A Software Process Framework for Guiding the Construction Specification of Geospatial Databases

Yolanda M Fernández-Ordóñez1, Jesús Soria-Ruiz1 and Antonia Macedo-Cruz1
1Geomatics Laboratory, College of Postgraduates in Agricultural Sciences, Campus Montecillo, Texcoco, Mexico
2Geomatics & Agro-climatology Laboratory, National Institute of Research for Forestry Agricultural and Livestock (INIFAP), Zacatecas, Mexico
3Hydro-science Department, College of Postgraduates, Campus Montecillo, Texcoco, Mexico

Abstract
Geospatial databases are in increasing demand for serving a variety of applications and user expectations. The general requirements for building such databases are understood, but technological tools and approaches continue to be developed because of the specific requirements of operating environments. However, in many situations, the implementation of domain-specific databases still presents major challenges. In interdisciplinary research environments where groups work on a common geographical region, several issues arise from the use of diverse sources of geospatial data. Such issues need the input of software specialists, and researchers usually lack the time and/or expertise required to become substantially involved in informatics projects. Groups investigating natural and social processes related to agriculture, ecology or natural resources management need to solve data incompatibilities in order to share and reuse information. Accordingly, this paper presents MP-Geo, a framework for a software process specification for guiding the construction of a specialized geospatial database and its management system. The salient characteristic of the framework is its emphasis on and inclusion of software engineering good practices and standards. The framework emerges from the real needs of academics and students working in the field of agricultural and natural resource management. The framework can be used to build useful geospatial databases in diverse domains.

Keywords: Geospatial databases; Software development framework; Software engineering

Introduction
General context
Geospatial databases serve a variety of applications and user expectations, and the demand for such databases is increasing. Earth observation initiatives using remotely sensed methods and instruments have contributed to a rapidly growing volume of geospatial data made available to both experts and casual users. Many initiatives have addressed the various problems of organizing geospatial databases and making them available for different types of applications and environments [1]. For example, Latre et al. [2] highlight that ‘the level of maturity or sophistication of e-government services is not improving in those areas that require geospatial information’. The same authors mention that although geospatial data are becoming increasingly pervasive, managing and using them is complex. The principles for building geospatial repositories are understood, but recommendations, standards, and management tools are continuing to be developed in view of the impacts and potential impacts of natural hazardous phenomena and of other environmental threats such as climate change and water scarcity [3].

In scientific research contexts, the users of geospatial data to be acquired from global, national, or regional sources face both practical and theoretical challenges regarding the production of knowledge [4]. Among these, a major issue is locating the best data sources with guarantees of quality and appropriateness for a specific use, because the selection of the data source(s) has an impact on the quality of the results. The tasks of determining the right data, and of preparing, integrating, and sharing them, require the investment of considerable effort and time. Even when researchers are preparing to study phenomena that lie within the same geospatial area, steps in the process are often executed as piecemeal actions. The main concerns include solving heterogeneity issues regarding format and meaning, and accounting for data gaps where the required items are either nonexistent, not sufficiently up to date or not at the desired spatial or temporal scales. Many computational processes and treatments precede the readiness of both geospatial and non-spatial data for their analysis within scientific research environments.

Software research and development issues
The implementation of software frameworks to support common geospatial related tasks is a major endeavor [5]. Proposals to include value-added management functionalities in geographic information systems are being explored to address recent challenges in the use of both geospatial and non-geospatial data [6]. With respect to the context outlined in Section 3.1 above and to the economies that could be gained by reusing geospatial data elements, we contend that an appropriate first step is the construction of specialized repositories that satisfy the data management requirements for defined research communities.

Therefore, the objective of this paper is to present a framework for software specification to aid in the development of a domain-specific geographic database and its management system. Stakeholders' participation, allowing their information requirements to be captured, and adherence to software engineering standards are salient elements of the framework. The context from which the present work emerges is a research endeavor in agricultural and natural resources management. A project called ‘Geographical Data Base - College of Postgraduates’ (GeoDB-CP) in an agricultural sciences research institution, the College of Postgraduates in Agricultural Sciences (CPCA) in Mexico, has been developed to facilitate current and future activities foreseen.

