GEOGRAPHICAL DATA SUPPLY FOR LOCATION BASED SYSTEMS (LBS) AND WEB APPLICATIONS

Mati Tee
mati.tee@regio.ee, matitee@ut.ee

Tanel Ilves
tanel.ilves@regio.ee, tanel.ilves@ut.ee

Estonia, Tartu Riia 24 51010, AS Régio

Abstract

A location-based service (LBS) is an information service, accessible with mobile devices through the mobile network, enabling people to position mobile phones in geographic space. LBS enable also to send custom advertising and other administered information to mobile device users based on their current location. LBS and Web Maps (WM) are applications that utilize geographical data to provide additional value to a user of the service. The quality of this “additional value” is dependent from service provider’s knowledge and administration about place of interest. LBS and map services in Internet need a permanent feeding with up-to-date and accurate geographical data.

The LBS and WM operate with large-scale maps (up from 1:30000), that are very time sensitive. The average ageing per year for 1:10000 scale topographic data is approximately 4%. This value is calculated for whole Estonia, for rapidly growing suburbs areas it may be ten times higher. But LBS and WM weren’t only the applications to show map images. Great part of them was based on address data usage (geocoding and reverse-geocoding), searches in points of interest (POI) datasets (e.g. “find nearest neighbor”) and route calculations on navigation databases. The ageing rate for addresses, POI’s and traffic rules with road descriptions is much higher, so the frequent updates are needed.

Presently there are thousands different spatial data sources for each country: public sector, private companies etc. Many of them provide information related to LBS and WM, but few manage the complete set of geographic data required for Service provider. Commonly the geographic databases were built to match specific purposes – by this reason the data collecting and processing requirements between them were different. The LBS/WM GIS integrates the various spatial datasets into one collection without duplication, with seamless matches between corresponding objects, without any gaps or divergence. Selection criteria for data sources, logical data model for integration and processing (and a presentation model for end-user, where needed) are imperative for the development of LBS/WM.

Quite often the LBS and WM providers (network operators) aren't the geographical data producers. By all means, there is need to organize a dataflow into service providers information system. This dataflow must guarantee the regular or permanent updates with high quality of data content. From process point of view this denotes an intensive communication between different organizations and their parts.

Regio/Reach-U has been in role of data supplier for LBS/WM applications for ten years. Our experience shows that *a specific geographical data model* for LBS and Web applications is very helpful as an instrument of communication. We developed out a geographical data management system (Reach-U GIS-Builder) that consist specifications, models, tools, know-how for spatial data handling and global network of data resources.

Reach-U GIS-Builder as service can act as a clearinghouse for aggregating maps from different map provides. For each data request, datasets can come from either Reach-U GIS/Cartography department or from external provider. The decision can be made based on required detail level (general data from one provider, detailed data from other), map content (map images from one provider, aerial images from other), or territory (map images from country X from one provider, for other areas from other provider).

Introduction

From location-based services (LBS), a fundamental benefit comes in form of information products (IPD). IPDs from location based services may be: maps, lists, graphs, reports, messages, multimedia files etc. An information product is data transformed into information (Tomlinson, 2007) particularly useful to somebody – for example, economic data analyzed in relation to a specific location – and delivered to user. For LBS this delivery happens via mobile network and mobile devices. The geographic data is the raw material for LBS's information products. The data quality is crucial for services quality. From end user point of view geographical data must be rich an up-to-date. Other quality aspects for LBS data content are richness, relevancy, thematic and temporary accuracy. By this means a well organized supply with data is sign of vitality for LBS as information system.

Regio's experience with LBS

AS Regio (Regio Ltd.) has been a key player on Estonian cartography market for twenty years. Founded as first map publisher in Estonia, AS Regio quickly transformed from paper maps producer into information systems creator and software company. The Regio brand was still used for hard-copy maps, survey and geographic database products, for software solutions and web-applications a Reach-U brand was developed out. Close ties to University of Tartu – the Regio started as spin-off enterprise – were helpful in technology change: in transition to the all-digital production process.

University of Tartu was the pioneer for digital cartography and GIS in Soviet Union (Koshkarev, 1999). A scientific conference "Problems of geoinformatics", held by University's geography department at 22.-23. September 1983 in Kääriku (Estonia) may be the first public demonstration of digital map in former Soviet bloc (Koshkarev, 2003). More wide attempts of digital cartography in Estonia were made at the end of the 1980s, when the computerizing of society began. After collapse of USSR, the digital maps and databases found practical use as tools for land (property) reform (Tee, 2007).

