The Feuropean Forum for GeoStatistics

Modelling with grids (Issues for discussion)

A summary (Notes towards aVision)

(This summary is very preliminary; a note for discussion. For inspiration please see http://www.koreus.com/video/powers-of-10.html)

Models: Our overriding task is to produce integrated and useful systems of information to serve as a foundation for Policies, programs, plans and projects for sustainable development on all levels of public administration from local to Global. This model must, we suggest, be based and based on a a fixed system of scale intervals following powers of ten. These models should in their structure and processes respond to state of the art in applied systems analysis and theory.

Object oriented and layer based: This requires an object based spatial data infrastructure SDI based on an object approach built according to the layer method and processed in with GIS technologies. Each object should be seen as a combination of object components; Feature objects, Attribute objects etc. One layer for each class of integrated objects. Each model as a stack of layers structured in response to shared semantic models. Progress in this field will require an integrated approach to an information system to describe whole man- environmental systems (MES).

One model concept for all purposes: This hierarchical system of information (or model) should serve be flexible enough to serve all stages of the process of action from data capture, analysis, policymaking, project design and implementation as well as the evaluation of its effects through a new data capture process to serve as a foundation for the next iteration. It is not efficient to produce more than one infrastructure to serve all purposes.

Subject to the rules of the Scientific Method and the method for Iterative Development: Models must be presented as the result of an applied science that focuses on the integration of information for analysis and synthesis. Descriptions must follow the rules of the scientific method to serve both the development of theory and praxis.

Accessible over the Internet: Models in this sense are hierarchical systems of information made available over internet portals where spatial data are presented as integrated wholes. Services to locate data in response to efficient metadata and cataloguing services. Complex objects (Integrating Geographic features and statistics) may be compiled by combining available spatial features with suitable systems of attributes. Adequate services are available to explore the results of a search both in terme of the features and the statistic (through diagrams, tables). Business services are awailable acquire/download or order the datasets required.

Stockholm September 29, 2010

Lars H. Backer

Table of contents

1. Integrated Models using complex objects (from the GGIISS proposal)

In the current proposal for a Vision for a integrated approach to the building of information systems (models) for statistics, we take it for granted that we need to discuss the context into which e.g. a system of statistical grids shall fit. For this purpose we will provide a rough concept towards a Global infrastructure of spatial information (the GGIISS proposal)

1.1. If you cannot describe it you cannot manage it.

The motive behind our current efforts to provide a spatial data infrastructure for Statistics is part of a global effort to provide systems of information that provide models of the global system for operational use. With “real user needs” we understand the needs of a whole string of information processes that lead from observation (data capture) analysis, policymaking, action programming, decision support and decision-making, planning, project design and implementation, use, and back to the capture of information to serve to revise and hopefully improve the process in the next iteration. (See the Cybernetic method described in the Geostat Hypothesises)

This information is probably the most important raw material for development the artificial equivalent of the evolutionary process.

1.2. Systems Approach

The best strategy we have at our disposal for building efficient systems of information is to found it on the use of systems theory. Basically a system is a set of objects with relations. This allows for two perspectives to systems that are always inter connected.

1. Systems as “Kits of parts” (With focus on the objects)
2. Systems as “Interacting wholes” (With focus on relations)

In the current report we will focus only on the descriptions of systems as kits of parts. We will leave the discussion of the use and development of these information systems to describe dynamic (temporal) patters to a later stage of the current project.

1.3. Hierarchical modeling our world(s) and the three tier idea.

The objective for the current excise is to provide a set of operational multipurpose sets of information that may be used on all levels of public authority from global to local. (INSPIRE Drafting Team "Data Specifications" 2007). This formulation is based on the three tier idea that is based on the idea of the world as a hierarchical system of systems. This in turn requires an hierarchical set of “Chinese box” or “Russian doll” models . The three tier idea forms in praxis the basic structure for all planning and development efforts including those of e.g. spatial planning and development where three scale intervals are always involved in the development of every project. The idea is that the scale interval used for the focus of the project is the middle scale; this has to be designed and developed with reference to the next higher level, and also describe the consequences for the lower. Thus we have

1. Datasets for European development policies and actions
   For the planning and evaluation etc. of projects to serve the sustainable development of the European Union as a whole, we need a the triad of 1) Global (higher, context) /2) European (focus)/3) National (lower, details) datasets. Thus also:
2. Datasets for National development policies and actions
   For the planning and evaluation etc. of projects to serve the sustainable development of EU Member states, we need a triad of 1) European (higher, context) /2) National (focus)/3) Regional (lower, details) datasets. Thus also:

3. Datasets for Regional development policies and actions
   For the planning and evaluation etc. of projects to serve the sustainable development of National regions, we need a the triad of National (higher, context) /Regional (focus)/Local (lower, details) datasets.

