The Need and Development for Dynamic Integrated GIS Enhancement and Support Tools (DIGEST) – The Geospatial Project Management Tool (GeoProMT)

Chris S. Renschler and Laercio M. Namikawa
National Center for Geographic Information and Analysis (NCGIA)
University at Buffalo – The State University of New York
105 Wilkeson Quad, Buffalo, NY 14221, USA.
Tel.: +1-716-645-2722 x 23; Fax: +1-716-645-2329; Email: rensch@buffalo.edu

Abstract:
Historic GIS data and more recent Remote Sensing (RS) images are widely available at different scales from diverse sources. The intelligence, engineering, and science community frequently use these available, but often outdated or misinterpreted data sets to support their decision-making procedures. Even more important is the guided update and assessment of dynamic environmental processes in a landscape after a drastic environmental change, such as a wildfire, a volcano eruption, or an earthquake. Appropriate and effective use of the available data sources involves taking into account the data limitations and uncertainties the present and use latest RS imagery for real-time updates. Effective use depends on understanding the strengths and limitations of RS sensors and data processing methods, but also utilizing knowledge-based data fusion algorithms to cross validate various data sets in the context of the existing relationships of observed environmental properties and processes (e.g. one can take advantage of the tight relationship between elevation, topography, and other land surface properties with processes in hydrology, land cover/land use, etc.). Such additional knowledge allows minimizing related uncertainties and maximizing the confidence level of a particular data set for the desired application or decision-making process. This is especially important during a time of crisis when either natural or anthropogenic changes in requisites or account for sudden transformations of basic information in decisions made under time pressure (e.g. combat fighting forest fires, assessing flood risks, analyze earthquakes, landslides, volcanic eruptions, and other natural or man-made hazards). To overcome the limitations of the data due to uncertainties, the labeling of changes and critical regions within target areas of interest will give decision-makers valuable intelligence for better assessment of a particular situation and of possible scenarios. This paper outlines the conceptual framework for a collaborative effort of process modeling, GIScience, and RS engineering researchers to develop a series of Dynamic Integrated GIS Enhancement and Support Tools (DIGEST). The proposed tool platform integrates real-time remote sensing to detect changes, quantify the existing uncertainties, and enhance already existing multi-temporal geospatial data sets. The Geospatial Project Management Tool (GeoProMT) is a prototype that represents the core of DIGEST. GeoProMT allows multiple users to access a centralized repository of quality assured geo-spatial data such as GIS data and RS Images. GeoProMT is accessible through the Internet using an internet interface and is linked to the data repository to prepare appropriate data for environmental modeling and/or project management. DIGEST and GeoProMT will be assessed in the following ongoing research projects with federal agency involvement/sponsoring: Sensor development for fire detection and fire fighting, post-fire soil erosion assessment for forested areas, (both U.S. Forest Service), and assessment of volcanic hazards (NSF).

I. Introduction

Geographic Information Systems (GIS) data and Remote Sensing (RS) imagery are widely available at different scales from diverse sources. The data varies in format, scale, and in particular the method and time of gathering. Users often combine these multi-temporal datasets for decision-making purposes even if the data used was not gathered within the same period (Tab. 1). The effective use of these and other geospatial information depends not only on the tools to analyze the data by applying integration and transformation methods on them but also on the consideration of the uncertainties and limitations of the original data.

Table 1: Available Data by Type

Information	DEM/Topography	Land Use/Cover	Hydrography	RS Imagery
Source Data	• Field survey • Photogrammetry • Interferometry • LIDAR • Topographic Maps	• Classification of Landsat images	• TIGER database • FEMA Flood Data • Water Bodies • Hydrologic Units	• Landsat TM • Landsat ETM • National Aerial Photography Program (NAPP)
Collection Date	• 2000 (SRTM)	• 1990, 2000/02	*various*	*Historic/Actual Images*
Scale/ Resolution	• 3-arc-sec • 1 arc-sec • 1/3 arc-sec • 1/9 arc-sec	• 30 m	• 1:24,000 • 1:100,000 • 1:250,000	• 30 m • 15 m • 1 m
Format	• Binary • ARCgrid	• GeoTIFF	• Shapefile	• GeoTIFF
Available from	• USGS	• USDA-NRCS	• NOAA	*Variuos platforms*

