

Basic Spatial Data According to User's Needs Aspects of Data Quality

Elissavet Pontikakis and Andrew Frank
TU Vienna, Institute of Geoinformation and Cartography

Foreword. This paper was written as part of the European Territorial Management Information Infrastructure project. It is reporting on a workshop aiming in identifying user's needs on base geographic data in Austria, Germany and Switzerland and it has been part of the project's position papers.

ABSTRACT: The collection of information about potential use is extremely difficult, because there is no systematic way to ask the 'non-users' and the 'not-yet' users why they are not using geographic information. As part of the European Territorial Management Information Infrastructure (ETeMII) the Institute for Geoinformation of the Technical University in Vienna Austria conducted a German workshop on user's needs for basic spatial data in Salzburg, Austria in 2000. In addition to the workshop report where, as indicated by the participants, the need for metadata and policies is mostly stressed, we offer in this paper some observations collected through discussions with various providers, consultants and software companies from the German speaking community.

1. Introduction-Description of problem

The systematic collection of information about why a certain product or service is not used is among the most difficult tasks in social statistics and marketing. Actual users can be identified and addressed, but it is nearly impossible to identify the potential users, who do not or not yet use a product. The option of surveying/asking everybody is naturally excluded for its high and unjustifiable cost. It is therefore necessary to devise alternative methods in order to collect information about impediments and potential markets. We had various contacts with

producers, distributors, consultants, and software companies in the geographic data field and we were present at numerous discussions about potential uses of geographic information (GI). One can learn from both: realized and not realized initiatives and projects. In general, such information is proprietary and not widely shared – another reason, why public discussion does not get easily at the heart of the problem. This document offers a synthesis of the information we have accumulated.

We will first point out a change in the character of the GI market, which we observe as beginning now and which will have a fundamental influence on the use and yielded quality of geographic data in the future. Then, we will describe the effects on the data requested and justify our opinion using the GI market in the United States as an example. Many of the commercial players known to the GI groups, are not yet aware of the change, which is just appearing now. It could also be that these players are not yet prepared to participate in this new GI market and cannot therefore adequately report about the impediments which stand against this development.

2. Big GIS vs. Small GI

A geographic information product has value because it improves decisions. Small GI products are driven from a large market. They are attributed the name Small GI because the amount of information sold for each individual case is very small irrespective of scale. This by no means implies that the GIS are smaller or require less data collection and management. The difference lays in the distinction between the party which develops and operates the GIS, and the user of the geographic information produced (Frank and Raubal 2000). The future of GI is in Small GIS project, because it is market driven. Naturally, this change in the focus of the GIS business has substantial effects on the demand for data. It affects both the data elements and the ways in which data are requested.

We can start from the known statement that 80% of all decisions are somehow related to space (Albaredes 1992). The contribution of Geographic Information can reach 16%. A rough 20 % of improvement for 80% of all decisions is 16% of Gross National Product. This is not a

short-term goal, but it can be achieved over the next 20 to 50 years. We argue that the data distribution and the business around it needs substantial reorganization.

3. A small set of data is at the core of GI usage

In this section, we will concentrate on the new applications and not discuss the traditional GIS for cadastre, public utility, administration and planning etc. We assume here, that these new, rapidly growing markets will become much more important than the current GIS markets.

A very small set of data is at the core of a very large number of applications, namely:

- street addresses,
- yellow pages,
- street network and driving rules,
- postal codes,
- administrative boundaries, and
- demographic data.

Most applications from the commercial side start with the street address as the primary location indication. Additionally, the use of location measurements from GSM or UMTS mobile phones is important. In both cases, an efficient and highly usable software and data to convert between street address and coordinates is highly desired. The availability of such services – either web based or distributed and installed at the users' site with regular updates – is crucial for the development of a large number of applications. This requires street addresses which are complete and correct, and regularly updated to include new streets. Resolution must be within a few meters, to differentiate between different doors of buildings etc. The semantics of such systems are non trivial, as a single interface for all of Europe (preferably the globe) must be capable of handling the many local specialties in street addresses. There are not only any conceivable way to number streets but also the concept of building varies between countries.

In Austria for example all of the buildings in the 1200-meters long, 1325-apartment Karl Marx Hof -- are a single street address which is serviced by five monumental archways and a large number of

consecutively numbered entryways. As a result of this particularity of the Austrian addressing system, many temporary residents in Vienna who live in compounds never learn their real home address since they never learn that the street number of their address is not the number that they see outside their entrance door. In other countries the same building may have two or more street addresses, while landmark buildings without street address are a common fad of modern business centers.