For Regio, the technology transition into all-digital happens between 1994-1996. The digital datasets give ability for new products: a web-map "*Regio Veebikaart*" (1997) and "*Regio CD-Atlas*" (1998). Next fundamental breakthrough comes with millennium when Regio Ltd. under Reach-U brand entered into global LBS market. Nowadays, our services run in approximately twenty countries around the world, for example in Argentina, Mexico, Ukraine and Saudi Arabia.

This paper summarizes our experience in this kind of business.

Main problems of data supply

The data quality is crucial for services quality, it is fact. The end-user wants a service on the base of rich and up-to-date datasets. This is a fact, too. But another essential part of quality of service is prize. This led to key question: *how good is good enough*?

Ten years ago, at the standing of our web-map, we see ideal in on-line service with permanent up-dates from our main database. This seems to be optimal solution for Regio as to the owner of one of the biggest database (DB) for Estonia and Baltic region. This point o view made significant evolution after Regio's entrance into global LBS market.

First discovery is that there wasn't fresh and ready-for-use geodata for LBS applications on the global market. A big international data vendors and producers, for example Teleatlas and Navteq, are oriented to supply navigation devices. Local provider's datasets are heterogeneous and quite often didn't contain all needed content.

Second discovery is that in reality the biggest available data content refresh rate is once per quarter. These discoveries lead us to do R&D work to specify purposes for data component in our LBS system. Questions to answer are:

- What is necessary in a minimum "must be" set of geographic data?
- How accurate it must be?
- How often it must be updated?
- Which kind of structure is easy to integrate heterogeneous datasets and to do partial refreshing?

- How to communicate our needs to different people from data vendors to software developers and support team?

Answers to this questions result a *Reach-U GIS-Builder*, a geographical data management system, that consist specifications, models, tools, know-how for spatial data handling and global network of data resources. R&D works that were performed 1997-2001 and 2004-2006, give us understanding about requirements for LBS as specific GIS geodata structure. Following usage of system shows a growing role of *a specific geographical data model* for LBS and Web applications as a communication tool between different groups of interests in LBS business.

Reach-U GIS-Builder

Reach-U GIS-Builder is comprehensive package of spatial information services:

- GIS-Engine, Software components and developing system
- Data Management, know-how for GIS data handling and global network of data resources
- Map design skills with wide practice in map appearance.

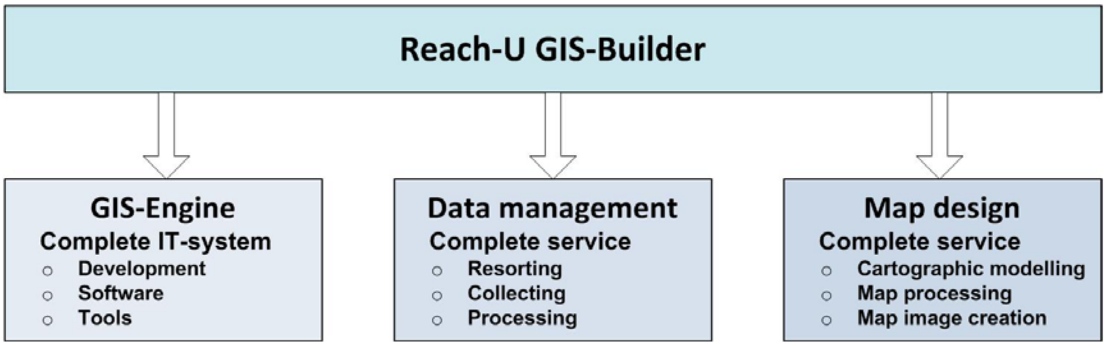

Figure 1. Reach-U GIS-Builder.

The GIS-Builder package is available either as a complete set (described above) or as standalone services. Practical training and support is available for all services and tools.

Reach-U GIS-Engine architecture

Reach-U GIS-Engine is software toolset for delivering GIS and mapping functionalities, designed specially to suit the unique requirements of LBS and web-GIS. The main functionality includes superior map image presentation/rendering for LBS/web-GIS Internet/WAP/MMS pages, geocoding and reverse geocoding (converting from coordinates to place name/address and vice versa), “find nearest” search and route/traffic assignment.

