Knowledge Discovery, Thematic Maps, and e-Discussion as Means for a Sustainable Decision Making in Developing Countries

Wolf Göhring, Stefanie Roeder, Volker Rudolph, Stefan Salz, Alexandr Savinov, Simon Scheider¹ and Cornelio Hopmann²

Keywords:
Knowledge discovery, spatial data analysis, geographical information system, online mediation, on-line discussion, decision-making, sustainable development, North-South technology and knowledge transfer, bridging the Digital Divide, poverty alleviation, Development Gateway.

1. The Spatial Distribution of Economic, Social and Environmental Factors

How can one integrate “the economic, social and environmental factors at the policy, planning and management levels” (Local Agenda 21, chap. 8.2), especially in a least developed country like Nicaragua? We present 3 tools, which help the different social actors to realize this task and to find common views on their problems. The tools support the efficiency and sustainability of development oriented decisions. Decisions on public investments like water, travel, communication, energy, environmental protection, waste, health, and education interfere with private ones in plants and buildings, with economic activities, the development of labor market, poverty alleviation, etc. Local and regional aspects, i.e. the spatial distribution of the mentioned factors, influence the decisions. Intended decisions have to be discussed in order to build consensus. In developing countries organizations of civil society play an important role as intermediates between the great majority of the people and the digital media which are accessible or usable for a few people only.

² Del PALI 4e al lago, 2c abajo, Lindavista, Managua, Nicaragua. Phone/Fax: +505 268 2888, cornelio@enicaragua.org.ni, http://www.enicaragua.org.ni.
2. Nicaragua: a Least Developed Country in the Tropics

Least Developed Countries in tropical zones present specific challenges for development. Local conditions, climate e.g., may vary drastically even between nearby places. The lack of infrastructure – from transport to housing, from energy to communications, from education to health, from organized municipalities to global law-enforcement and public administration – adds complexity (Arce/Hopmann 2002). One has to take into account the spatial distribution of those data (Figure 1). Identifying local factors for success or failure of development projects becomes crucial for effective development. One needs data concerning geographical structures, socio-economic and environmental factors down to the municipalities including type of buildings, type of farmland, crop areas, forests, water, rainfall, water flood areas, wetlands, epicentres of earthquakes, volcanoes, landslip threads, natural catastrophes, but also crimes, gender, income, the distribution of educational facilities and language differences, and time series of such data. One needs tools to visualize and to analyze such facts and their geographical distribution.

In Nicaragua, 5.2 million people live on 130,000 km². Some regions in the East have no road connections with the West of the country. It counts more than 600 vol-

![Figure 1: Nicaragua’s uneven distribution of infrastructure: Main roads, town centers, and ICT infrastructure (phones, mobiles, television).](Data source: TELCOR Nicaragua)
canoes with high risks. Since 1973 have occured 150 earthquakes with more than 5 mb G. Nicaragua’s balance of trade is negative though emigrants support their families with credit transfers. Dept service amounts to 35 % of the national budget. About 85 % of its population are poors with less than 1.5 $ income per capita and day leaving no room for extra consumption. Extreme poverty, i.e. people are suffering from famine because a lack of appropriate incomes, amounts to 30 % of the population. 54 % of it are not elder than 19 years. Population density varies extremely: in Managua it reaches thousand or more people per km², in rural areas it may drop down to 2 or less per km².

In extended areas agriculture is based on burning down the forests with a rapid degradation of soils yet the dwellers lack the knowledge about better methods. Specialized training could increase production and quality. Functional illiteracy reaches 50 %. The average grade achieved in public schools corresponds to 4th grade of primary education. Though the number of students in public education almost doubled during the last 15 years, still about 28% of the respective age groups don’t attend school at all. As contrast, there are about 11,000 university-students in computing or related subjects and about 4,000 professionals with an university-degree. In rural areas the access to a phone may be measured in a walk of 8 or more hours – one way! In contrast, in Managua and other larges cities there exist more than 500 telecenters with qualified personnel (university degree in computer science).

If such differences are not treated appropriately, they may cause serious conflicts. In Wiwili e.g., a municipality in the West, arose a sharp conflict when one part of the town got nearly almost of the aid after the hurricane Mitch while the other, on the opposite riverbank, got almost none. The conflict was solved by subdividing Wiwili into two municipalities each responsible for its negotiations with the government. Managua, the capital, is crossed by a dozen earth-faults. Urban planning depends hence on spatially differenced risks for the population. 100 meters of spatial distance may triple construction costs.