*Corresponding author: Yolanda M Fernandez-Ordonez, Geomatics Laboratory, College of Postgraduates in Agricultural Sciences, Campus Montecillo, Texcoco 56230, Mexico, Tel: +52-59-5952-0200; E-mail: yfernand@colpos.mx

Received December 01, 2017; Accepted December 18, 2017; Published December 24, 2017


Copyright: © 2017 Fernández-Ordóñez YM, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
within the institution.

The research activities considered in test-bed developments call for interdisciplinary groups to work in a coordinated manner on geospatial phenomena within defined common areas. Such areas might include a watershed, a municipality an agricultural district, or an arbitrarily defined geographic region. The precursors of the Geo DB-CP project produced versions of a geospatial database that acted as the initial test-beds for our proposal of the framework [7,8]. The interdisciplinary groups in the CPCa change over time, and so do the projects and research lines according to academic policies [9].

The remainder of the paper is organized as follows. Materials and methods constitutes Section 4. The rationale of our proposal and an overview of the problem through examples are included here. Section 5 describes the framework. Section 6 contains results and discussion and some final remarks, including research perspectives.

Materials and Methods

This section explains our rationale for defining the framework through two real examples that raised our awareness of and justify the need for comprehensive shareable geospatial databases. The examples concern research projects conducted within the CPCa and by the Ministry of Agriculture in Mexico. Our work was planned as a guide for the design and implementation of geospatial databases based on a unified software modeling perspective [10], bringing into play recognized good software engineering practices and standards. Those requirements addressing the reuse or modification of geospatial data are the main drivers of our conception, the outcome of which is a block-based process framework leading to a software system specification.

Motivating examples

Multi-temporal land-use change: The objective of this example project was to analyze multi-temporal land-use change in the Metztitlán watershed in Mexico (Figure 1). Studying phenomena within this area holds importance because of the rich biotic and abiotic components and because it is a major agricultural center. The watershed also contains the Metztitlán Municipality, irrigation districts, rainfall-dependent agricultural areas, and protected natural areas [11]. Unplanned growth and population increase have negatively modified the land-use pattern, promoting social conflicts that impact land stability. This study by a multidisciplinary group required hard-copy maps, ancillary satellite images, and other official socio-economic data for three periods (1973-1985, 1985-2000 and 2000-2007) to obtain Markov-based predictions for the year 2030 for the land-use pattern and its consequences [12]. The required data were not readily available in a centralized repository, so the research project participants gathered and processed the data in a piecemeal fashion, an inconvenience mentioned before.

Compact agricultural areas: Precisely defined agricultural areas in Mexico are monitored with respect to their behavior regarding production and rural productivity by researchers appointed by the Ministry of Agriculture [13]. These pilot areas will be organized through a network of agro-technological observatories (AOTs). Figure 2 shows the example of the 160,000-ha AOT 19, whose agricultural study area covers 101,955 ha, as determined using a series of geomatics processes [14]. To optimize agricultural productivity in the pilot areas, a multidisciplinary and holistic research approach is proposed at the national level. Four lines of activity (agro-ecological, technological, economic, and social), and ten specific observational actions have been defined. One such line is committed to gathering data about land-use/cover in the country by applying remote sensing techniques using SPOT 5 satellite images. The research products include field data reports, classified satellite images, the corresponding cartography, and statistical data. The products and data that were used in the project constitute a considerable volume. The data must be made available for use by all lines of research activity and also for determining the agricultural frontier of the country on a periodic basis. The data are not currently held in a single repository, so the magnitude of the bookkeeping endeavor is evident. A specialized geospatial database would be a boon for such a situation. To our knowledge, the kind of project described in the study we present here, namely, an informatics development to build a geospatial database, as would be needed to support comprehensive agricultural research, has not been attempted until now.

