MOBILE APPLICATION FOR ACQUIRING GEODATA ON PUBLIC TRANSPORT NETWORK

Lenka, ZAJÍČKOVÁ[1]

[1]Department of Geoinformatics, Faculty of Science, Palacky University in Olomouc, 17. listopadu 50, 771 46, Olomouc, Czech Republic

lenka.zajickova@upol.cz

Abstract

Following the detected deficiencies in managing geodata on public transport network in the Olomouc Region, potential solutions of the situation were analysed. On the basis of extensive research of global standards in public transport data standardization, a data model of entities and their attributes according to the needs of public transport management and control in the Czech Republic was established. Further, a data warehouse was created on the base of the newly created data model by means of the special software specializing in work with lines and network elements. A mobile application for a tablet to collect field data was designed to supply the warehouse. The concept of public transport geodata collection, management and updating was created in cooperation of Asseco Central Europe, Coordinator of the Integrated Transport System in the Olomouc Region and the Department of Geoinformatics of Palacky University in Olomouc in reaction to the unsuitable data situation in public transport in the Olomouc Region. Current filling of the data warehouse by means of the designed mobile application will later be followed by creating an interactive transport plan and spatial analysis based on the data about the managed area.

Keywords: data model, data warehouse, geodata, mobile application, public transport

INTRODUCTION

The issue of collection, management and long-time sustainability of current information about the public transport network in regions is, in conditions of the Czech Republic, dealt with especially by coordinators (organizers) of the integrated transport systems (integrovaný dopravní systém, hereinafter referred to as "IDS"). The majority of the information is – due to the nature of phenomenons which take place in space – related to spatial data; they are therefore predestined to be processed and stored in GIS. With respect to the expanding possibilities to publish and present digital data, there is an ever growing need to deal with currency, accessibility, particularity, good arrangement and accuracy of public transport (hereinafter referred to as "PT") data. Most of the IDS control bodies realize the importance and benefits of quality register of spatial network data in GIS. However, the absence of a unifying standard results in many different approaches, different data quality and incompatibility in case of their exchange. The question of data updating is also often discussed because, though the network is rather stable, there is a great number of objects to register and many of their attributes are subject to frequent changes.

The goal of this text is to introduce the conception and launching of the mobile application for collection of the PT network geodata in the area administered by the Integrated Transport System Coordinator of the Olomouc Region which is widely extensible and applicable in other regions. The partial goal of the text is to introduce the data model of the registered entities established on the basis of global standards accounting for the effort to standardize in the Czech Republic, newly established data warehouse respecting this data model and the updating method of this extensive database.

THE SCOPE OF AVAILABLE PUBLIC TRANSPORT GEODATA AND THE BASE OF THEIR STANDARDIZATION

The quality and scope of the PT network data differs in various regions of the Czech Republic. In some regions, the data are processed in GIS, in others graphic software (e.g. Corel Draw) is used; in regions served by carriers without central administration there are no geodata at all (Drdla, 2008). However, data warehouses of IDS organizers in GIS are often incomplete, the objects are not located correctly, the layers lack a coordinate system and errors like undershoots, overshoots and missing attributes frequently occur.

The logical structure is often inappropriate as the data are not based on a quality data model and also the way of data processing is not ideal.

The only present solution to keep the basic information regarding the movement of PT vehicles in national scope is the unified data format (hereinafter referred to as “JDF”) defined by the Decree of the Ministry of Transport No. 388/2000 Coll., on Regular Public Transport Timetables, and the Decree No. 175/2000 Coll., on Transport Rules for Public Railway and Road Transport of People (CHAPS, 2005). It is a pre-defined data format designed for regular PT road and railway carriers in order to process the timetables (hereinafter referred to as “JŘ”) for the purposes of the National Information System on Timetables (hereinafter referred to as “CIS JŘ”). The carrier processes the timetables in an electronic form, is responsible for them and submits them to the appropriate transport authority for an approval and to be advanced to CIS JŘ the management of which was delegated to CHAPS Ltd. (CHAPS, 2013) from 26th October 2001. However, no spatial information is related to the data inserted in the CSV files, attributes to individual entities are insufficiently registered and hardly anyone fills in the voluntary files. For passengers' needs, CHAPS allows not only searching for the appropriate connection but also the possibility of graphic display. It involves coordinate visualization of the stops on the base of a source map and their interconnection by straight lines. From the point of view of GIS, this solution is unsatisfactory and incorrect, however, CHAPS primarily deals with timetable administration and the visualization and spatial information are not the subject of their business.

