Ways to integrate real World into GIS

Lorenz Fanger

Socrates – Erasmus Summer School, Brno 2006 : Full Integration of Geodata in GIS

1 Introduction

In the last years GIS developed to widespread and widely used Toolset, used in wide variety of fields in a similar wide variety of tasks. Therefore many different approaches were realized to integrate real world in a Geographic Information System (GIS). However, to get an idea of the varied approaches, to understand the concept of this approaches and to evaluate the characteristics of them it is worth to have a closer look on some typical once of them to may get an overview of the state of art an to get an idea of upcoming developments.

In first Step it is necessary to clarify the different steps of bringing real World into a GIS. What ever software or data we use a certain common down stepping from real world to a GIS is common to all approaches, fields and the tasks the GIS should be used for. Having this common stairway its going to be interesting where some crossroads are inherent, where a certain decision have to be made and what kind of possible decisions are known and in use up to now or what kind of roads can be chosen.

Since Data are the fuel of GI-systems we also need to have a look what kind of Data can be incorporated. How can we find and or get them and finally with which criteria we have to select or produce them.

After all this we will have a look on different approaches, how the real world already have been introduced and integrated in GI-Systems, with which we can find the right information for questions of a certain field.

Since no System who tries to incorporate the world can ever be perfect, we like to look for the potential of further development which can be expected or asked from the existing approaches.

2 References and source materials

The Knowledgebase of this text are lectures, practical trainings and distributed material of the Summerschool 2006 in Brno with the Titel "Full integration of Geodata in GIS". Lectures on this Summer School where held by:

- Dr. Adrijana Car, Centre of GeoInformatica – UNIGIS Salzburg University, Austria
- Doc. RNDr. Jaromir Kolejka, CSc, Mendel University of Brno, Czech republic
- Tomáš Mikita, Mendel University of Brno, Czech republic
- Ing. Martin Klimanek, Ph.D., Mendel University of Brno, Czech republic
- Pavel Vranka, Masaryk University of Brno, Czech republic
- Antal Guszlev, College of Geoinformatics – University of West Hungary
- assoc.prof. Jüri Roosaare, University of Tartu, Estonia
- Lluis Vicens, SIGTE University of Girona, Spain
- Andrzej Kotarba & Jacek Kozak, GIS Laboratory, Institute of Geography and Spatial Management, Jagiellonian University Krakow, Poland
- Prof. Ing. Ján Tuček & Ing. Andrea Majlingová, skTechnical University Zvolen, Slovak republic

In the following text the lecturer where only cited in connection with special Ideas and concept they brought up. Common Knowledge and concepts were used without any reference. Concepts and Knowledge which were introduced from other sources were cited and listed at the End of the text.

3 Concepts

3.1 Modelling real World

To bring Reality into an information system (IS) and even further in a digital format in a computer is one of the challenges of Information technologies (IT) which originates at its very beginning and reside it till this days. If this is true for IT in general it is even more for geoinformation technologies. The specific thing on Geoinformation within the broad field of information is its spatial reference. Never the less there are some common steps when bringing any reality to any IS. First we have to create a conceptual model of the reality, of the real world. We have to name, to specify the things, the Objects, the Entities we want to describe and to include in our IS. Further we describe them with their behaviour and possible interrelation they can have. This step is descriptive in that way, that we mainly can use words, text or possibly tables. In a second step we create a logical model in which objects, their attributes and relation are defined in detail. Often graphic tools or methods such as UML or so were used to catch the reality as a more or less independent system. In a third step we have to think about how the logical model can be transformed into a physical model which in GIS also can be seen as the Representation. Here we have to be aware of the important characters of our objects for our model. Are they of continuous field character or are they discrete objects? Do they change in time and does that matter to our system we want to model? Do we have also to take in to account the third Dimension? The output of this last third step should be a physical data structure in which the data can be filled in. This possibly includes among others raster data models, TIN's, vector data models or/and Databasemodels. Once the model has established, the Data are gathered and filled in the model, we are ready to start using this integrated Data as a Base for Analysis to produce further geodata which may can be included again in the Model.

3.2 Data sources

As mentioned in the last abstract every datamodel needs data to fill in. Sources of such Data are various and reach from measuring oneself in the field till downloading third-party data from internet. As self collecting Data is expensive and is a shrinking market since many datas can be found in Internet and can be often bought cheaper somewhere else, in this abstract we want to concentrate on questions like how we can find the right data and what have to be considered to get right data? With the consolidation of the internet technology and searching-tools the idea raised, that geodata may can be searched and find on the internet like homepages. In order to search geodataset on specific criteria, structured Metadata (data about the geodata like descriptions of its content, quality, condition and other characteristics) has to be collected for every geodataset and stored in catalogue services similar like for homepages. Further if all the metadata are structured in a standardised way, both the exchange of such datasets and the search in them can be standardised and therefore catalogue services can communicate with each other and can built up a network of a huge metadatabase. With the ISO Standard 19115 such a standard was established and published in the last years and will now more and more be integrated in the actual metadata catalogues. Up to now in many cases every single metadata catalogue has to be known and searches have to be made in every catalogue of interest individually. The Outputs of such searches are lists of datasets which fits the one or the other way to the criteria we inputted in the catalogue service. While browsing this list we can check the results and to check further criteria. Criteria we should take into account earlier or later in our data search are:

