GIS – WHAT CAN GO WRONG?

Christoph Feldkötter
SMRP

Hanoi, Vietnam, October 1999

FOREST COVER MONITORING PROJECT

- an MRC / GTZ project
- Initial Objective: To provide the MRC with basin-wide homogenous information on Forest & Land Cover for analysis and basin-wide strategic planning.
- covered the entire Lower Mekong Basin
  4 countries
  Cambodia, Laos, Thailand, Vietnam
- mapped Forest & Land Cover from Landsat TM satellite images at 1:250,000 scale in 1992/93 and again in 1996/97
- ended in 01 / 1999
- Post-Project Support 1999 - 2000 provided through Sustainable Management of Resources Project (SMRP - MRC / GTZ)

WHAT IS GIS USED FOR?

IN THEORY ...

The IMAP Model

- Input
- Management
- Analysis
- Presentation

"A GIS is a system composed of hardware, software and procedures for the

- capture
- storage
- manipulation
- analysis
- modeling
- output
of spatial data to solve complex planning and management problems."

... AND IN PRACTICE

The ImP Model

- Input ++
- Management +
- Analysis --
- Presentation ++

"A GIS is a system composed of hardware, software and procedures for the

- capture ++
- storage +
- manipulation -
- analysis --
- modeling --
- output ++

of spatial data to solve complex planning and management problems."

CONCLUSIONS

Not only data input or capture (from existing sources), but DATA GENERATION probably constitutes the most important component of GIS applications in the region.

This is justifiable given the lack of accurate and / or up-to-date base data.

"It is grave mistake to draw conclusions before one has data." (Sherlock Holmes)

"In the 1970s, one could, with reason, have argued that GIS was essentially an extension of cartography with other tools."

Has this really changed?

The fact seems to be that GIS units have taken over the job of traditional cartography.

Are they capable?

GIS is no religion. It is a tool, which, if used in the right way, can make things faster.

WHAT DO WE EXPECT FROM GIS DATA?

ACCURACY

GIS data should reflect the situation on the ground.

- **Positional**: the location of objects on GIS should depict their location on the ground as accurately as possible.
• **Thematical**: the classification of objects on GIS should depict their character on the ground as accurately as possible.

**UP-TO-DATENESS**

GIS data should be recent.

**AVAILABILITY**

GIS data should be available.

1. Technical Aspects
2. Human Resources Aspects
3. Data Use Aspects

**PIXEL SIZE AND INFORMATION CONTENT**

PIXEL = smallest unit of an image

### High Resolution Satellite Image

- Pixel: 30 * 30 m
- Information Content: 1

### Very High Resolution Satellite Image

- Pixel: 5 * 5 m
- Information Content: 36

### Aerial Photo (scanned)

- Pixel: 1 * 1 m
- Information Content: 900
The maximum scale depends on the pixel size.
The information content does not grow proportionally, but with the square of the resolution (or scale) ratio.

**DIGITIZING TIMES AT VARIOUS SCALES**

<table>
<thead>
<tr>
<th>source scale</th>
<th>250,000</th>
<th>100,000</th>
<th>50,000</th>
<th>25,000</th>
<th>10,000</th>
<th>5,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>boundary on the ground (m)</td>
<td>175 Mio</td>
<td>437.5 Mio</td>
<td>875 Mio</td>
<td>1,750 Mio</td>
<td>4,375 Mio</td>
<td>8,750 Mio</td>
</tr>
<tr>
<td>boundary at map scale (m)</td>
<td>700</td>
<td>4,375</td>
<td>17,500</td>
<td>70,000</td>
<td>437,500</td>
<td>1,750,000</td>
</tr>
<tr>
<td>digitizing speed (mm / sec)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>overhead factor</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>digitizing speed (m / h)</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>total digitizing time (h)</td>
<td>389</td>
<td>2,431</td>
<td>9,722</td>
<td>38,889</td>
<td>243,056</td>
<td>972,222</td>
</tr>
<tr>
<td>digitizing time (h / week)</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>digitizing time (weeks / year)</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>total digitizing time (weeks)</td>
<td>19</td>
<td>122</td>
<td>486</td>
<td>1,944</td>
<td>12,153</td>
<td>48,611</td>
</tr>
<tr>
<td>total digitizing time (years)</td>
<td>0.65</td>
<td>4</td>
<td>16</td>
<td>65</td>
<td>405</td>
<td>1,620</td>
</tr>
</tbody>
</table>