Modeling tools, software engineering practices and standards

The good practices of software engineering, the modeling elements and standards we use are known in the field. They are enlisted below:

Software engineering best practices [15]:

- Stakeholder communication
- Compliance with standards
- Administration of requirements
- Component-based architecture
• Iterative process realization and iterative implementation
• Visual abstractions
• Managed versions of specification documents.

Modeling Elements:
• UML - Unified Modeling Language [16]
• MADS - Modeling of Application Data with Spatio-temporal features [17]

Standards:

Software Process Framework

Regarding software processes, major considerations concern the tasks of user requirements analyses and of determining the demands of the external environment, while bearing in mind the policies of the organization or business.

Design and development

A general panorama of how geospatial data requirements can be viewed according to different kinds of user groups is schematized in Figure 3 [20]. The top part of the left side of the diagram indicates lower levels of detailed item descriptions. This is the case for common usage by general user groups, who exist outside the current operational environment. These external user groups do motivate our framework but to a lesser extent than do specialized users of our project in the short term. Such general user groups are interested mostly in standard-quality information, such as item identification attributes, location, and topology included in metadata standards. The bottom part of this same side indicates higher levels of detail as required by specialized groups. Individuals within these groups have different and more precise data information requirements according to intended use and to academic or professional profiles. In this case, clarification of the varied ways of how requirements are expressed plays an important part in arriving at consensus or compromises regarding the contents of a geospatial database. For example, workflows and item-derivation processes are involved, and thus describing them constitutes part of the formal requirements specification.

The right side of the diagram in Figure 3 refers to cases where organizations or other steering special interest groups set the requirement guidelines. Here, standard development objectives or fitness-for-immediate-use are the main quality criteria, as is the case, for instance, in the development of national data infrastructures. Such criteria contrast with the properties expected by certain groups, such as users of volunteered geographic information [21]. In cases where the purpose of a geodatabase, its context, and its user community are established, stakeholders should participate in order to arrive at a useful and sustainable repository. We agree with other authors that though somewhat neglected, user involvement is a basic good practice of software engineering [22]. We contend that this is of particular relevance when the objective of a database is to support research. The relevance is highlighted by Dick et al., who state that ‘without a relatively stable requirements base, a development project can only flounder’ [23], while Becker et al. maintain that for systems to be sustainable requirements, engineering is a major issue [24].

In the case of specialized geodatabases, which this paper addresses, user teams implement workflows in a variety of software systems relying on high-quality data. The workflows in terms of geospatial processes, ideally consist of a description of each process, the objectives of each process, input/output data, and the transformation of input data into output data. A software process specification consists of defining these parts, including the roles of people therein. Therefore, an analysis of workflows, though difficult, can provide basic information for a process specification framework that can be used to build a comprehensive geodatabase. This is the rationale for our work, in which the documentation and production of a geodatabase are supported.

Stakeholder specification process

The structure of the framework for the stakeholder specification process is summarized in Figure 4. The larger block ‘A’ within the dashed line refers to requirements and their transformation into specification documents. ‘A’ is the core component of requirements engineering and identifies the pertinent standards. The other phases of the framework for a software process include design, implementation, verification, validation, and maintenance. These phases are not explicitly referred to in Figure 4 but should be addressed considering the mission of a database in order to arrive at a comprehensive specification. Block ‘B’ represents the culmination of the specification process. We have considered not only an informatics technical perspective but also the social context that might be embraced by the use of the database, either because of the environment, which provides input data, or because of the social actors who will rely on derived information products. The following subsections describe the elements in Figure 4.

**Starred box:** Generally accepted good practices inside the starred box provide guidance applicable to specification steps and to the production of accompanying documents. Requirements are not written in stone; rather, they are subject to review in any of the phases of a software process. This is a characteristic of non-waterfall software development models that promote the practice of iterative process implementation. Other relevant practices that address the need for documentation and follow-up are stakeholder communication, administration of requirements, and component-based architecture.