Installing a telecenter e.g. depends on the goals one has to reach: If there is a short-term need to refinance the investment and to accumulate stocks for further investment then the telecenter with Internet access has to be installed where the purchase power of the surrounding area will support such a center. This contrasts to a long-term strategy where one would seek to facilitate access to communication for poor rural farmers. Here a multipurpose-communication-center, which combines traditional phone-services with digital services, might be more adequate. Here, most people have plenty of emerging needs, such as improving their housing and education. Telecommunications needs are not at the top of the list, so investment in Internet access has to bring concrete benefits, particularly in the short term. It is important to offer users of a telecenter a vision for development. (Mannila 2003 6, 43). To achieve this vision, a telecenter needs appropriate local operators – like local non-governmental organizations, cooperatives or other community-groups.
3. The Need for an Analytical GIS-Tool

The introduction shows the difficulties when one tries to alleviate poverty by means of information and communication technologies (ICTs) as envisaged in the 8th Millennium Development Goal. In Nicaragua the presidential secretaries for strategic planning and coordination, the National Institute for Territorial Studies and the National Institute for Census and Statistics are in charge to monitor and evaluate the results of the Nicaraguan Poverty Reduction and Economic Growth Strategy, convened by Nicaragua and the donor community as part of the debt reduction of the Heavily Indebted Poor Countries HIPC. In most cases this task implies a spatio-temporal analysis of socio-economic and environmental data, not only the visualization of bare connectivity or availability of grid electricity e.g.

Actually there are lots of geo-referenced data available in Nicaragua but they are managed by different institutions, using different formats and maps with different scales and projections, different levels of aggregation, and different zoning of regions. Hence, in Nicaragua as in many other places, Geographical Information Systems GIS have been introduced to present the complexity geographically. Yet most of these tools permit only an overlaid presentation of localized facts as in figure 1, or – when used as planning tools – of projected localized objectives and their constraints. To become an active part of the planning process, tools are needed that permit the analysis of different scenarios with data from different sources – what-if type of analysis – and the dynamic comparison of projected outcomes between different regions.

The Fraunhofer Institute for Autonomous Intelligent Systems AIS has started a project in cooperation with the Asociación Internet de Nicaragua AIN and its operating entity eNicaragua. AIS provides the analytical GIS-tool CommonGIS together with a pilot demonstrator. CommonGIS enables an efficient, highly interactive, exploratory spatial analysis of attributed geographical data even for less experienced users. It allows to visualize and to manipulate the spatially differentiated results. Its features are intuitive and easy to learn. CommonGIS is Web-enabled and implemented in Java. Using CommonGIS one will reap the benefits of a longstanding experience with applications of this tool (and its predecessors) in Western Europa.

AIN is a not-for-profit association joining most of the Internet-Service-Providers (80% of traffic) and the four public Universities in Nicaragua. By grant of the Development Gateway Foundation DGF (Worldbank Group, http://www.dgfoundation.org, http://www.developmentgateway.org/, July 27, 2004) eNicaragua is going to implement the Nicaragua Development Gateway niDG. Similar to the Gateways in other countries, the niDG aims to act as a national clearinghouse for information about development and development projects in Nicaragua, promoting the collaboration among local actors themselves and their counterparts from abroad. The niDG intends to improve efficiency and effectiveness of development projects by providing knowledge-management tools, which enable active
leverage of the accumulated information from planning through monitoring while implementing until final impact evaluation (project-cycle). It will provide basic information, tools and methodologies, which each actor then may apply according to his needs and objectives. As example, it will provide information about ICT usage in education and by businesses, but also the basic educational and economic statistics for each municipality. The niDG will use the so-called DiGi-platform of the DGF. On this platform the DGF gathers on international scale information about development from almost all development-agencies.

4. **Overview on CommonGIS**

The combination of visualization and data analysis in CommonGIS empowers common people to make full use of geographical knowledge. CommonGIS provides a better basis for adequate decisions and later for the evaluation of the real success of which. CommonGIS makes the related geo-data including time-series interactively usable for local, regional or country authorities, for people in enterprises, NGOs, cooperatives, trade unions or other people trained in using ICT, and from each access point to the Internet. The automatic generation of thematic maps unburdens the user from complex mapcreation tasks. CommonGIS can be offered in a Web environment in a reduced and customized way. ([http://www.commongis.com/html/viewlets/advertising/advertisingcgis_viewlet.html](http://www.commongis.com/html/viewlets/advertising/advertisingcgis_viewlet.html)) Equipped with laptop and beamer one is able to demonstrate online and interactively the analysis of geo-referenced data. Especially a functionally illiterate auditorium as one meet it in many places in a least developed country like Nicaragua could have a realistic insight into statistical relations which could not be given through written text or endless tables. The auditorium could better understand its situation – presupposed the demonstration is mediated by trained people. The tool helps to discuss alternative solutions of planning questions:

- Where’s the best place to locate a factory needing clean water, near roads, close to raw materials and electricity?
- How to assess soil erosion and its impact on production? What is the best place to locate a tourist hotel?
- Where should a city place a sewer treatment system? A hospital? A school? Where to locate an industry and protect the environment? (CGIS 2004)

Translated into the philosophie of CommonGIS such tasks comply with the functionality of CommonGIS (Andrienko/Andrienko 1999, Andrienko/Andrienko 2001, Andrienko/Andrienko/Voss 2003) of which we give an incomplete overview:

- Effective multidimensional description of spatial objects: “How can the considered space be characterized by a set of attributes?”
• Multidimensional query and search for spatial objects: “Where can one find a special type of space?”

• Multivariate spatial analysis (e.g. dominance analysis): “Where are regions with exceptional attribute properties in comparison to all other properties and regions?”

• Spatial hypothesis finding: “Is there a spatial/non-spatial relation between attributes?”

• Multicriteria decision support: “Where is the best region according to several weighted criteria, and how sensitive is the decision when criteria weights are changed?”

5. Applying CommonGIS in Nicaragua

The first steps of the project concern a few of the named aspects only. AIS had to inspect a lot of files delivered by the partner in Nicaragua. The huge amount of files was due to a high redundancy of the scattered, inconsistent and mostly undocumented data (different zoning, different ordering, varying data types and projections). This scattered and incomplete information was sufficient when scattered institutions used only their small slice of data. The project implements a “demonstrator” which may serve as a proof-of-concept for those state-agencies in Nicaragua, that already have a stock of geo-referenced data, yet don’t use any analytic or visual exploratory GIS-based tools.

The demonstrator, first with an English, later with a Spanish user interface, bundles some of the given data as “projects” in CommonGIS, namely:

• Geo-data like maps of Nicaragua, the zoning in regions and municipalities; rivers, roads, volcanoes, earthquake epicenters; average rainfall & temperature; telecommunication lines and nodes, television lines and nodes.

• Attribute data like population (1995), poors, extreme poors; communication capacity & users; voters (1996), invalid votes, resulting votes; risk of volcanoes (high/medium); intensity, date & location of earthquakes; technical data on telecommunication facilities.

Some of these data allow only overlay visualization as in figure 1. Others are analyzed using the features of CommonGIS as shown in the examples of figures 2 to 5. There, we first search for a spatial hypothesis explaining the distribution of poverty in Nicaragua (< 1.5 $ per capita and day). CommonGIS computes the percentage of poors w.r.t. population. Figure 2 shows the result as a “dynamic classification” where one assigns interactively value classes to value intervals, shown on the dot plot diagram in the upper part of the right-hand manipulation window. One sees the value distribution together with the (coloured) class breaks. CommonGIS automatically distributes the value classes according to the value distribution with least in-
formation loss; the user sees a thematic map: Poverty is non-equally distributed from the West to the poorer East (the poorer the region, the darker its colour) (Data source: TELCOR). In general, the Nicaraguan cordilleras and the East coast are poorest, Managua is richest. Some East coast regions are exceptionally less poor than the typical East cost. Figure 1 shows in these regions (Puerto Cabezas, Bonanza, Rosita) a developed communication infrastructure.

Figure 2: Analysis of poverty distribution in Nicaragua

Figure 3: Analysis of communication capacity distribution in Nicaragua

Is this observation significant? Figure 3 shows in a second window the distribution of communication capacity (the darker the color, the higher the capacity) which is similar to the poverty distribution: the mentioned East coast areas have an exceptionally high communication capacity. Does a statistical relation between infrastructure and poverty exist throughout Nicaragua? Can one find exceptions from this...
rule? How to detect these outliers? Where are they on the map? The interactive linking and brushing technique of CommonGIS helps to answer these questions. Figure 4 shows a correlation diagram of the attributes “poors in % of population” and “communication capacity”. There is a medium negative correlation coefficient (-0.56) with some outliers highlighted (here as black circles) by mouse click. The interactive display technique immediately highlights all corresponding regions on the map in figure 3. Managua itself is an outlier to this relation. It has high poverty in relation to its infrastructural equipment. The histogram of the right-hand manipulation window in figure 3 shows most of these outliers in regions with middle or low poverty. They are located around the lake Nicaragua. (Data source: TELCOR)

In a third example we use again multidimensional analysis. Figure 5 shows in the left the result of the 1996 election. The conservative PC won very few regions (medium grey), the liberal party PLC dominates the election results. Small parties did not win any region. The mountainous regions, some urban regions and the North East were won by the sandinista FSLN. What can be discovered from the electoral statistics?

