A "BIG PICTURE" DESCRIPTION OF PRODUCTION PROCESS(ES) FOR STATISTICS

SUMMARY

This paper represents an effort to provide a broad- or big- picture of the production processes serving as a background for the "technicalities" for the production of National Geostat dataset that in the "Big picture" appears mainly in section 4 of the following process overview.

1. Data model
2. Data capture
3. Object- based statistical databases
4. <u>Non- spatial, Spatial and Temporal- analysis</u> (Production of geostat datasets proper)
5. Compilation of National, European or Global Geostat dataset(s) for dissemination
6. Dissemination
7. Feedback on real user needs
8. Overall critique and quality assessment

I think that this two level approach is important for our discussions about the development of the spatial dimension of the European Statistical System. I suggest that a discussion of all issues related to the Geostat datasets should refer to these two perspectives:

I. Focus on the whole process
(all issues 1-8)

1. Big picture process description (Paper by Lars H. Backer under preparation)
2. Big Picture quality assessment (Paper by Lars H. Backer under preparation)

II. Focus on the production of spatial datasets
(predominantly issue 4 Spatial analysis)

1. Small picture process description (Paper by Geostat 1B partners under preparation)
2. Small Picture quality assessment (Paper by Geostat 1B partners under preparation)

In this light, the author have three reports under preparation:

1. A "big picture" description of production process(es) for statistics. (The current "overview")
2. A "big picture" approach to production processes issue by issue
3. A "big picture" approach to a quality assessment of the production process.

Stockholm Monday, Thursday, 13 December 2012

Lars H. Backer
lars.backer@mdmapping.se
Valhallavägen 145
115 31 Stockholm
Sweden

THE BIG PICTURE

This paper considers the production of spatial statistics as a part of a larger process that is started by the so-called UN Seoul declaration that marked the start of a process to build a GGIM system of information that may reach over three levels.

1. Global Geospatial Information management system (GGIM)
2. European Geospatial Information management system (EGIM)
3. National Geospatial Information management system (NGIM)

The EFGS has argued consistently that this hierarchy of statistical systems should be built according to a real object approach that is based on the integration of geography (or rather cartography) and statistics. This means that a "big picture" Geostat perspective must relate its effort(s) to this work. The production a Geostat dataset for this scheme as part of the upper three levels of a global hierarchy of statistical systems. This hierarchy consists of the NSS (national statistical systems), the ESS (European statistical system) and the ESS (European statistical systems. In this light please consider the "upper" and "lower" strata of the 6 layer hierarchy of statistical systems below.

1. Lower levels
    1.1. Local Statistical system (LSS)
    1.2. Municipal Statistical System (MSS)
    1.3. Regional Statistical System (RSS)
2. Upper levels
    2.1. National Statistical System (NSS)
    2.2. European statistical system (ESS)
    2.3. Global statistical system (GSS)

From a bottom- up perspective the foundation for all statistics is based on data produced by the individual NSI's. It is however, our hypothesis that the different items on each level are in general terms (in their outline) repeated as we go from bottom to top. If this holds, then it we suffice to discuss the whole system based on this general process description.

1. Data model
2. Data capture
3. Object- based statistical databases
4. Non- spatial, Spatial and Temporal- analysis (Production of geostat datasets proper)
5. Compilation of National, European or Global Geostat dataset(s) for dissemination
6. Dissemination
7. Feedback on real user needs
8. Overall quality assessment

It is also our hypothesis that the same scheme may be used for quality control of the whole product/ production process both regarded as an integrated whole, and a kit of parts.

THE NARROW PICTURE

The small picture description of the production of spatial statistics is in the focus of the Geostat projects. This perspective will discuss the issue 1) from a point of view that focuses on the production of a European grid dataset as a part of the European statistical system (ESS). On the other hand it will focus 2) on the part in the general process that in our scheme is described as 4. Non- spatial, Spatial and Temporal- analysis. This issue is described in another report from the Geostat project.

WP1_A Geostat 1B production process_1.docx (Bloch, Vilni Verner Holst)

"BIG PICTURE" PROCESS DESCRIPTION

The following process overview will be discussed issue by issue in more detailed report currently under preparation. This description will serve as a foundation for a two layer process description and a two level quality assessment.

THE PRODUCT

1. <u>Modelling</u> (of a dataset in response to real user needs)
   The building of a statistical system must be regarded in the perspective of an integrated system of information needed for all levels of actions from local to global. This modelling effort will require the active cooperation between a" more practical science" and a "more scientific practice". The modelling effort will for obvious reasons have to be based on the object approach. Its scope must be holistic. This means that be useful for the description of the Earth as an integrated man environmental system. It must respond to the observation that we experience our world as a 4D universe, an interaction of objects in space and time.
   1.1. Modelling as product
       The result of cooperation between a more practical science and a more scientific practice
   1.2. Modelling as a process
       The modelling of man environmental systems as a process of continuing development
   1.3. Quality control of the modelling product and processes
       1.3.1. Output (product)
       1.3.2. Production Process

THE PROCESS

2. <u>Data capture</u>
   2.1. Methods used for data capture
       2.1.1. Registers
           i   Data capture tools and methods used in register systems
           ii   Critical production (quality?) bottlenecks
       2.1.2. Census- based systems
           i   Data capture tools and methods used in census based systems
           ii   Critical production (quality?) bottlenecks
       2.1.3. Sampling
           i   Data capture tools and methods used in sampling- based systems
           ii   Critical production (quality?) bottlenecks
       2.1.4. Other
           i   Data capture tools and methods used in sampling- based systems
           ii   Critical production (quality?) bottlenecks
       2.1.5. Mixed systems for data capture
           i   Mix 1
           ii   Mix 2.
           iii Etc.
   2.2. Cleaning (correction) of raw data
   2.3. The production of statistical data tables
       Building of Statistical database tables for integration national data into the NSS.(the object based statistical databases described in next section 2 below.). On the European level this process involves the building of the ESS from compiling information provided by the NSI's. On the European level this process involves the building of the GSS from compiling information provided by the NSI's as well.