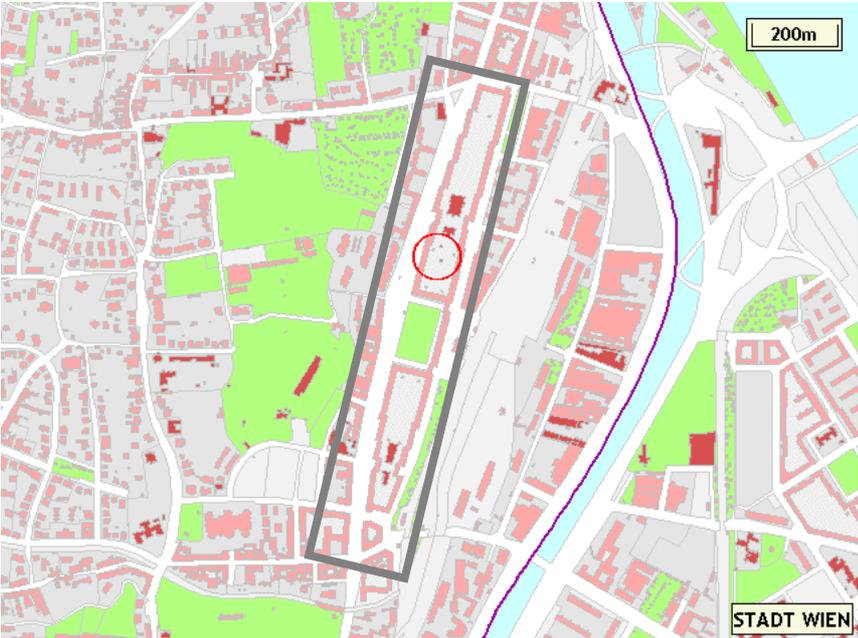


Figure 1. Karl-Marx-Hof building in Vienna, Austria.
Source: Wien online, Stadt Wien, Wien-Grafik Redaktion



Figure 2. Karl-Marx-Hof building in Vienna, Austria-View of the court area.
Source: 1999 Vitruvio Architecture in the Web

Next in importance are the Yellow Pages, which connect the street addresses to activities and thus are a key for guiding people to services. This is typically a private business, but often related to the telephone directories etc.

The growing market of navigation and guidance systems targets people's need to find a location of interest in the real world. The development of such systems combines the use of street network information elements, such as localized rules for driving and pedestrians walkways, with Yellow Pages and location based services. The same system combination is also used in the development of applications that determine the effectively felt distances, which influence a user's choice for a number of services.

Administrative boundaries are a small but very important element. They are used by geomarketing in combination with geo-demographic data collected by the statistics service bureaus. In the same group, one must place the postal codes, which unfortunately do not often have clean spatial boundary, because they are not defined as areas, but as service routes (Rhind, Raper et al. 1989).

For many commercial applications, the use of the demography of an area is becoming increasingly crucial to the decision-makers. Detailed knowledge about the population, age groups, average income etc. in context with the area they occupy is widely used in geomarketing. There

are currently commercial services which use official statistics and other sources to establish detailed and properly classified data. Such data is available for a fee. The three mostly requested geomarketing services as ranked by Experian Ltd, UK and Microm Bureau Services are: customer profiling, area ranking/profiling, site selection and shortest route to nearest dealer (GISMO - Verwoert 2000).

4. Data sets of secondary importance

Widely used but of less importance to non-specialized market are the following geographic data:

- topography,
- stream networks and watersheds, and
- land use and land cover.

A number of technical applications, but also applications from agriculture and forestry need topography with a high resolution of 1 m. Watersheds and stream networks belong to the same collection and the two datasets must be coordinated. Although such data are extensively used by environmental institutions in most phases of environmental analysis, are often generally used as backdrop to locate other data by a more general market. This usage as widespread it is, contributes little value and is 'nice to have' at a low price.

Land use and land cover are two different, but very similar concepts. Similarly to topographic and watershed and streams network data, the land use and land cover data are extensively used by environmental and planning institutions. Often, the land use and land cover analysis is a key to the formulation of environmental regulations that are related to a particular area. This characteristic makes this type of data very useful, however, in most cases this data is seen only as an 'add on' to help illustrate a situation for the general market. Few applications, especially technical applications from the telecommunication side, use land cover data. One particular burden in the use of land use and land cover data is the fact that they are often subject to interpretation which varies depending on the land use semantics used to classify the land use types. Another impediment is the currentness of the data set. As a result to the classification or "version", land use and land cover data which are

collected at some point of time are not suitable for many other applications.

5. How to assess importance

We suggest that a serious prospective investigation is made to assess the tasks such data can be used for and then assess the total number of uses multiplied with the value each such usage creates for the user. We have found that such assessments – per tasks – can be made following standard methods for the development of business plans; it should be possible, to classify tasks into groups of similar importance and then estimate totals. The identification of the value of data requires understanding of user needs. This way the data producer is able to focus on the data features which bring additional value to the user (Krek and Frank 2000).

Quality is a crucial factor of the use and consequently of the value of the data. Some important components of quality of geoinformation products are accuracy, consistency, completeness, currentness. It is however often the case, in geographic information as well as in the general market, that limited (lower=compromised) quality yields wider but not necessarily better use (Krek and Frank 1999).