"We have aerial photos of the whole country, so why don't we just make a map?"

"Topo maps are there, so why not just put them on a GIS?"

The workload of data generation is usually heavily underestimated.

The initial project (unit) design is dramatically under-dimensional.

The project (unit) never reaches its objectives.

**GEO-REFERENCING (I)**

- The process of referencing spatial data to a map coordinate system.
- This process is of crucial importance for the positional accuracy and the spatial compatibility of data.
- Criterion used to evaluate the quality of geo-referencing: the RMS Error (Root / Residual Mean Square).
- Simplified: RMS Error is the average distance between the actual location of a point (after geo-referencing) and its desired location.
- **Guideline**: the lower the total RMS Error, the more accurate the geo-referencing.
- **In practice**: RMS Error is reduced by subjectively altering coordinates. This may result in very ill fitting data layers.
- Example: interpreted satellite images (1:250,000), positional errors of up to ± 500 meters.
GEO-REFERENCING (II)

- In some cases geo-referencing is done by manually transferring data to base maps (e.g. from aerial photos).
- This works best in flat terrain with sufficiently dense infrastructure or hydrographic features, if these features are used as references.
- In practice: This technique is applied in any terrain. Little use is made of reference features. This may result in very ill fitting data layers.
- Example: interpreted aerial photos (1:25,000), positional errors of up to ± 250 meters (1 cm at source scale!).

DATA GENERATION / INTERPRETATION

- GIS data generation is time-consuming, strenuous and often tedious work.
- "People who have a sufficient qualification think it beneath their dignity to do the routine work (of interpretation). So this work is given to newcomers or less qualified staff."
- Supervision is not done properly.
- Disastrous effects on data quality in terms of both positional and thematic accuracy.
- Example I: satellite image interpretation (1:100,000)

In a full coverage repetition experiment only 64 % of the area was assigned to the same class in both rounds.

Even if the interpretation was simplified to only 2 classes (Forest and Non-Forest), only 88 % of the area were assigned to the same class in both rounds.

- Example II: satellite image interpretation (1:100,000)

In a sample repetition experiment simplified to only 2 classes (Forest and Non-Forest) only 52 % of the area was assigned to the same class in both rounds.

This almost amounts to random assignment!

DATA BASE MAINTENANCE
GIS data bases can be huge (thousands of files, several gigabytes).
GIS data bases are commonly used and updated by various operators simultaneously.
Operators tend to keep the data sets they are currently working on in a rather unorganized way.
Data sets are not named after their content, but after the operator (or his daughter …).
Different updates are simultaneously applied to copies of the same data sets.
Data sets are split for map printing purposes.
Insufficient or no documentation is kept.
It becomes extremely difficult to retrieve information. Enormous time has to be spent on organizing data before the fabled analytical capacities of GIS can be put to any use.

SYSTEM SETUP

- **Over-Dimensioned Systems**

  - Example: UNIX based Systems, ArcInfo Systems
  - Expensive Hardware & Software
  - Complicated System Maintenance
  - Local Support often not available
  - Too much time is spent learning, understanding (and playing with) system & software ® less time spent on production
  - High risk of individuals gaining hidden executive power due to their system management abilities

- **Under-Dimensioned Systems**

  - Example: Slow (Standard Office) Computers, Insufficient Storage Space, Outdated Software (PC ArcInfo), No Networks
  - Frequent Mistake: Cheapest Computers are purchased. Education-oriented systems (IDRISI) are used in a production environment.
  - Too much time is spent on routine tasks and data swapping ® less time is spent on production