WP1_A Geostat 1B production process_1.docx (Bloch, Vilni Verner Holst)

2.4. Quality control of the data capture processes
    2.4.1. Output (product)
        i   The statistical data tables
    2.4.2. Production Process
2.5.
3. Object- based statistical databases (Microdata database system)
    3.1. Non- spatial attribute data (attributes to describe object according to model requirements)
    3.2. Spatial data (coordinates for observations)
        3.2.1. Smallest possible spatial reference (Point- Based foundation)
          i   Street Address
          ii   Building
          iii   Real Estate unit
          iv   Census are or similar Municipality sub- division
    3.3. Temporal data (timestamps for observations)
       (Not discussed here)
    3.4. Quality control of the Microdata system
        3.4.1. Output (product)
        3.4.2. Production Process
4. <mark>Non- spatial, spatial and temporal analysis</mark>
    <mark>This part of the over- all production process is the focus of the Geostat effort to describe the processes used for the production of Geostat datasets.</mark>
    4.1. Analysis of non- spatial object attributes (according to conventional statistics)
       (Not discussed here)
    4.2. Analysis of the spatial dimension of object attributes (focus on the spatial dimension)
        4.2.1. Introducing new grid data sets
          i   Bottom- up procedures
- Table aggregation method
  - Method 1 (SAS method)
  - Method 2 (Mapinfo script method)
  - Other.
- GIS spatial method for aggregation
  - Method 1 (ArchGis method)
  - Method 2 (Mapinfo method)
  - Other

          ii   Top- down procedures
- Disaggregation method 1(Corine data)
  - Method 1(JRC method 1)
  - Method 2 (JRC method 2)
  - Other
- Disaggregation method 2 (soil sealing)
- Etc.

          iii   Hybrid procedures
        4.2.2. Transformation of existing grid data sets
          i   Data transformations
- Transformation method 1
- Transformation method 2
- Etc.

    4.3. Temporal analysis
       (Not discussed here)
    4.4. Quality control of the Microdata system
       1.1.2.Output (product)
       1.1.3.Production Process

5. <u>Compilation of National, European or Global Geostat dataset(s) for dissemination</u>
    5.1. Contents in response to real user needs
       5.1.1.Data for public authority use
          i   Indirect action (information required for the design, implementation and evaluation of national laws)
          ii   Direct action (information required for the design, implementation and evaluation of direct action (e.g. the building of physical infrastructures)
       5.1.2.Data for private sector use
          i   Data needed e.g. for production process
          ii   Data needed e.g. in the products themselves
    5.2. Quality control of the dataset compiled for dissemination
       5.2.1.Output (product)
          i   Control of the quality of the dataset (compared with promised specifications)
          ii   Confidentiality control (or compliance with other institution regulations, rules or national laws)
          iii   etc.
       5.2.2.Production Process

6. <u>Dissemination</u> for national and international use
    6.1. Spatial data infrastructure (for spatial statistics) for dissemination of statistics over the internet
    6.2. Business model
    6.3. Quality control of the Dissemination system
       6.3.1.Output (product)
       6.3.2.Production Process

7. <u>Feedback on Use (in response to "real" and "assumed" user requirements)</u>
    7.1. Public authority use
       7.1.1.Direct action
       7.1.2.Indirect action
    7.2. Private sector use
       7.2.1.Data needed for production process
       7.2.2.Data needed in the products themselves
    7.3. Quality control of the Feedback system
       7.3.1.Output (product)
       7.3.2.Production Process

"BIG PICTURE" QUALITY ASSESSMENT

This product and process is as all iterative processes dependent on a highly dependable and constantly learning quality control. This concerns both the quality at a given point in time (a product) and the quality of the single parts of the process item by item.

1. Quality Control
    1.1. Quality control issue by issue
        1.1.1. Quality control of the modelling product and processes
            i   Output (product)
                • GGIM, EGIM or NGIM component
                • Product designed for private sector consumption
            ii   Production Process
        1.1.2. Quality control of the data capture processes
            i   Output (product)
                • The statistical data tables
            ii   Droduction Process
        1.1.3. Quality control of the Microdata system
            i   Output (product)
            ii   Production Process
        1.1.4. <mark>Quality control of Non-spatial, Spatial and Temporal analysis</mark>
            i   Output (product)
            ii   Production Process
        1.1.5. Quality control of the dataset compiled for dissemination
            i   Output (product)
                • Control of the quality of the dataset (compared with promised specifications)
                • Confidentiality control (or compliance with other institution regulations, rules or national laws)
                • etc.
            ii   Production Process
        1.1.6. Quality control of the Dissemination system
            i   Output (product)
            ii   Production Process
        1.1.7. Quality control of the Feedback system
            i   Output (product)
            ii   Production Process
    1.2. Overall Quality control
        Assessing the quality of the whole process.from modelling / data capture to Use and evaluation of how the data satisfies real user needs.
        1.2.1. Output (product)
        1.2.2. Production Process