Thus development projects require that the datasets used as their foundation must be scalable in this sense all the way from Global to local.

1.4. The Problem of scales

This involves the problem of scales and scale intervals. In praxis it may be argued that the use of map scales is a left over from the former days of analogue maps. For work with digital mapping systems it is more important that date are efficiently described in datasets that correspond to the hierarchical system of public authority systems. /Global, European ,National, Regional, Local (Commune/district), Locality(Village/urban block)/ etc.

In praxis it can be observed that the scale intervals between these harmonise roughly with the so-called "Power of ten" strategy, which allows for bridging the Global and the Local over 6 intermediate steps.

1. 10^7m objects (10000km windows) (To describe Europe, Global Hemispheres),
2. 10^6 objects (1000km windows) (To describe Large National State, Macroregion),
3. 10^4 m objects (100km windows) (L To describe arge national Region small National state),
4. 10^3 m objects (10km windows) (To describe Commune, large urban system).
5. 10^2 m objects (1km windows) (To describe Locality, Urban distict),
6. 10^1 m objects (100m windows) (To describe locality, village, Urban block)).
7. 10^0 m objects(10m windows (To describe ideal points in space?)

For each of these steps we suggest the use of the term "windows" to data that in terms of resolution, variables etc. fit reasonably well together for practical purposes.

1.5. Semantic modeling

Semantic descriptions form the foundation for all modeling efforts. When using the "Object approach" to modeling Nouns are generally transformed into Objects, verbs into methods etc. It is of primary importance that all data producers and data users use the same or similar 8translatable) semantic descriptions and codes.

A semantic model provides the foundation for any modeling effort. It must have a hierarchical structure. We need a semantic strategy to describe the Earth as an integrated man- environmental system. This involves integrated semantic modeling for both human and environmental systems. There are a lot of possible taxonomies in general use such as:

- Society
  - Sociocultural system(s)
  - Economic system(s)
- Environment
  - Built environment(s)
  - Natural environment(s)

We have not progressed very far towards defining a standard approach to an international accepted semantic taxonomy of this kind. This report will build on the layer method as a foundation with a method that allows for multiple variations for different purposes. This could be excemplified with a structure to emphasize the dichotomy between natural and artificial systems (this would please

Gianbattista Vico). This structure may be compiled using the same set of layers and the same semantic foundation.

- Natural systems
    - Man as a biological animal (the product of evolution albeit with logos)
    - Nature (the product of evolution)
- Artificial systems
    - Scocio- cultural and economic systems (virtual systems)
    - Manmade environment (real systems)

In our approach the layers are the key, not the structure for their assembly. This principle will be further discussed in connection with the Layer method below.

1.6. Complex Object Structures

The object strategy to model semantic objects and methods, must respond to the idea that objects are complex and including:

- **Object Name** (with coding system)
  Developed from a shared semantic description
- Object Features /Feature Objects(including metadata for object features)
  These are often presented as spatial (geographic) "objects" in their own right with attributes and methods that relate to the geographic dimension only.
  (Generally have developed their own solutions for metadata and other services)
- Object Attributes /Attribute objects (including metadata for object attributes)
  These are often presented as statistics "objects" in their own right with links to features (points, administrative areas) and methods that relates mainly to the attribute information.
  (Generally have developed their own solutions for metadata and other services)
- Object Methods /Method Objects(including metadata for object methods (e.g. delineation methods))
  These are often presented as "objet" methods in their own right without direct reference to particular methods limited to a specific set of objects.
  (Generally have developed their own solutions for metadata and other services)
- Object Reference /Reference Objects (including metadata for bibliographical sources of information.

In praxis the different aspects of objects are produced in different environments to be compiled by the user for different purposes. Thus:

- Systems of objects as Features (provided by mapping agencies) and
- Systems of objects as Attributes (provided by statistical institutes).
- Systems of objects as Methods (generally developed and discussed in isolation by the academia (research institutions))
- Systems of objects as References ("Clouds" of references often provided by libraries and other systems of informatorics.