**Box 1:** For stakeholder requirements to be formally administered, they are captured, analyzed, elicited, and documented. The Unified Modeling Language (UML) is the preferred tool for this. Arrow 182 indicates that requirements specification is itself a process that produces documents, as indicated in Box 2.

**Box 2:** Two documents constitute a specification: StRS (Stakeholder Requirements Specification) and SRS (Software Requirements...
Specification). These documents capture results to be revisited within the established life cycles of the development project. ISO/IEC/IEEE standard 15289 provides a mapping of documents to information items. This standard was under revision in 2017 [25]. Arrow 2◊3 refers to the architecture design process, for which component-based architecture is suggested. IEEE standard 42010 addresses architecture frameworks and description languages [26] and is currently under review in view of emerging proposals for different environments. Once the architecture process ends, new versions of StRS and SRS will be produced, as noted in Box 3.

Box 3: The main models in this element concern the design and implementation of the geospatial database repository and management software. The repository will contain a collection of complex multi-dimensional objects that possess temporal–geospatial and non-spatial data components. This repository can be perceived conceptually as an associative network where the associated software system manages links and paths among objects. A progression in design from conceptual to logical and physical data models is the basis for the implementation of a database (Figure 5).

The modeling tool MADS -Modeling of Application Data with Spatio-temporal features provides generic elements based on types. This is a convenient feature for geospatial databases. For instance, the types in MADS for defining objects explicitly address the heterogeneous and temporal nature of geospatial objects. Other modeling approaches exist and could certainly be used within the framework [27].

Results and Discussion

The basis for the implementation of the geospatial repository is the logical model, which prescribes the registration of identification, temporal and spatial location, lineage, textual or numeric components, and object mutations in both form and content. The conceptual model from which the logical model is derived, can be built hand-in-hand with MADS and UML use-cases derived from researcher interviews [28]. This conceptual model is in correspondence with a metadata model to prescribe what is required of an object prior to its insertion in the repository. The fundamentals for geographic metadata are included in ISO 19115-1:2014 [29]. In the GeoDB-CP project, two local metadata norms are also considered [30-32].

Block B of the MP-Geo framework indicates the end of a specification cycle has been attained and all is set for implementation activities. These will be guided by the chosen software development cycle, most likely iterative as recommended by recognized good practices. To further guide developing teams, and based on an example provided in [33], the scope and levels of the documents to be considered and produced for the environment in which the geospatial database system will operate are presented in Figure 6. The administrative level encompasses the operational level, which is the innermost level in the referred.

- The Systems Requirements Specification (SyRS) is the integral specification. Whereas SRS defines the software component in terms of its functional capabilities, SyRS describes the non-functional requirements and other technical specifications, including interfaces with external environments. These documents appear in the innermost level in Figure 6.

- The scope of the SRS document crosses all levels, as it embodies organizational policies, quality criteria, and the requirements of all user roles. As explained above, the framework contains stakeholder requirements and the data management environment as essential components. Final user groups express different information requirements according to their needs. The administration of updates is based on this document.

Appointed development teams will be aware of the new trends in software development and will take into account the reality identified by Fuggetta and Di Nitto that ‘activities are intrinsically dynamic and continuously evolving entities that cannot be frozen or defined once and for all’ [34]. An overall software development approach can handle this together with a component-based architecture supported by UML, visual abstractions, and iterative implementation. To this end, the UML-based specification documents will most likely exist in different versions and will be managed to maintain consistency. The ISO/IEC 27000 standard, currently under review, presents an overview of information security management systems and is applicable to all types and sizes of organizations [35].