- Coverage
- Currency
- Restriction and Price-Policy
- Distribution details (Distributor, Media, Dataformat, prices)
- Accuracy and Precision of positioning and attributes (Precision: of measurement scale, Accuracy: measurement compared to reality)
- Consistency and Completeness
- Way the data were generated
- Spatial reference systems, Encoding methods (coordinates, addresses…)
- Dataformat, Datamodel, Datastructure
- Physical data size (KB, MB, GB…)
- Use and meaning of codes

4 Ways to integrate geodata on selected examples

As we now have clarified the meaning and relation of Datamodel and its (Source-)data we can start to discuss examples of datamodels, its sourcedata and their integration.

4.1 Digital Landscape Model (DLM, introduced by a talk of J. Kolejka)

Most geodata and models follow an analytic approach to describe the world. In contradiction to that most geodata analyses require a synthetic approach.

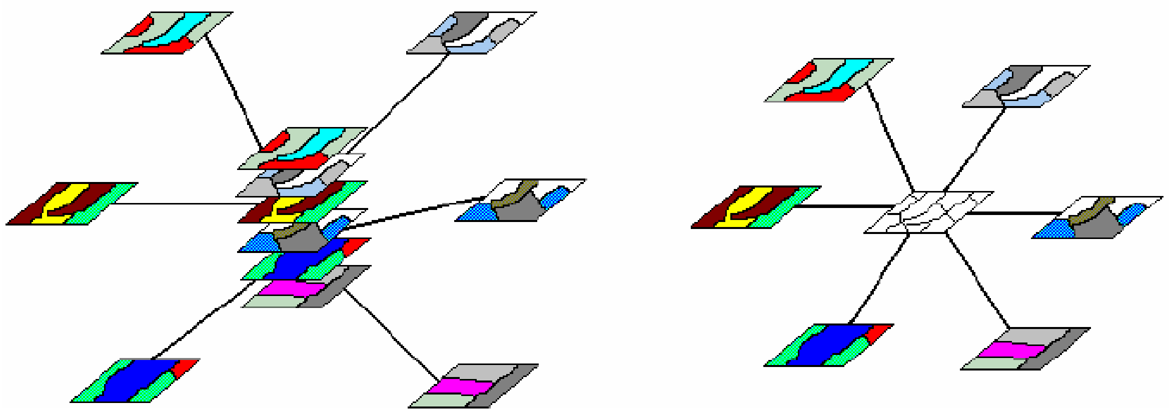

Fig 1 Difference between collecting (left) and intergating datas (right) according to a Tal of J. Kolejka

This is why in this digital landscape model (DLM) the most time consuming part of synthesis is already done by integrating source datas in one layer and not just collecting and overlaying them (see Fig 1). A DLM is normally composite out of 4 sub themes. 1st it consist of a Natural Background with homogenous natural landscape units as reference areas. 2nd we introduce Products of Human Impacts with parcels and/or subparcels as reference areas into the model. 3rd layers of human and social interests are defining social limits with parcels and/or subparcels as reference units. 4th a digital elevation model builds up the skeleton for the whole model. Following on the example of the natural background theme the integration of different source datas are described. Having a set of classified thematic maps which influences the landscape we can integrate the different maps as Layer into one Model. Such thematic maps could contain classified data such as:

- Local climate
- Biota
- Drainage net connecting events
- Land forms
- Disposable water
- Soils

- Surface Geology
- Etc.

The Integration as such should be realized as followed. Each map get scanned, georeferenced and digitized if the source is an analogue map. If we have already digital featureclasses we can load them directly to ArcGIS or another GIS. Having after that a pile of vectorized layers we ad at last a layer with vectorized topographic information. Further we put the layer in an order were independency of the layer theme (Topography > Geology > Biosphere) and its assumed ore known reliability (accuracy) relative to each other increases. From the most reliable and accurate layer, which is in most cases the topography and or the geology we check all the other layers on their geometric and logic consistency relative to the upper layers. For example if we have in the geology layer river gravels showed in areas where it is not following the river valleys and creaks of the topography and or drainage network. This work needs expert knowledge as we need some ideas about logic and non-logic combination of different attributes from different layers. With tools like trace, snapping and topology in ArcGIS the geometries can be adjusted relatively easy and fast. Once having a consistent pile of layer with the same coverage we can intersect them each other. Finally we have one featureclass which covers the hole common area with a mosaic of polygons. These Polygons possess all attributes of the different layers in this part of the covered area. This type of integration can be applied to all the other subthemes if necessary. Only the DEM most probably stays as it is and can serve as mentioned as a skeleton of the whole model. At the end the DLM with its four subthemes can be used as a base for any Landscape analysis. Further the DLM can be disintegrated at any time by classifying and if necessary merging on the different attributes of the source layers.