In praxis there exists a one - many relationship between object Feature, attributes, methods etc. This means that each of these systems must be stored separately and assembled according to need.

In spatial terms systems of Objects with focus on attributes are used for two purposes.

- The creation of spatial patterns as a result of spatial analysis (The Grid issue)
  (the creations of patterns based on point information).
- The comparison of spatial patterns (The GGIISS issue)
  (the creation of patterns of patterns)

This means that the idea of "patterns" is crucial to the idea of spatial analysis.

1.7. Layer method

It is taken for granted that models will be digital and layer-based. This means that objets of the same class are kept in the same layer / database table. These are the linked together with traditional relational database techniques. The ordering / assembly of the different layers are open

There are proposals here as the customary division of:

- Society
  - Sociocultural system(s)
    - Layer1
      (combination of Feature objects, Attribute objects, Method objects Reference objects etc (each with metadata)
    - Layer 2
      (compilations of parts as layer 1)
    - …
  - Economic system(s)
    - Layer1
      (combination of Feature objects, Attribute objects, Method objects Reference objects etc (each with metadata)
    - Layer 2
      (compilations of parts as layer 1)
    - …
- Environment
  - Built environment(s)
    - Layer1
      (combination of Feature objects, Attribute objects, Method objects Reference objects etc (each with metadata)
    - Layer 2
      (compilations of parts as layer 1)
    - …
  - Natural environment(s)
    - Layer1
      (combination of Feature objects, Attribute objects, Method objects Reference objects etc (each with metadata)
    - Layer 2
      (compilations of parts as layer 1)
    - …

This principle should be demonstrated on the proposal towards an European Infrastructure of spatial information to be compiled in connection with the INSPIRE annexes.
The key is to keep Feature objects and attribute object separate, to be combined (or not) according to real user needs.

1.8. Digital spatial data infrastructure (SDI)

The challenges of the digital method requires special infrastructures. That goes for all the major parts of an object model (features, attributes, methods, and reference information)

1.9. Datum

The question of datum seems to be settled as we have a shared Datum for European Use. The author does not know if there is a similar global standard. However, datum as also projection requirements will vary over the hierarchy from local to global. We need to use one standard that is flexible enough to work over a broad spectrum for general use.

- Datum (WGS84 for Europe)
- Datum for the Earth (What is used here?)

1.10. Projection(s)

The use of Projections is very important du to the distortion of grid systems when datasets are migrated from one projection to another. The crucial discussion is about the projection that should be used for the capture and storage of data.

UTM for a global solution

- For global use and extensive national and local purposes UTM projections are used.

Lamberts conical Equal Area projection

- The INSPIRE project has settled on the use of a local projection for European purposes. This descision is controversial.

1.11. File formats for Complex Object Structures)

Formats for complex (integrated- (muiltidimentional-) objects: It is taken for granted that when discussing datasets (complex objects) that requires the integration of databases and maps, we will probably prefer to work with data in one of the many GIS formats. We have to discuss which to use in this case. Here is generally few problems because it is relatively straightforward to translate a file from one format to another.

Howeverit is only in a file format like Internet webb pages where it is possible to build true integrated models models that integrate all the necessary components needed to describe complex man- environmental systems. Therefore the form of the classic webb page system structured for the use with fullfledged information systems (like the GGIISS vision)

However this is an extensive issue, and besides, in most cases complex objects seem to be assembled on the internet via websites specially build for the purpose. (See the INSPIRE hierarchy of portals)

2. Modelling with attribute objects / object attributes (from the Grid proposal)

In our proposal for a vision for a infrastructure of statistical information and a search for a SDI to support it we have proposed that we treat and build complex objects from simple components (Feature objects, Attribute objects, Method objects and Reference objects) there may be many more but we are here (in the EFGS) primarily interested in Attribute objects.

2.1. Statistical system as a system of object with focus on attributes

The complex objects idea rest on the given situation where object systems have involved in different environments without necessary coordination. So also within the Statistical community. The Semantic model used as foundation for most national statistical systems are not mutually harmonized, and these models do not correspond to semantic models used in other fields e.g. by the mapping agencies.