Compare this picture with the suggested impediments by the traditionally accessible user and producer groups:

The view expressed here contrasts with the information collected from current or close to current users. The users with whom we can discuss their data needs are closer to the BIG GIS applications, are typically more technical, and desire many more data sets. We argue here, that while this is certainly true, and these potential users are important, this is not the largest potential user group for GI in the future. More important is the unmet demands of the very large number (essentially everybody) who could benefit from geographic data services provided as Small GI.

To justify our view, that a very small number of widely used datasets are the bulk of the future GI business, we offer an analysis of the US experience. Over the past decade a very sizable use of GI has emerged in the US and with some delay the same appears in some parts of Europe. GI is widely used for logistics, for car navigation, tourism etc. It is not

often sold directly to the public, but it is used for advertising, as part of service of other companies etc.

This market emerged with the advent of the TIGER file. The U.S. Census Bureau's TIGER System aims in automation of mapping and other related geographic activities required to support the census and sample survey programs of the Census Bureau starting with the 1990 decennial census. The Census TIGER System provides support for the: creation and maintenance of a digital geographic database; production of maps; and assignment of individual addresses to geographic entities and census blocks (ESRI: TIGER Documentation, 1995 TIGER/LINE, U.S. Census Bureau). The system provides complete coverage of the United States.

The wide availability of geographic information at a very low cost, even when the quality was relatively low, made it possible for companies to experiment with the use of GI in various applications. Once novel applications were established, users realized the additional benefits of quality in the data. Investment in improving data quality in special directions became economically viable. A large multi-billion US dollar industry was born – out of two data sets.

In these initial phases, little other data was available and required, and very little other data was available free of charge – local data is typically not included in the ‘no-copyright’ policy of the U.S. Federal Government.

6. The situation among the German speaking countries

From the results of the 2000 Salzburg workshop ‘Basic Spatial Data According to User’s Needs’ and the ‘Kartenwerke in der Bundesrepublik Deutschland’ provided courtesy the Landesvermessungsamt (NMA) Nordrhein-Westphalen in 2000, we conclude that completeness of data is a quality factor that is definitely missing. The lack of completeness becomes especially dominant when we are looking into European-wide reality. The implementation of ETeMII’s vision for minimum level of pan-European geographic data will boost the momentum of the GI market in Europe. This will occur with the development of GI applications for areas which are currently being utilized but also for those where gaps in the geographic data prohibit their exploitation. It

seems from this perspective that few core data sets available for all Europe is more important than many data sets which are not complete.

7. Conclusions

The observations of desirable data sets collected from the currently known users of GI are just one aspect of the picture. It must be completed by a prospective assessment of the needs of a very large number – if not everybody – of potential users, which could benefit from Geographic Information Services of various kinds

It is not possible to invite these users to meetings and have them explain their needs – they are interested in services which are useful, not in geographic data. The availability of geographic data to the – not yet existing and therefore not heard – providers of the services is crucial.

Data sets necessary, are – in estimated order of importance:

- street addresses,
- yellow pages,
- street network and driving rules,
- postal codes,
- administrative boundaries, and
- demographic data.

It is not only the availability of the data – and we speak of updated data – but also the form under which it is available crucial for this use of GI. The organization of the business must be streamlined. Central points which mediate commercial access to a large collection of data for very large areas (preferably all of Europe) is important, as it reduces the cost for organization of the data providers.

An analysis of the situation in the U.S.A. and the developments over the past 15 years justifies this viewpoint: from a few commonly available datasets an industry has sprung with a multi-billion US dollar turnover. Even today, nothing comparable is visible in Europe. The benefit of this novel GI industry is, however, mostly to the millions of users of these services – from hotels to business and leisure travelers: availability of geographic information makes many tasks much more efficient.

Acknowledgements

This work has been supported by the project ETeMII sponsored by the European Commission.

References

- Albaredes, Gilles. A New Approach: User Oriented GIS. Proceedings of EGIS '92. Munich: EGIS Foundation, 1992. 2: 830-837
- Frank, A. and M. Raubal. (2001) "GIS Education Today: From GI Science To GI Engineering." Journal of the Urban and Regional Information Systems Association (URISA). 13(2): 5-10.
- GISMO - Verwoert, C. (2000). GISMO White Paper, ESPRIT project.
- Krek, A. and A. U. Frank (1999). Economic Optimization of Quality for Geoinformation product. AGIT 99, Salzburg.
- Krek, A. and A. U. Frank (2000). "The Economic Value of Geo Information." Geo-Information-Systeme - Journal for Spatial Information and Decision Making 13(3): 10-12.
- Rhind, D. W., J. F. Raper, et al. (1989). First UNIX, then UGIX. Auto-Carto 9, Baltimore, MA, ASPRS & ACSM.