The first attempt to implement a standard for PT data in the Czech Republic was the project called “The unified public transport data system with respect to a standard form application with the possibility to interconnect the existing systems to a unified software platform” (JSDV). The project was realized in 2011-2013 by the Transport Research Centre (Centrum dopravního výzkumu v.v.i.) in cooperation with CHAPS Ltd. and APEX, Ltd. The ground of the project consisted of the existing situation in public transport information systems, valid Czech and European legislation and first of all the Service Interface for Real-time Information (hereinafter referred to as “SIRI”) (CEN/TS 15531:2011, 2011). The general goal of the project was to support public transport competitiveness by implementing the uniformly arranged telematic system with a standardised interface which should have allowed later integration of other information systems of carriers, integrated transport systems and/or transport route operators/administrators. The partial goal was to create the central data warehouse and to prepare standardised data formats for data exchange. Further, the project should have used the CIS JŘ information to create a national information system in real time (so-called CISReal). The outcomes are the central information system, drafts of the architecture, feasibility study and the methodology for building up the information system. In course of the project, the Czech technical standard ČSN was being formulated (Centrum dopravního výzkumu, 2012). Nevertheless, the involvement in the project outcomes is still self-imposed; no binding duty is implied for the carriers and organizers.

DRAFT OF THE SOLUTION FOR ACQUIRING COMPLETE DATA ON THE PT NETWORK

Asseco Central Europe (hereinafter referred to as Asseco), Coordinator of the Integrated Transport System in the Olomouc Region (hereinafter referred to as KIDSOK) and the Department of Geoinformatics of Palacky University in Olomouc developed a solution to acquire complete GIS data on PT network for the needs of KIDSOK and defined the rules for their updating. The aim was to create a solution for data collection, processing and updating that could easily be repeated in any other region or area, keep the compatibility with the existing exchange formats and respect international standards in PT data standardization. KIDSOK detected incorrect topology in individual layers, insufficient scope of the attributes and poor structure of the existing data which was based on the analysis of the administered geodata, the scope and nature of the data for the needs of PT management and control. That was why a new data model was drafted for the PT network geodata, a way of supplying the newly established data warehouse was specified and rules for current data sustainability were defined.

THE DATA MODEL OF THE PUBLIC TRANSPORT NETWORK

The drafted data model is based on the extensive Network Exchange (hereinafter referred to as NeTEx) standard which is a new, still developing standard in the PT data standardization (PRE/CEN TC 278 WG9, 2012). It is based on the Transmodel V5.1 EN (12896:2006, 2006), on the CEN technical standard "Identification of Fixed Objects in Public Transport" (IFOPT) (EN 28701:2012, 2012) and on the SIRI standard. The aim of NeTEx is to provide pan-European standard for exchanging data from timetables and related data and information. It is a complex and extensive standard describing statistic elements of the PT network (stops, stations, access areas, equipment etc.) but also operational descriptive elements of the network (e.g. transfers). Only the parts applicable for Czech Republic conditions were selected, adapted and completed from this standard when formulating the data model. Compatibility with the JDF format was also taken into account. The initial concept was completed by code lists from the national level (JDF) and by KIDSOK specific requirements.

The basic element of the data model is a stop as a point feature and also an umbrella term for all other spatial entities (entities with spatial relation to the stop, see Fig. 1).