2.2. Object Variables (For attribute objects / object attributes)

For the Statistical system (of systems) the focus is on the variables. We need a much deeper focus on the variables that are needed and required on different levels of public authorities. (See the six windows or scale intervals proposed in this note.)

- Day-time population
  For most purposes we will need to model economic systems with information on the day-time population. This will ideally have to consist of individuals participating in business and industrial enterprises based om mappings of places of work.
    - Adequate variables to describe and analyse day-time populations and economic systems (business- industrial clusters etc.)
- Night-time statistics
  In addition to this we will be well served with qualified information to describe and analyse social and cultural systems based on night- time statistics
    - Adequate variables to describe and analyse night-time populations and social and cultural systems.

2.3. Hierarchical structure (For attribute objects / object attributes)

The need for a working hierarchical structure for the system of objects with focus on attributes must naturally fit the hierarchical structure required for different levels of public administration etc. from Global to local.

1. Global
2. Hemisphere (e.g. Europe)
3. Large country or Macro region
4. National region or small national state
5. Commune or large urban system
6. Locallity or urban Block

There has some discussion about the need for a Global vs. a European system of grids. This question is primarily related to the projection and datum used for the storage of data. The main problem is one of distortion when different projections and geoids are used.

2.4. The question of scale (For attribute objects / object attributes)

The Hierarchial structure throws up the question of scale. This question has been avoided in the INSPIRE project. This is a pity due to the scale issue is quite critical for the integration of datasets in a digital environment. It would help a lot if scales could be solved by agreeing upon a series of standard scale intervals. These could also follow the power of ten structure.

Windows (Window grids or Scale intervals)

- The "Windows" or scale intervals should for practical reasons generally be in the scale between 100 to 1000 primary grids to the side of one window. This will guaranty a good foundation (resolution) for spatial analysis. As an example a 100km window for regional development will require a system of primary grids of either 1 km (+ secondary grids when needed) or 100m (+ secondary grids when needed)

2.5. Points as grids (For attribute objects / object attributes)

The idea of spatial analysis in the Grid proposal is firmly based on the idea that spatial analysis is conducted primarily with point information. In this grid-based proposition however there is no such thing as a "point" because all positions are given with coordinates, that in turn describes a smaller or larger grid cell, depending on the accuracy given. Thus in a digital system of coordinates as proposed by the "Grid paper" there are 7 digits to get down to 1m accuracy. Thus a point grid with different accuracies may be described as:

Xxxxxxxyyyyyyy (for one meter accuracy)
Xxxxxx0yyyyyy0 (for a ten meter accuracy)
Xxxxx00yyyyy00(for a one hundred meter accuracy)
Xxxx000yyyyy000(for one kilometer accuracy)

When data are aggregate are aggregated to administrative areas the resolution of the data are naturally destroyed, or rather decreased. Given a system of small administrative areas (a system of irregular tessellations) then the accuracy of that dataset is roughly equal to a square that circumscribes the area or even larger because we are forced to use the areas point of gravity to represent the whole.

It is very difficult to disaggregate data that has been aggregated in this manner. Therefore it is preferred that microdata are captured and stored at the highest possible resolution, or if confidentiality rules prohibit it, are <u>captured automatically to the highest resolution allowed</u>. This would mean that microdate would be captured as part of both a system of regular tessellations (grid structure e.g. 1000m grids) andat the same time (based on the position) aggregated to the smallest unit of administrative- and / or other area system (water catchment areas etc.).

However, ideally, data should be captured and stored with the highest possible accuracy because the efficient use of statistics require that they can be aggregated to any system of regular and irregular tessellations.

2.6. System of Grids for aggregations (For attribute objects / object attributes)

Primary grids

- Spatial analysis requires a reasonable level of data resolution. Promary grids should also follow the "Power of ten" rule (1m, 10m, 100m, 1km, 10km, 100km, 1000km, 10000km etc.) for both primary grids and Windows.

Secondary (Quadtree grids)

- Division of a grid or window into, 4, 16 or 256 parts (e.g. a km grid into 4x500m grids, 16x250m grids and 256x125m grids. The quadtree hierarchy should stop here.