Reach-U GIS-Engine has support:

- to different middleware’s
- to diverse applications
- to various map projections
- to Multilanguage usage.

Reach-U GIS-Engine components are (Fig. 2):

- MapEngine – Minnesota MapServer + Reach-U FlashTile Technology + scripts for automated map image generation and indexing
- Geocoding service:
    - JGC – Java Geocoder:
        - Geocoding and address search tool
        - Reverse-geocoding
        - Rich content search
        - Usage of spatial ontology algorithms
    - Geocoding tool:
        - Mass geocoding
- PoiEngine – service for points of interest handling:
    - POI search
    - POI management interface
    - Rich content POI’s
- RoutingEngine – tool for route/traffic assignment and shortest/fastest path choice.

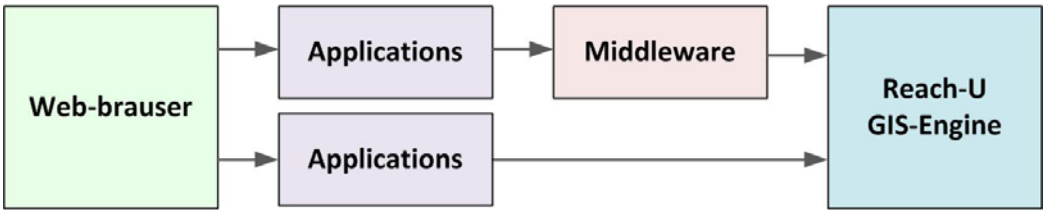

Figure 2. Reach-U GIS-Engine in LBS system architecture.

GIS data management

Reach-U GIS-Builder service can act as a clearinghouse for aggregating maps from different map provides. For each data request, datasets can come from either Reach-U GIS/Cartography department or from external provider. The decision can be made based on required detail level (general data from one provider, detailed data from other), map content (map images from one provider, aerial images from other), or territory (map images from country X from one provider, for other areas from other provider).

Map Design

Map design began from describing of map concept:

- map scale
- projections
- label placement.

After that map layers structure and map elements are created. Map conceptual designing ends with creation of map presentation template.

Data design for LBS

Spatial data is relatively complicated, so the rational data design is major factor for success. Reach-U LBS applications do not use map data directly. Maps and other geographic data are delivered to LBS by map/GIS server on per request basis. Every LBS has different requirement for the data, and the map server must satisfy all needs of all LBS applications.

To orchestrate data needs for all three Reach-U GIS-Builder components, the coherent data design was developed. In general, Regio/Reach-U LBS geographic data are divided into 5 components: maps, addresses, points of interest and routing data (Fig. 3). A "special information" can be added as independent thematic layers. Depend on the kind of services some of components may miss. Main advantage of this design is flexibility and possibility to integrate data from different sources.

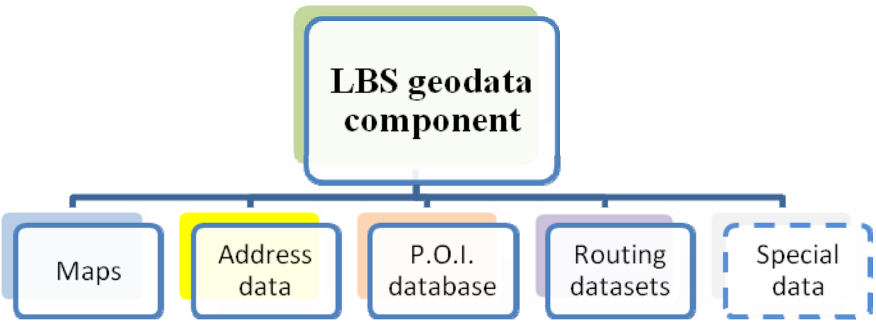

Figure 3. Data components for Regio/Reach-U LBS.

In main configuration, the maps were designed and created via technique of image pyramid and hash indexing (Kervi, 2007). Large map image is first extracted into different levels of detail to form a hierarchical structure or pyramid. Each level image is further cut into pieces or tiles. When the user is requesting data, image is combined of the tiles inside the area of interest and transported to the client side. Using pre-generated images instead of generating map from spatial data on demand gives a considerable performance improvement (Chaowei *et all*, 2005). So, the most static component of LBS geodata component is set by technology.