II. Background

Efforts in the GIS and RS user community are mainly focused at data integration and transformation. Analysis of uncertainty and studies to determine the reliability of the methods applied to the data are not common and only rudimentary tools are included in GIS software packages. Intensive studies since the mid 1990s initiated by (Hunter and Goodchild 1995) provide a series of error estimates using probability theory. DEM data are the basis of a series of products, usually with the use of derived data, such as the slope and the aspect information. Therefore, many of the uncertainty studies deal with DEM. Uncertainty is known to be present in DEM (Hunter and Goodchild 1997), for example, the USGS DEM uncertainty is stated for the various scales, with the best defined to have a root mean squared error (RMSE) maximum of one-third of the contour interval (USGS 2005). However, this measure of uncertainty in USGS DEM is not of practical use as spatially distributed information or for mathematical modeling because the RSME is defined over a small number of about 30 sample points on distributed locations.

(Renschler et al. 2002) designed new RMSE and Model Efficiency (ME) filter values (MEFV) that provide spatially distributed measures of uncertainty raster data models based on multiple sources of elevation data. DEM from other sources, such as the one from the Shuttle Radar Topographic Mission (SRTM) are not better in modeling uncertainty, with their data specification demanding that 90% of the points lie within 16 meters accuracy in the vertical dimension and 20 meters in the horizontal plane (Kretsch 2000). In addition, the currently released data often has data gaps and inconsistencies.

During a time of crisis when decision-makers must rely on this potentially uncertain data, the knowledge of the reliability of the data concerning the area of interest will help achieve a better assessment. The definition of the conditions for a data set to be adequate for input in a given model are not based only on uncertainty and but also on the scale of the used data. In the case of the elevation data, scale is related to the effective resolution of the DEM. In hydrological applications, several studies have demonstrated that DEM resolution plays a major role (Hardy et al. 1999; Horritt and Bates 2001; McMaster 2002). One method to define the adequate scale/resolution for a given problem is to use different scale/resolution combination and select from one of them (Brasington and Richards 1998; Horritt and Bates 2001).

Given that the adequate scale and quality for the data to be used for the assessment of a particular phenomenon is known, if data characteristics do not fulfill the minimum requirements for a proper analysis, methods to enhance a particular data set will be crucial. Enhancement of a data set can be achieved by using additional information from other sources, either a pre-existing easily available dataset or a specially gathered dataset. The integration process is dependent on the georeferencing of the newly available data and synergy with the main data. Georeferencing is dependent on the correct transformation between the reference systems of the main and the new data. The widely used World Geodetic System 1984 (WGS84) reference system relies on controls points common to the other reference system to minimize differences (Kumar 1988).

The correctly referenced additional data can be integrated into the existing data through conflation methods and it can be available in different formats. The first separation for a topographic data is based on the existence of height information. Data without height information can be used to position characteristic features of the terrain such as peaks and pits points or ridge and valley lines, and require some procedure to define an estimate of their heights (Namikawa 1997). Integration of data with height information in punctual, linear, and rectangular grid format should be done considering the relative reliability of the different sources (Kyriakidis et al. 1999).

The additional information from this latest imagery can be used to improve topographic data quality if the images are correctly georeferenced. Imagery can be georeferenced through automatic registration techniques (Fonseca and Manjunath 1996) using wavelets or contour matching approaches (Fedorov et al. 2003). From these images, linear features that will help improve data quality can be extracted through either segmentation (Munoz et al. 2003) or mathematical morphology techniques (Candeias 1996).

The integration in a georeferenced database of data, its quality information, scale of the event to which the data is more suitable, and methods to improve quality by incorporating additional information will provide a reliable platform for simulations required for the decision-making process.

