs, tools which allow the manipulation of geographical data, and which can be considered as future functionalities of GIS-tools. In this book, we will present GENEXP-LANDSITES a software dedicated to the simulation of virtual landscapes. It aims at exploring the variability of agricultural landscapes and considers different cases for the spatiotemporal organization of agricultural production. So GENEXP-LANDSITES belongs to the sixth "A" (Anticipate) of the GIS-tools. Let us emphasize that software innovation in geomatics is also due to other actors rather than researchers, such as the military or private companies. We can, for example, mention the GOOGLE MAPS API that offers a functionality for new users: integrating a map into a website with eventually a specific overlay. This functionality was already available through Web extensions for classic GIS software, but the innovation was to offer it to geomatics novices due to use of simple language.

Thus, change in geomatics is partly tied to the evolution in computer science, it follows them, and improves them. The main software innovations that have stood out in the field of geomatics in the last few years are in part the evolutions of architectures distributed toward the Web, *grid computing*, *cloud computing*, ubiquitous computer science, and ambient intelligence, as well as the phenomenon of the semantic Web, robotics, and miniaturization. In the last few years, for example, we find distributed GIS, especially on the Internet. These distributed architectures favor the implementation of participative GIS, which create new problems beyond the pooling of software components [MAR 08, TUR 08], due to the rise of ubiquitous environments, localized services and ubiquitous cartography that also rise in importance.

1.2. Pooling

The term “pooling” is derived from the verb “to pool”, which can be defined as “to combine (as resources) in a common fund or effort” [MIS 93]. The term was used for information technology applications, as early as the introduction of these applications in small businesses and communities, to essentially mean the sharing of upkeep and update costs. The term “information technology pooling” is also used in research and training about data and resources, such as linguistic resources [PIE 08]: the goal is to offer access to all the information and knowledge produced by every person and thus promote knowledge dissemination and progress. In this book, we consider the term “pooling” as meaning the pooling of resources that come into play during the design and development of software, aiming for shared benefits. These resources can be varied: abstract models, code, programming interfaces, financing, or yet experience in project management.

1.2.1. *The need for pooling and its relevance*

The relevance of pooling is true for any field of research focusing on innovation. Indeed, a specific type of pooling is sharing methods, making one’s methods accessible to others and vice versa. By sharing methods, we promote their improvements as well as the comparison between the methods, and thus progress. It also allows the pooling of effort on certain components, and thus enables us to go faster. This book holds such an example: the WebGen project aims to facilitate the comparison of different implementation with the same function of introduction, to facilitate the progression in this field of research. Another example of innovation pooling is the European project SPIRIT, whose goal is to design a search engine based on geographical knowledge. The design and implementation of the engine required the collaboration of teams specializing in research on information, spatial analysis, and visualization. The pooling of the software contributions of the various teams took place within a service-based architecture whose interface contracts were defined during a joint project [FIN 03].

We should also note that the research teams use and sometimes improve other pieces of software necessary to their activities in higher education and research in general, such as article writing, presentation preparation, sharing courses, setting up websites for conferences, as well as all the management activities required by an institution which relies on digital information systems. This book does not focus on these tools. That said, the necessity for pooling solutions to support these activities has been proved and an answer has actually been provided by the PLUME[2] project, or by the implementation of the university and higher education and research institution pooling agency[3]. Other initiatives focus on digital documents such as the HAL[4] or ARXIV[5] archive sites – which gather researchers' scientific publications – or even the ORI-OAI[6] software that creates digital document sharing portals between education and research institutions.

1.2.2. *Reflection opportunity on geomatics pooling*

A reflection on the possibilities of pooling software development projects carried out in geomatics research teams

2 http://www.projet-plume.org

3 http://www.amue.fr/

4 http://hal.archives-ouvertes.fr/

5 http://arxiv.org/

6 http://www.ori-oai.org/

is all the more timely now that the techniques allowing us to interoperate software components, to cooperate on the design of a module, to design reusable components, or even to reuse existing components have improved and are widespread in software development.

These techniques are first and foremost, in geomatics, norms and standards concerning interfaces between components manipulating geographical data. In the field of geomatics, these standards mostly come from the ISO and its technical committee TC211 as well as the OGC. Specification may concern exchanged data, as in the *Geographic Markup Language* norm for instance, or functionalities, as in the *Web Feature Service*, *Web Map Service*, and *Catalogue Service for the Web* norms.

These techniques also cover methods and correlated collaborative development tools, OMG method [OMG 08], software project management tools, such as *Enterprise Architect* as well as middleware techniques aiming to encourage the reuse of software components with mediation architectures or component architectures [KRA 06]. A key architecture is, for example, the Web service architecture that corresponds to an architecture based on loosely coupled components on a widely accessible network. Another proof of the maturity of middleware techniques is ubiquitous architectures [WEI 93, WAL 97].

A particularly interesting standard for us is the *Web Processing Services* standard proposed by OGC. It focuses on the online availability of geographical data processing to promote sharing and reuse.

Another element promoting pooling is the success of *open source* software projects. Indeed, having access to a software's sources promotes its understanding and reuse due to the code and debugging documentation.