The routing data, traffic conditions are labor-intensive datasets. Delivery of navigation data requires careful testing for logical and contextual consistency. Obviously the street-road network has close ties to other map layers (aligning) so it is rational to do synchronized updates for map images and routing data.

Most flexible are address and points of interest datasets. House numbers and objects symbols can by created from database and integrated to the map image by map server utilities.

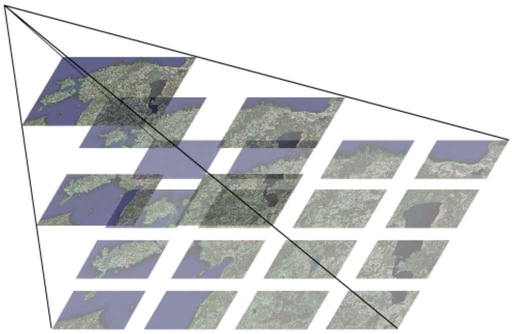

Figure 4. A pyramid constructed from Estonian satellite image (Kervi, 2007).

Logical data model for LBS

A logical data model describes those parts of real world that concern our customers. Choosing applications to use, the customer figure out a *minimal set of geographic data*, necessary to feed the services (Table 1).

Table 1. Data requirements for various location-based services (sample).

Application	**Short description and requirements for the data**
FriendFinder	Finds location of other people (usually per request) **Street-level maps** for showing the location on graphical map (web – big image, WAP and MMS – small image) **Address data** for giving place name of located person (or approximate address). Application has SMS, MMS, WAP and web interfaces.
Mobile Map	Basic application which uses user's own location information: Finds nearest points of interest (e.g. businesses), send my location to friend, find addresses/placenames, find shortest routes. Detailed **building-level maps** for guiding the way to the business. **Address data** for address search (to find businesses at manually given locations). **Placename/address data** for giving user location as short text. **POI data** with rich set of attributes (e.g. opening hours, phone numbers, descriptive information). Optionally also **routing data** for calculating shortest/fastest route to business and for giving step-by-step instructions for that route. Application has SMS, MMS, WAP and web interfaces.

Spatial accuracy and precision levels

Measured circular error for mobile positioning varies from tens and hundreds meters in dense mobile network areas up to hundreds and kilometers in areas with thin mobile network. By this reason there isn't necessary to large scale maps (1:2000 and up) in LBS. In other hand, the mobile devices displays are small and had quite limited possibilities to show detailed map images. These two factors set up base requirements for LBS map data. According to this, our researches result map data dividing into three main precision levels:

- Overview map (low-level this level datasets are used for the map images generation at low-detail zoom levels. The supposed map scale for overview data is smaller than 1:1,000,000).
- General map (for rural areas): for the general precision level datasets the map features must be interrupted on the separation line between rural (less detailed) and urban (more detailed) areas (e.g. index box or administrative boundary). General map's features within urban map area boundaries must be cut off or have special attribute. For best results the index boxes are preferred. The optimal map scales are 1:1,000,000 to 1:50,000 for general maps.
- Street-level maps (for urban/city areas): the default standard implies that layers for urban area (this level of precision) contain more details than the general datasets. The levels of precision/accuracy (map scales) are subjects to negotiation. The optimal map scales are 1:30,000 to 1:10,000 for urban areas.

The corresponding map layers on rural and urban apportionments must be connected accurately and seamlessly to each other, without any gaps or divergence (e.g. rural main road must be connected seamlessly to the urban main street).

Necessary frequency of updates

The LBS and WM operate with large-scale maps (up from 1:30000), that are very time sensitive. There was too few known about data ageing for Estonia and other countries of interest. For Estonia, for example, we found out only one research on this theme (Gauk, 2008). So we had only pinch of base facts: (1) in soviet-era an average update frequency for large scale topographic maps was from 10 to 15 years; (2) average ageing per year for 1:10000 scale topographic data is approximately 4%; for whole Estonia, this estimate was used for Estonian Base map planning in 1990nties; (3) Estonian Land Board set 4 years as optimal time interval for updates of Estonian 1:10000 digital topographic database (Maa-amet, 2009). On this basis we began to build up our data supply service.