STAFF SELECTION

- It is a frequently told (and believed) myth, supported by the industry, that GIS can be used by anyone. This may be true if the objective is just running a couple of queries on a ready-made database. The situation is entirely different if the objective is data generation.
- GIS is a powerful tool. However, one can not expect that it can be operated just like any other computer software.
- Using GIS efficiently does require a certain degree of mathematical, geometrical and logical understanding plus organizational skills.
- GIS staff often are (have to be) selected from an existing pool of available staff. This may be no problem if a GIS is set up in an organization with a geographical background (such as cartography). The situation may be entirely different if a GIS is set up in e.g. agriculture or forestry (“the tree identification specialist turned geographer”).
- Background qualifications of available staff can make or break the success of a GIS installation.
INTRODUCTORY TRAINING

- Introductory training is often held in national or regional training centers.
- These institutions will train people in basic operations. They will train people in the software packages they have available on the training site. The training is normally done without regard to specific (institutional or project) requirements. Result: staff normally have to be re-trained on-the-job later. This normally happens with a considerable delay, so skills acquired during the introductory training have already been forgotten.
- People are rarely, if ever, trained to read (the Help).
- Introductory training is not planned carefully. In many cases it does not produce the expected results. Expensive follow-ups become necessary.

UPGRADING OF SKILLS

- The life span of software packages is not more than 2 - 3 years.
- With every (major) new software generation the productivity increases significantly (example: PC ArcInfo - ArcView).
- GIS staff, however, have a tendency to stick to procedures once established, thus giving away major advantages that could be gained from software upgrades.
- It is frequently done, but normally not sufficient to buy software upgrades without providing additional training.
- Additional training, if provided, is often held in national or regional training centers. Frequently, this is just a repetition of the basic training.
- Upgrading of skills is mandatory. It must be tailored to the institutions / projects specific requirements. It should best be held in-house.

CAMBODIA: AVAILABLE MAPS

Topographic Base Maps

- **1:50,000**
  - compiled by US DMA in 1960s - 70s
  - partly updated & reprinted by Vietnamese
  - positional accuracy about ± 50 meters
  - infrastructure, irrigation, populated places largely outdated

- **1:250,000**
  - compiled by US DMA in 1960s - 70s
  - positional accuracy about ± 500 meters
  - infrastructure, irrigation, populated places largely outdated

- **1:100,000**
  - compiled with Japanese assistance in 1998
  - positional accuracy about ± 50 meters
  - covers only populated areas
- infrastructure, irrigation, populated updated
- availability?

• **1:5,000**

  - compiled with Japanese assistance in 1998
  - highly accurate
  - covers only Angkor conservation area

**CAMBODIA: AVAILABLE GIS DATA**

• **Roads**
  - based on 1:50,000 topographic maps
  - available

• **Rivers**
  - based on 1:50,000 topographic maps
  - available

• **Administrative Boundaries**
  - based on 1:50,000 topographic maps and 1998 census
  - country, provinces, districts, communes
  - available

• **Populated Places**
  - based on 1:50,000 topographic maps and 1998 census
  - available

• **Elevation Contours**
  - based on 1:50,000 topographic maps
  - restricted access

• **Others**
  - based on various surveys
  - varying quality and availability

**DATA OWNERSHIP / DISTRIBUTION**
• Who owns data?
  - The regional agency?
  - The national agency?
  - The producing (consulting) company?
  - The donor who supported their generation?

We may all have our private answers to these questions. However, quite often, there are no written and binding regulations.

• Copyright

Regulations do not exist.

If they exist, there are no legal instruments to enforce them, or existing legal instruments are not used.

• Conceptual Mistakes

Donors funding data generation fail to tie their funding to clear commitments that the data sets will be put to public use.