Our work considers diverse forms of geospatial objects that are amenable to populating a comprehensive and shareable database. Different tools, such as geographical information systems or image analysis toolboxes in common use give rise to these kind of objects. Our approach does not propose to replace such tools but to accept that different users conceive, gather, transform, and register objects in different ways. For instance, due to the nature of research in remote sensing, objects exist in different digital forms and versions, and they are transformed, discarded, kept, or deleted during the course of analysis workflows. In our approach, the final annotated versions of objects to be inserted in a geospatial database to support scientific
research must fulfill quality requirements. This is different from initiatives where only the spatial properties of geographic data, such as their geometry, are considered. In our view, quality checks should consider whether objects are to conform to external standards such as a national data infrastructure or to other criteria set for an institutional database. Annotated forms include digital maps showing, for example, the displacements of the agricultural frontier over a defined period of time. An instance of a national project that would require just this has been presented in Section 2. It must be pointed out that researchers are not generally motivated to produce annotated forms of objects owing to the extra efforts needed to do this, disregarding that they might need to recall such objects in future projects. Software developers will enforce annotated forms of objects via default values or allowing users to pick from lists the pertinent annotations. These can include, for instance, selection of a named study area, thereby linking the objects therein for eventual sharing by other users working on the same geographical area.

Conclusion

Geospatial databases are a major component of scientific research into natural and social phenomena. There exists a general understanding of the content of such databases, and knowledge and technological tools for building these data repositories are continually improving. In spite of this, real day-to-day projects are challenged by conflicting interests about the breadth and scope of the contents of the repositories. It is therefore commonplace that researchers build anew what they require, even in cases where the study areas fall within the same geographical confines, and much data and information could be shared and reused. Considerable time, effort, and resources are consumed by repeating work already done, and projects are frequently hampered by limited access to storage and processing capabilities. Hardware, software, and knowledgeable human resources can be prohibitively expensive to develop computerized solutions, especially if time constraints are also present. However, in the medium and long term costs can be lowered and non-negligible returns on investments can be envisaged if concerted efforts towards a solution are considered. It is clear that arriving at a consensus in all cases with respect to the content of a database in terms of completeness, durability, fitness-for-use, and other properties is a difficult task, but one that can be made easier by adopting an agreed perspective. Given the above, MP-Geo, the software system framework we have presented, allows joint analyses and supports decisions concerning all stakeholders in specific contexts of design and development around a geospatial database. When interested parties agree on the relevance of relying on standards for building a common database, a substantial step towards reaching such a goal will have been taken. A software development team is then able to arrive at a system specification based on recognized best software engineering practices.

This MP-Geo framework has arisen from an academic environment where interdisciplinary research in the agricultural and natural science domains is promoted. Here, researchers and students analyze a variety of phenomena that are frequently related to common geographic areas. While we agree that even in a confined research environment a geospatial database cannot be of the one-size-fits-all kind, the work we have presented here results in a comprehensive specification for designing and building a shareable, useful repository, one that can potentially be adapted to changing situations. Such changing situations might include new groups of users (whether researchers or students), new projects, interest in other geographical areas encompassing those defined for previous projects, changes in official or local regulations applicable to data, different computational platforms, and updated versions of software components. The MP-Geo framework supports the production of a complete specification model for the design and implementation of a geospatial database and its management software. We think that the approach may be useful for other similar geospatial database projects. A salient feature is the incorporation of researcher requirements derived from use-cases that gather information about real experiences. From these use-cases, important information about the meaning of requirements is captured in a form that aids the integration of glossaries and, in the longer term, a form that can sustain building of ontologies and semantic queries to the database. Concerning our future work, we are proceeding towards an implementation of a geospatial database using the MP-Geo framework. A user-friendly interface will pre-insert objects to temporarily store them until checking of requirements marks them as acceptable in a shareable form. Metadata will register the status of objects so that users can make decisions about their usefulness. In due time, we will consider the potential economies of porting the geospatial database to the cloud.

Acknowledgment

The funds for covering open access publication costs were covered by College of Postgraduates and by the National Institute of Research for Forestry Agricultural and Livestock in Mexico.

References

30. www.inegi.org.mx/geo/contenidos/normasTecnicas