4.2 ArcHydro

With the ArcHydro module of the Spatial Analyst of ArcGIS and its provided tools a predefined data model will be built up and filled with data in the form of a object based, structured geodatabase, wherein all Features are managed. Data, featureclasses which were saved in the database can be connected by relations with each other. Like that a more complex world can be modeled. In case of the ArcHydro module the input data are a DEM, a vectorized drainage network and time series datas of defined measuring points in form of tables. From these datas following datamodel can be built up which can be divided logically into 5 data components:

- Network: connected sets of points and lines showing pathways of water flow.
- Drainage: drainage areas and stream lines defined from surface topography.
- Channel: a 3D line representation of the shape of river and stream channels.
- Hydrography: the base data from topographic maps and tabular data inventories such as for example bridges, dams, lakes, water discharge or withdrawal points etc.
- Time series: tabular attribute data describing time-varying water properties for any hydro feature
- The first 2 Components can be almost entirely derived from the DEM and if available from the vectorized drainage Network. The other data have to be loaded to the database considering the consistency of the data model.

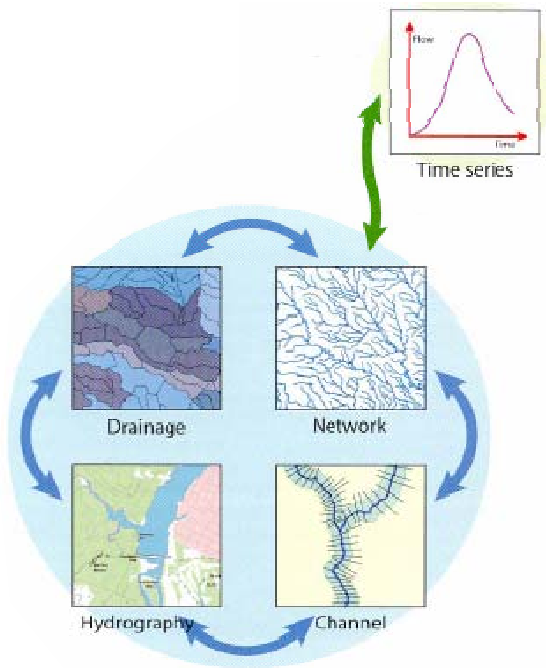

Fig 2 Relation of the five data model components in ArcHydro

With this Geodatabase we are able to achieve complex analysis of the hydrological system of a certain area.

4.3 Land Use and Cover Change (LUCC) modelling

To model landscape changes we have to deal with space (2-2.5D) and time (4^{th}D). To do so we need to discretesize the Dimensions such as using time steps and rasterized space. Such models can be classified according to their complexity as followed:

	Model				
Criterion of complexity	Simple evolution models	Local dynamics models	Coupled dynamics, single-system models	Coupled dynamics, multiple-system models	Models with dynamically changing structure
dynamic interactions of a number of local parameters	-	X	X	X	X
Interactions between neighbouring grid cells	-	-	X	X	X
different model for different cells	-	-	-	X	X
character of interactionsbetween neighbouring grid cells can change	-	-	-	-	X
local models can be replaced with some new properties	-	-	-	-	X

Tab 1 Models vs. criterion of complexity according to a talk of J.Rossaare

The words giving the abrivation of LUCC include land cover and land use whereas land use implies human activity and land cover is the physical coverage of a space unit which observed by remote sensing. With MARKOV, CA_MARKOV, GEOMOD and Land Change Modeler (LCM) IDRISI provides several powerful tools to model Landscape changes. To get an Idea of the power and the complexity of such tools in the following an overview of the possibility of the LCM are given.

- finding changes in past (as a basis for prediction)
  – what kind of transitions have been taken place?
  – what might be the driving forces of changes?
- generating model for these changes
  – to identify explanatory variables
  – to select suitable modelling method
- predicting land use changes
  – to find change demand
  – to make change allocation
- assessing the impact of changes
  – for ecological sustainability
- attaching human interventions
  – to introduce infrastructure changes
  – to add constraints / incentives

Beside the complex modeling, data sources and data integration is on of the key challenge in land change modeling. You have to deal with problem like:

- Incomparable historical maps
- Scale and generalization problems
- Different feature definitions with changing time and people
- Lacks of explicit data, many historical data are just descriptive

This means the most data have to be integrated connected with expert knowledge. However with methods and tools like the LCM in Idrisi the past can be modeled. Adding some assumption we maybe even able to extrapolate the modeling into (near) future.

5 Discussion and a possible View into future of data integration

The summer school tried to give an impression of the vide and complex field of data integration in many different fields of GIS. There will never be single approach to this topic, most of all as GIS is a advancing technology. Some of the most important actual developments in GIS and its closer technical neighborhood are:

- Full Integration of GIS in DBMS
- Location based services (LBS) and mobile GIS in general
- The need and development to handle, analyse and visualize real 3D and 4D (time) data
- Further developement of Computer technology most of all increasing the speed and the capacity of computers as well as of communication networks

These developments will guide but also offer possibilities to data integration as well as to the whole GIS and will make them even more versatile and complex than it is nowadays.