2.7. Gradicule and co-ordinate systems (For attribute objects / object attributes)

It seems self evident now that we will use a meter-based gradicule. For some international datasets, a Long Lat based gradicule has been used. This is generally judged to be quite impractical due to the somewhat cumbersome mathematics.
Meter

- The European system is based on a gradicule of primary and secondary grids according to the power of ten strategy. Theer has been little discussion about the gradicule to be used. There is a general acceptance for the metric system.

LongLat

- Often used for navigation, moving objets as points over large territories. Often in arch Degrees, minutes and seconds, or degrees and decimal degrees. (e.g. the CIESIN has produced a statistical gridmap of the world showing the distribution of the global population on 30'' grids). This has not been an obtion for the European system of grids.

2.8. Coding system (For attribute objects / object attributes)

The coding system for the meter coordinate system is very simple and intuitive at least when used on systems of regular (primary) grids based on powers of 10.
For the secondary grids ther are two proposals made in Europe (see the proposals by Albrecht Wirthman and that by Lars Backer in preparation for the Eurogrid discussions at the JRC).

- Coding system for primary grids
  This is constructed from the coordinates for the lower left corners (south-west corner) of a grid.
- Coding system for Schondary grids (Quadtree structure).
  The coding system for the sechjondary (quadteree) system is not (to my knowledge) solved. There have been two proposals one from Albrech Withman (EUROSTAT) and another from Lars Backer (then statistics Sweden)

2.9. Metadata (For attribute objects / object attributes)

We have had long discussions about the use and form of metadata. There seem to be two alternative strategies. The key issue has been the question of integration.

- The first strategy is to have one <u>metadata structure for both features and attributes</u> in the same metadata xml sheet.
  This would be to decide for a simple object structure with <u>one set of metadata for all spatial information</u>. This seems problematic due to the fact that there are already in use metadata systems for systems of statistical objecs, systems of geographical objects etc.)
- The other to have a <u>separate structure for the two</u> due to the one to many relationship between the two. This would mean that the problem of harmonization could be reduced.

This may in both cases be solved through the development of a more complex cataloging service. This strategy could be used also for the integration of other reference data like bibliographic information.

2.10. Projection (For attribute objects / object attributes)

The use of Projections is very important du to the distortion of grid systems when datasets are migrated from one projection to another. The crucial discussion is about the projection that should be used for the capture and storage of data.

UTM for a global solution

- For global use and extensive national and local purposes UTM projections are used.

Lamberts conical Equal Area projection

- The INSPIRE project has settled on the use of a local projection for European purposes. This descision is controversial.

The question of projection is very important. The three tier structure will require that at least for the higher realm of the datasets required for international cooperation, one commonly shared projection should be used. The Lamberts Conical Equal Area projection should therefore not be used because it is not a Global projection. The global UTM system is currently in use in many corners of the world and so also extensively in Europe (e.g. by statistics Norway and others).

2.11. File formats (For attribute objects / object attributes)

Formats for simple (one dimentional-) objects (Attribute objects, or Feature objects etc.)
On the other hand, when discussing objects that do not involve geographic features, e.g. statistics with positions given as strings of coordinates, we might use any database format. In this case it is of course perfectly possible to store the data in any database format. The grids are then generated from the coordinates directly (on the fly). Methods may be separately stored in other formats and Reference literature in formats used for searching for literature between libraries and archives.

Formats for complex (integrated- (muiltidimentional-) objects
It is taken for granted that when discussing datasets (complex objects) that requires the integration of databases and maps, we will probably prefer to work with data in one of the many GIS formats. We have to discuss which to use in this case. Here is generally few problems because it is relatively straightforward to translate a file from one format to another.

However this is an extensive issue, and besides, in most cases complex objects seem to be assembled on the internet via websites specially build for the purpose. (See the INSPIRE hierarchy of portals)

1. Modelling with object Features /Feature objects

This topic, although an integral part of a comprehensive integrated system of information, is beyond the scope of the current project. This issue is being developed in Europe through the INSPIRE project.

2. Object Methods /Method objects

This topic, although an integral part of a comprehensive integrated system of information, is beyond the scope of the current project. However, it is clear that standardized methods are needed to generate patterns (methods for delineations e.g. of urban areas, or the simulation of a sysetmes dynamic behavior over time.

3. Object Reference /Reference Objects

This topic, although an integral part of a comprehensive integrated system of information, is beyond the scope of the current project. However, it is a key requirement to all work conducted according the rules of the scientific method, that a serious contribution to knowledge must make reference to key literature.