Figure 4: Correlation between poverty and communication capacity

CommonGIS can classify regions by dominant attributes, which are statistically outstanding attribute characteristics compared to all other attributes (by mean, median, quantiles or standard deviation e.g.). Figure 5 shows on the right an interactive dominance map of the electoral statistics. Here, median-quantile dominance was calculated over the number of votes for the 3 major parties plus “others” in each region. The histogram on the right shows, that those regions where the PC dominates
are the most frequent (white), even though they did not win any of these regions. The PC doesn’t show any spatial pattern except around the lake Nicaragua, were it clusters. The mountainous regions in the North did vote mostly for the FSLN and the PLC. The most striking spatial pattern is the distribution of minor party voters (“others”, dark), which is the second most frequent dominance class. Obviously in the Eastern poor parts of Nicaragua, people vote exceptionally frequent for minor parties in comparison to all other regions.

Figure 5

*Left:* Winners of the 1996 election

*Right:* Interactive dominance classification of the 1996 election

6. **Data Mining and CommonGIS**

The examples given above indicate the effective extraction of geographical knowledge from statistical data through CommonGIS. We consider now another kind of problem: In some places in Nicaragua one introduced the concept of backyards in order to provide people with fresh vegetables. This concept failed somewhere, elsewhere it succeeded. Are there hidden reasons for this paradoxical result? Could one find an explanation through mining in the data on the socio-economic context?

Data mining is the partially automated search for hidden patterns in typically large and multi-dimensional databases. So far, data mining and GIS have existed as two separate technologies. Recently, their integration has become attractive as various organizations possessing huge databases began to realize the potential of information hidden there.

SPIN!, a product of AIS, integrates an analytical state-of-the-art GIS, namely CommonGIS for interactive visual data exploration and data mining functionality in a closely coupled, open, and extensible system (May/Savinov 2002, 2004). The data
mining functionality is adapted for spatial data. A spatial database is used to execute
the spatial queries generated by the analysis algorithms. The final system integrates
data mining methods for spatial data such as multi-relational subgroup discovery,
rule induction and spatial cluster analysis with the interactive functionality of visual
data exploration, thus offering an integrated, distributed environment for spatial data
analysis. SPIN! links cartographic and non-cartographic displays together through
simultaneous dynamic highlighting of the corresponding parts. The SPIN! platform
connects data mining to an analytical GIS like CommonGIS whose strengths lies in
the dynamic on-line analysis of spatial statistical data. As the user navigates in the
list of spatial rules, the corresponding geographic objects in the map window are dy-
namically highlighted. With the simultaneous visualization of the results one can
prepare non-trivial decisions. Decision makers may back up their intuitive insights
by sound statistics, and automatically explore patterns in the data that are invisible to
the eye because they live in high-dimensional spaces.

Figure 6: Visualization of spatial rules:
Map and other views in SPIN! are shown simultaneously and interactively.

Figure 6 shows an analysis applied to UK 1991 census data for Stockport, a dis-
trict in Greater Manchester, UK. The analysis was carried out at the level of enu-
meration districts (the lowest level of aggregation) characterized by such attributes
like persons and cars per household, migration, long-term illness, unemployment.
Spatial data were given as coordinates and borders of objects like enumeration dis-
tricts, water, roads, streets, railways, and bus stops. In the shown example the spatial
rule induction algorithm (Savinov 2003) finds a relationship between districts with high migration rate and high long-term illness.

The mentioned paradoxical result for the backyards attempt in Nicaragua could be analyzed in a similar way. A tool like SPIN! supports decision makers to retrieve the efficiency of development projects and the reasons for success and failure.

SPIN! is Java based. It has a plug-in architecture so that data mining algorithms, data access and visualization components easily can be integrated. The user can manipulate various types of components via easy to use graph view where they can be connected according to the task to be solved.

7. **E-Participation: Remote Discussions of Analytical Results from CommonGIS and SPIN!**

In the previous examples the spatial dimension plays a substantial role in understanding given phenomena and forecasting possible scenarios. Yet, analytical GIS alone cannot solve all the problems in spatial planning processes whose complexity is frequently driven by the lack of representation. The decision makers are expected to decide on the basis of various information sources (tables, maps, debates and recitations), considering the values and interests of individuals and groups involved. Yet, it is almost impossible to gather the multitude of stakeholders at the same time and place. Due to time constraints not all interests can be heard and not all aspects be discussed elaborately. As a consequence, the ideas and suggestions of some citizens and stakeholders can be lost to the process, making it both less informed and more likely not to be accepted.