III. Dynamic Integrated GIS Enhancement and Support Tools (DIGEST)

The main objective is the development of a series of Dynamic Integrated GIS Enhancement and Support Tools (DIGEST) to assist intelligence analysts in the use of geospatial data by providing the most fit-for-use data by reducing inaccuracies of existing data “on-the-fly” thus providing an effective and reliable environment for crisis management. The DIGEST system will be integrated in an existing geographic database, with the tools targeted to create quality information for the available data, and to define the appropriate scale required for an event simulation based on the limitations of existing geospatial data. The tools will also dynamically extract features of relevance from the latest non-photogrammetric visible, infrared and thermal imagery, detect changes to features in the already available geospatial data, and integrate additional information through data conflation algorithms for quality enhancement of the data based on the desired scale of interest. A typical data flow (Figure 1) will begin with the analysis of existing geospatial data to create existing data quality information and provide data scale suitability information. Based on the most recent gathered imagery, relevant features will be extracted and integrated in order to determine and enhance the quality of the preexisting data and detect changes.

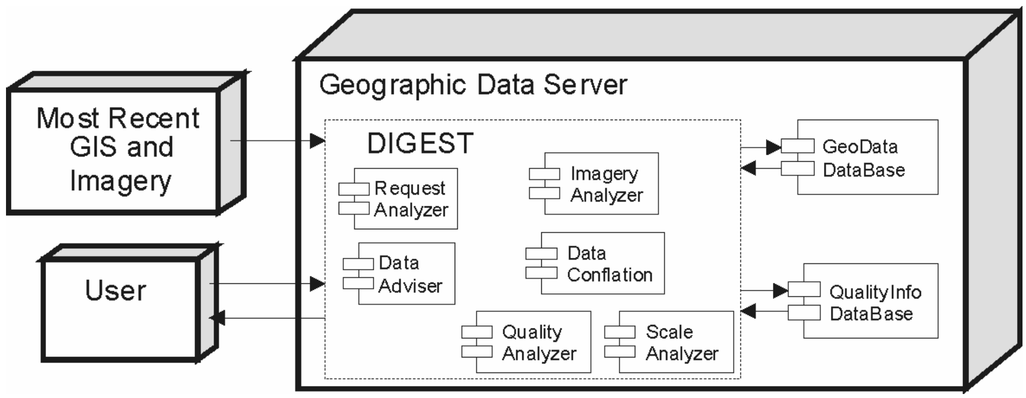

Figure 1. Proposed dynamic DIGEST system integrated in a Geographic Data Server.

DIGEST will be integrated in a Geographic Data Server that stores the available GIS data and RS imagery in a Geographic Data Database (GeoDataDB) (Figure 1). The data and images will have its quality information attached as defined by the Quality Analyzer Tool (QAT). The quality information will be stored in the Quality Information Database (QualityInfoDB). The Scale Analyzer Tool (SAT) will create a knowledge database built from a series of simulations on samples of stored geographic data in different scales. Request Analyzer Tool (RAT) and Data Adviser Tool (DAT) will process user’s data request and suggest acquisition of new data if there is no data in GeoDataDB that meets the quality requirements. Imagery Analyzer Tool (IAT) will register the real-time imagery and extract relevant geospatial features. Data Conflation Tool (DCT) will be responsible for integrating data from imagery and GIS data to existing entries in GeoDataDB. The DIGEST system will be built according to existing industry standards and can be linked to other geospatial analysis systems. The foundation of DIGEST is based on a geospatial project management component that is currently implemented in research and teaching in natural hazard and natural resources management projects at the University at Buffalo (UB) – The State University of New York (SUNY), New York.