What Regio has experienced during ten years in the specific business is that in LBS services data tends to age during the time span of 6 Months - this is the time by what clients usually wait for updates.

On the other hand we have seen that clients don’t pay much attention to updating of data – this is reflected in the statistics of using the applications (old data leads to decrease of usage). And according to usage statistics every announce of newer data leads to increase of usage (not all will remain regular users, but some still).

The changes in Baltic states for street names for example go up to 1500 changed or new names per year, addresses count in the same category up til 16 000 per year. The same goes for POI’s – location, name, phone number etc, can and will change.

Data supply processes for LBS

In recent times Regio/Reach-U has been in need of even better and even more up-to-date data because of business awareness from clients – they value the power of data. We have experienced much trouble in sourcing data good enough from one vendor only. Off cause, there are thousands different spatial data sources for each country: public sector, private companies etc. Many of them provide information related to LBS and WM, but few manage the complete set of geographic data required for LBS provider. In cases like Pakistan, Morocco and Saudi Arabia for example we used two or three different sources for different type of geographical data.

Commonly the geographic databases were built to match specific purposes – by this reason the data collecting and processing requirements between them were different. The LBS/WM GIS integrates the various spatial datasets into one collection without duplication, with seamless matches between corresponding objects, without any gaps or divergence. Selection criteria for data sources, logical data model for integration and processing (and a presentation model for end-user, where needed) are imperative for the development of LBS/WM.

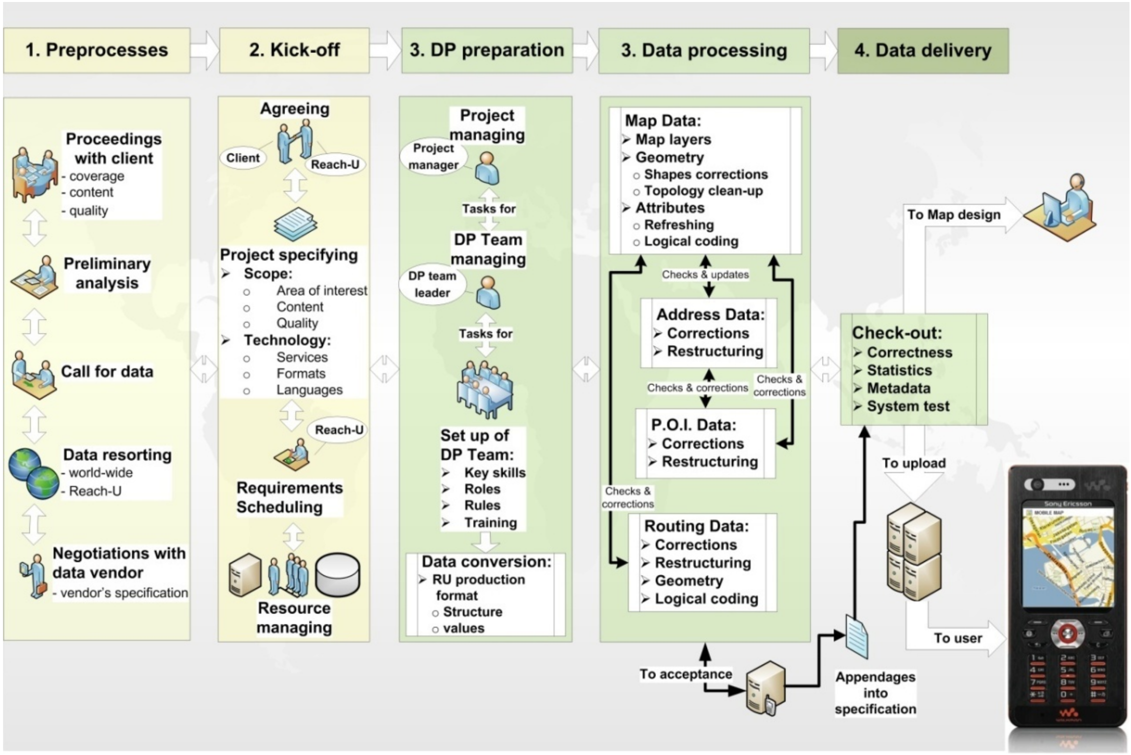

Figure 5. Reach-U Data Supply for LBS processes.