• Commercial Interests

Consulting companies producing data and officials in charge of data distribution are quite aware of the fact that public distribution may reduce their private profits.

DATA OWNERSHIP / DISTRIBUTION

• Mythology

"Data like these have been used to guide cruise missiles into Saddam Hussein's living room …"

- Consequences

- Data sets that were produced for public use are distributed hesitantly or not at all.
- Agencies are extremely cautious about releasing data sets officially.
- However: data sets are traded unofficially. The profit goes to individuals rather than institutions.
- Data obtained unofficially can not be used in public documents.
- Distribution becomes almost impossible to monitor.
- Institutions become even more cautious about releasing data sets.
- DATA PRODUCED AT HIGH COSTS ARE NOT USED.

META DATA BASES
• **Meta Data**: Data about Data.

  Data descriptions including scale, content, format, completeness, quality, producing agency, availability, price, …

• **Advantages**

  - Data availability situation becomes more transparent.
  - Information gaps are easier to identify.
  - Time and money is saved because repeated (project) investigations of data availability become unnecessary.
  - Even more time and money is saved because the redundancy of data generation can be reduced.
  - Selective inputs in data generation can be made.

**META DATA BASES**

• **Common Problems**

  - The workload of establishing a meta data base is underestimated.
  - The person(s) in charge of establishing the meta data base have little GIS experience. They are not qualified to judge the information provided.
  - Statements of producing agencies on data quality and availability are accepted without evaluation.
  - Agencies list everything they have, including copies of data sets obtained from other agencies. This leads to high redundancy.
  - Agencies make misleading statements on the completeness of data sets.
  - Agencies fail to provide information because their own internal documentation is incomplete.
  - Agencies withhold information.
  - The establishment of the meta data base is a one time exercise without proper establishment of update procedures.

**NEW DATA BASES**

• Setting up a "New GIS Data Base" is still a favorite among donor activities.

**What happens?**

• The time frame set for this activity is rather short.
• The project is under-staffed.
• Data collection from all possible available sources is started.

  These sources provide mainly data sets at small scales, not useful for anything but overviews.

  No or few quality checks are done on the data collected.
Occasionally, data generation at small scales is started to fill perceived gaps. (Which may not exist at all. Sometimes people are just not aware that data sets already exist.)

**What is the product?**

- An impressive, flashy looking "new" database, perhaps even on the Internet.
- This database merely contains copies of data sets that existed earlier.
- One may end up with several duplicate databases, which differ only in names and layouts, but not in their contents.

**RESEARCH AND PRACTICE**

- One finds quite a number of cooperations between projects and universities, wherein the universities provide the GIS know-how.
- Advantage: university know-how commonly comes at a lower price than the services of permanent advisors or private companies.
- Disadvantage: university know-how is often rather theoretical and a means in itself.

There is much fascination with technique, but little knowledge as to what information is required.

Their research approach may also differ widely from a approach oriented towards production.

- Example:
  
  Various vegetation classifications carried out in overseas universities with little or no ground truthing.

  "Scientifically" designed forest inventories.

- **Consequences**

  **Widespread dissatisfaction among institutions and projects as far as GIS is concerned.**

  The assistance offered by universities often constitutes the first GIS experience of an institution or project.

  Understandably, this may result in a "No thank you" attitude towards GIS in the future.

**SO WHAT DO WE NEED?**

- clearer identifications of the user’s information requirements
- better explanations of what GIS can do and what not
- better explanations of how information content is related to scale
- sound estimates of how much time and effort GIS operations may require
- more production-oriented approaches
- technical guidelines and procedures instead of repetitious data bases
- better quality control
- more structured training and upgrading of knowledge
- better formalized exchanges of meta data
- better coordination among institutions that generate data
- clearer definitions and more transparency of data ownership
- more open data distribution policies
- clear commitments from the donor side that if the generation of GIS data is funded, the products are expected to be used publicly

**PUBLIC MONEY CREATES PUBLIC DATA!**