IV. Geospatial Project Management Tool (GeoProMT)

The Geo-spatial Project Management Tool (GeoProMT) accesses a centralized repository of quality assured geo-spatial data such as GIS and Remote Sensing Images. GeoProMT is accessible through the Internet using a WEB interface and is linked to the data repository to prepare appropriate data for environmental modeling and/or project management. Integral to GeoProMT framework is role-based access control (RBAC), where data access permissions are associated with roles, and data users are assigned to appropriate roles, enabling efficient collaboration among participants of a large interdisciplinary geo-spatial project. The centralization aspect of GeoProMT permits data quality control, data access control, data usage tracking, and supervision of process steps, while the accessibility of GeoProMT on the Internet provides decentralization of computational resources. By assigning roles to the users of a data set, GeoProMT promotes efficient use of a project personnel expertise, given that each user could contribute only on steps where one's knowledge is the maximum within project participants, and that he/she does not need to be concerned about the others steps of the process. The various steps are a serious of data processing steps in the digital domain that are required in any geospatial management project (Figure 2; Renschler 2003). These essential steps are various levels of digital representations of natural properties and processes in space and time (Renschler and Harbor 2002).

In a course environment we simulate then project conditions: different roles are given to the instructor, the teaching assistant, and the students. The instructor is responsible for setting the project parameters or learning environment by defining the scope of the natural hazard or natural resources management projects. The instructor and the teaching assistant are assigned to control and monitor the technical issues of each project. The teaching assistant role defines the students that are members of the project and also handles the tasks associated with data quality issues. Only read access is given to quality approved data for a user in the role of a project manager of a consecutive data processing step.

The target course for GeoProMT is the Geography course Geo575 titled "Landscape Modeling with GIS". This course provides introduction to concepts, theories and applications of geo-spatial analysis and modeling tools in Geographic Information Science (GIScience). The interdisciplinary audience consists of graduate students in Geography, Geology, Civil Engineering, Chemistry, Biology, Planning or Environmental Studies, Business Administration and Management Science. There is a prerequisite of an intro course to GIS or by permission by instructor. The course contains exercises in a computer laboratory aimed to test models of surface/terrain spatial-temporal dynamics fluxes due to concentrations or costs gradients in space and time, using data collected from various sources. Since the learning experience must be based on realistic models, implying that the computational models of fluxes should be as similar as possible to real world fluxes, and handling of data is the weakest link, GeoProMT provides hand-on to approaches in data management, theory, concepts and a is practical tool to obtain the data set that is most fit-for-use in the simulation models.

PROCESS SCALE	TRUE PROCESS SCALE AND VARIANCE
⇓ *BASIC SCALING* ⇓	⇓ **MEASURING** ⇓
MEASUREMENT SCALE	*Observation Unit (measurement device)*
⇓ *1^{ST} SCALING* ⇓	⇓ **PRE-PROCESSING** ⇓
DATABASE SCALE	*Common Database Unit (data availability)*
⇓ *2^{ND} SCALING* ⇓	⇓ **DISCRETIZATION** ⇓
MODELING SCALE	*Modeling Unit (model requirements)*
⇓ *3^{RD} SCALING* ⇓	⇓ **MODELING** ⇓
PREDICTION SCALE	*Prediction Unit (model design)*
⇓ *4^{TH} SCALING* ⇓	⇓ **POST-PROCESSING** ⇓
ASSESSMENT SCALE	*Scale of Interest (user requirements)*
⇓ *5^{TH} SCALING* ⇓	⇓ **EVALUATING** ⇓
VALIDATION / MEASUREMENT SCALE	*Observation Unit (measurement device)*
⇑ *BASIC SCALING* ⇑	⇑ **MEASURING** ⇑
PROCESS SCALE	TRUE PROCESS SCALE AND VARIANCE

Figure 2: Steps for transformation of information (scaling steps) (after Renschler, 2003) in GeoProMT. Note that scaling and evaluation through scaling requires a transformation of information (=>) and within the domain of digital geo-spatial data handling (^^^).