Quite often the LBS and WM providers (network operators) aren't the geographical data producers. By all means, there is need to organize a dataflow into service providers information system. This dataflow must guarantee the regular or permanent updates with high quality of data content. From process point of view this denotes an intensive communication between different organizations and their parts (Fig. 5).

Reach-U GIS-Builder as service can act as a clearinghouse for aggregating maps from different map provides. For each data request, datasets can come from either Reach-U GIS/Cartography department or from external provider. The decision can be made based on required detail level (general data from one provider, detailed data from other), map content (map images from one provider, aerial images from other), or territory (map images from country X from one provider, for other areas from other provider).

Either way – using one or several sources – the first contact and sorting out the best provider on market needs a lot of effort (if the country is not covered by any of the larger data vendors) to sort out the best quality for LBS. We have seen that only face-to-face communication leas to fast results.

This leads to problems with data quality and needed GIS works before any usage of the data is possible. Despite all efforts to offshore data producing to data vendor, we still need production capacities for geodata processing. Specifications don't help much if you are in need – and not in the position of demanding. Anyway – to wait is more time consuming then to act for ourselves.

Conclusion

The LBS and WM operate with large-scale maps (up from 1:30000), that are very time sensitive. But LBS and WM weren't only the applications to show map images. Great part of them was based on address data usage (geocoding and reverse-geocoding), searches in points of interest (POI) datasets (e.g. "find nearest neighbor") and route calculations on navigation databases. The ageing rate for addresses, POI's and traffic rules with road descriptions is much higher, so the frequent updates are needed.

Our research gives update rates for LBS datasets. These average frequencies were a superposition of customer's requirements, technology ability, data vendor's capability to refresh content, prizes etc. Optimal update rates are:

- Map images:
    - Overview maps: from 1 to 5 years;
    - General maps: from 6 to 24 months;
    - Street-level maps: from 3 to 6 months;
- Routing data: from 3 to 12 months;
- Address data: at least once 6 months;
- P.O.I. data: at least once in 6 months.

LBS data supply needs coherent efforts from different teams: marketing, managing, data vendors, LBS vendors, cartographers, software developers and database specialists. This is impossible without intense communications between different people. For this purpose, simple but pithy models are useful as communication tools.

Last, but not least, Regio's/Reach-U's ten year experience in the business tells us that without full GIS and cartography competence, without production capability, it is impossible to guarantee high quality LBS services for customers.

References

(Chaowei *et all*, 2005) Chaowei (Phil) Yang, David W. Wong, Ruixin Yang, Menas Kafatos, and Qi Li. Performance-improving techniques in web-based GIS. In *International Journal of Geographical Information Science*, Vol. 19, No. 3, pages 319-342, March 2005.

(Gauk, 2008) Martin Gauk, The change in completeness and thematic accuracy in Estonian Basic map based on four map sheets. Bachelor Thesis ("Eesti põhikaardi täielikkuse ja temaatilise õigsuse muutumine nelja kaardi-lehe näitel”), Tartu 2008.

(Kervi, 2007) Priit Kervi, Applying Map Tiling to Web-Based Mapping Applications. Bachelor thesis. Tartu 2007.

(Koshkarev, 1999) Aleksandr Koshkarev, History of geoinformatics I (Кошкарев Александр Владимирович, История геоинформатики.) Информационный бюллетень №4(21) 1999

(Koshkarev, 2003) Aleksandr Koshkarev, History of geoinformatics II (Кошкарев Александр Владимирович, История геоинформатики. Окончание.) Информационный бюллетень №4(41) 2003

(Maa-amet, 2009) Estonian Land Board homepage www.maaamet.ee

(Tee, 2007) Mati Tee, Historical maps in internet GIS solutions QVIZ: an example of community-based WEB-GIS. In: XXIII International Cartographic Conference 4-10 August, Moscow 2007, Russia. Abstract of Papers, Moscow, 2007.: XXIII International Cartographic Conference 4-10 August, Moscow 2007, Russia.. Moskva: Roskartografia, 2007, 233 - 233.

(Tomlinson, 2007) Roger Tomlinson, Thinking about GIS: Geographic Information System Planning for Managers. Third Edition. ESRI Press, Redlands, California, 2007.