Although electronic media might not fully replace face-to-face meetings, the well organized use of ICT can help to enhance the participation of citizens and to achieve a balanced representation of people’s values and concerns, e.g. by representing analytical results appropriately and by supporting discourses between citizens and authorities. The integration of an analytical GIS with tools for e-participation yields the infrastructure for decision processes (Campagna/Deplano 2002). In doing so, throughout the discussion manifold perspectives can be considered and a broad spectrum of alternative solutions be highlighted. Conflicts may be mitigated already in preliminary stages of the decision-making. Thus, consensus based solutions which all participants can identify with, will crystallize almost inevitably. The decision making would become more effective, efficient, and transparent.

**DITO** ("as said") is an Internet-based public e-participation platform, which has been developed by Fraunhofer AIS, and it offers means for complex, text-based discussions (Voss/Schäfer). **DITO** is used for self-guided or *moderated* collaboration between persons that are distributed in time and space, making the discussion independent of temporal and spatial constraints. The written record of the contents and the structure of the communication make the whole process more comprehensible.
and resulting decisions more transparent to all participants. In a developing country, DITO could aid to create a participative setting regardless of the spatial distances.

DITO features user and group management, protected communication sections for structured discourses, and process control instruments for the moderator, as well as support for discourse awareness and voting. DITO sections contain networks of articles, representing, for example, discussions, rationales or idea maps. The articles and their links can be labelled according to configurable ontologies. The articles may have a title, a note, attached documents and links to resources on the web; they can be topics or responses.

DITO has been used for different application areas already. The largest public online participation so far was the idea contest about the future development of the city of Hamburg, Germany. Over 8,000 people read the 4,000 contributions made by over 500 participants resulting in 1,400,000 accesses in four weeks. DITO has also been applied to urban planning projects, e.g. for the Berlin Alexanderplatz, Germany. The distribution of resources and facilities has been the issue of a project in Esslingen, Germany, where the citizens discussed with the local government the options of spending the city’s budget (http://www.e-partizipation.org/).

Figure 7: Integration of DITO with CommonGIS and SPIN!

DITO can be used to discuss the spatial aspects, which have been explored via CommonGIS. Representatives of several communities or countries can submit comments or ideas directly referenced to the regarding spatial area or location. Or, representatives of different regions in a country can choose a moderator and, on-line, they can discuss possible decisions. Civil society may also participate, through digital media where these are accessible, through radio, television, printed media, and in
local assemblies. In a developing country with a high percentage of illiterates trained people would apply DITO and would act as mediators to their fellow citizens.

AIS is planning an integration of CommonGIS and DITO. Maps can then become the object of discussions simply by putting a flag at a point of interest and adding a contribution about it. This contribution can be discussed with all facilitation needed and can further reference other points of interest in the map. The points of most frequent discussion can be automatically highlighted in the map.

8. Final Remarks

This paper describes some steps of sharing tools between developed and developing countries. Additionally it shows thematic areas and ways of sharing and analyzing data within a developing country. The general aim of this approach is poverty alleviation. It fits into the 8th Millenium Goal of the United Nations: “develop a global partnership for development”.

In Nicaragua as an example country and in a first step, the analytical GIS-tool CommonGIS will be applied to map the Digital Divide inside the country. This application should show the usefulness and efficiency of this tool; people will be trained in its use. As a side-effect the example will show which kind of quality assurance of the collected and stored data is needed. The partners AIS and eNicaragua/AIN will continue to co-operate and to extend the application to environmental aspects also. The global network established by the Country Development Gateways at their meeting in Bonn in June 2004 (http://topics.developmentgateway.org/cgn, July 27, 2004) may help to disseminate the presented approach. Though different in many details the hosting countries have the same kind of problems with respect to spatio-temporal decision-making. From Bolivia, Guatemala, Kenya, Mongolia, Palestine, Romania, Rwanda to Venezuela and Vietnam there exist interests to apply an analytical GIS-tool. Additionally these countries have a need for knowledge discovery tools. E-participation is driven by a lack of transportation-means, which impedes face-to-face meetings. Now, we come full circle: The donor countries as well should be interested in the usage of such a modell (Commission 2004). It offers a chance for more integrated approaches to the environmental, economic and social tasks. It helps to better adjust the investments in developing countries to regional or even local conditions that supports sustainability.

References


Campagna, M., Deplano, G. (2002). Public Participation GIS for re-development support in European Historic City Centres. Computergestützte Raumplanung CORP, Vienna, Department of computer aided planning and architecture, Vienna University of Technology.