In a research oriented setting, GeoProMT targets management of data in multidisciplinary, multi-site, and multi-application environment. Such environment requires the existence of participants in supervisor, project manager, and project member roles, and the definition of friendship among projects. The supervisor is the technical administrator, responsible for creating the projects and assigning managers for each project. Project managers define its members, control data quality, and establish friendship relations. A project member has read access to the project data and to other projects that granted friendship, thus providing the sharing of data for different applications. A combination of education and research environment is the North American Earth Hazards Consortium (EHaz) of six universities in Canada, Mexico, and the U.S. (incl. UB). The students and faculty of the consortium collect, manipulate, generate, and exchange geo-spatial data. GeoProMT will be essential to provide an orderly handling of information for the many different data users, located in separated places, working in different applications for the same data set or collect data for a particular project site. The environment at the targeted UB Center for Geohazard Studies will be similar, with interdisciplinary group of faculty members accessing data for modeling of landslides, mudflows, volcanic flows, avalanches, and surface or subsurface water flows, and GeoProMT providing the information handling for the data users from the various disciplines, with their expertise.

V. References

Brasington, J. and K. Richards (1998). "Interactions between model predictions, parameters and DTM scales for topmodel." Computers and Geosciences **24**(4): 299-314.

Candeias, A. L. B. (1996). Drainage Network Extraction from a SAREX'92 RADAR image. Simposio Brasileiro de Sensoriamento Remoto, Salvador, Brazil, Instituto Nacional de Pesquisas Espacias.

Fedorov, D., L. M. G. Fonseca, C. Kenney and B. S. Manjunath (2003). Automatic Registration and Mosaicking System for Remotely Sensed Imagery. Simposio Brasileiro de Sensoriamento Remoto, Belo Horizonte, Brazil, Instituto Nacional de Pesquisas Espaciais.

Fonseca, L. M. G. and B. S. Manjunath (1996). "Registration techniques for multisensor remotely sensed imagery." Photogrammetric Engineering and Remote Sensing **62**(9): 1049-1056.

Hardy, R. J., P. D. Bates and M. G. Anderson (1999). "The importance of spatial resolution in hydraulic models for floodplain environments." Journal of Hydrology **216**(1-2): 124-136.

Horritt, M. S. and P. D. Bates (2001). "Effects of spatial resolution on a raster based model of flood flow." Journal of Hydrology **253**(1-4): 239-249.

Hunter, G. J. and M. F. Goodchild (1995). "Dealing with Error in Spatial Databases - a Simple Case-Study." Photogrammetric Engineering and Remote Sensing **61**(5): 529-537.

Hunter, G. J. and M. F. Goodchild (1997). "Modeling the uncertainty of slope and aspect estimates derived from spatial databases." Geographical Analysis **29**(1): 35-49.

Kretsch, J. L. (2000). Shuttle radar topography mission overview. Applied Imagery Pattern Recognition Workshop.

Kumar, M. (1988). "World Geodetic System 1984 - a Modern and Accurate Global Reference Frame." Marine Geodesy **12**(2): 117-126.

Kyriakidis, P. C., A. M. Shortridge and M. F. Goodchild (1999). "Geostatistics for conflation and accuracy assessment of digital elevation models." International Journal of Geographical Information Science **13**(7): 677-707.

McMaster, K. J. (2002). "Effects of digital elevation model resolution on derived stream network positions." WATER RESOURCES RESEARCH **38**(4).

Munoz, X., J. Freixenet, X. Cufi and J. Marti (2003). "Strategies for image segmentation combining region and boundary information." Pattern Recognition Letters **24**(1-3): 375-392.

Renschler, C. S. (2003). "Designing geo-spatial interfaces to scale process models: the GeoWEPP approach." Hydrological Processes **17**(5): 1005-1017.

Renschler, C. S., D. C. Flanagan, B. A. Engel, L. A. Kramer and K. A. Sudduth (2002). "Site-specific decision-making based on RTK GPS survey and six alternative elevation data sources: Watershed topography and delineation." Transactions of the Asae **45**(6): 1883-1895.

Renschler, C. S. and J. Harbor (2002). "Soil erosion assessment tools from point to regional scales-the role of geomorphologists in land management research and implementation." Geomorphology **47**(2-4): 189-209.

USGS (2005). USGS Digital Elevation Model Data. http://edc.usgs.gov/glis/hyper/guide/usgs_dem, accessed on 03/12/2005.