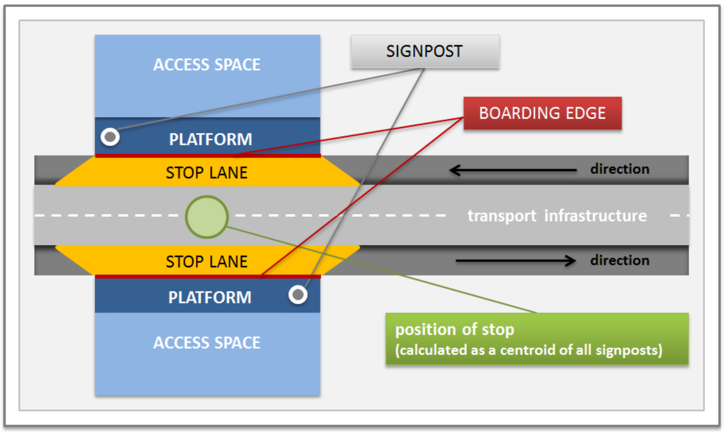

Fig. 1. Stop and its parts (Source: internal)

The stop is delimited by a boarding edge which is a border part of the platform and is defined as a paved area allowing safe movement of passengers when boarding or leaving a vehicle or waiting for a connection. Most of the stops have two boarding edges (one for each direction); larger transfer hubs can have more than two. Another adjoining entity is a stop lane delineating the area where the PT vehicles stop. Every boarding edge must contain at least one signpost, which is defined as a distinct upright marker designating bus, tram, trolleybus or other stop (Act No. 111/1994 Coll., 1994). The important change comparing to the existing entity registration is the rule for calculating the position of the stop which says that the position of the stop is calculated as a centroid of all signposts of the involved stop. Among outdoor captured spatial entities also belong other spaces and objects related to boarding edges or stops. These include especially the equipment or furniture of stops (e.g. shelters, concourses, waiting areas, benches, ticket machines etc.).

The result of standard research and element analysis of PT network is the data model of static elements respecting strict hierarchy of entities which are mutually interconnected by unique identifiers and which have various mutual relations (Fig. 2). The unequivocal identifier of a stop is the ID from CIS JŘ (e.g. 20580) which corresponds with the unique name of the stop. The code of the boarding edge consists of the stop ID in CIS JŘ + the index of the boarding edge sequence (e.g. NH1) + the orientation of the edge (e.g. N-S). The example of a unique code of the first boarding edge of the 20580 stop with the north-south orientation is 20580/NH1-N-S. The orientation of the boarding edge is derived from the position of the beginning (location of the signpost) and end of the boarding edge (the position where the vehicle enters the stop). If the PT vehicle enters the stop from the south and heads to the north, the orientation of the boarding edge is north-south (i.e. N-S). Possible boarding edge orientations are derived from cardinal points (N-S, S-N, W-E, E-W, SE-NW, NW-SE, NE-SW and SW-NE). Analogically, unique codes for IDs (20580/NH1-N-S/OZ1), stop lanes (20580/NH1-N-S/ZP1), stop equipment (20580/NH1-N-S/V1) and access spaces (20580/NH1-N-S/PP1) related to the boarding edge are derived from the boarding edge code. Tariff zones and parking possibilities are tied to the stop as a whole.

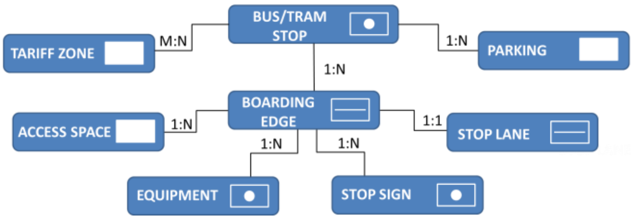

Fig. 2. The outline of proposed data model for data management (Source: internal)

DATA WAREHOUSE LIDS

The practical solution of the data model of transport network static elements is based on the LIDS geographic information system of BERIT, Brno division of Asseco (ASSECO BERIT, 2012), which offers wide functionality for registering spatial data (especially those related to networks), their maintenance, processing, analyses and evaluation. The GIS software solution of LIDS uses a three-layer architecture utilizing the XML standard. The ground is the Oracle spatial geodatabase where the geometry, semantics and topology of individual network elements are stored. LIDS allows for very effective graphic and non-graphic data management and their storing in a unique central database in compliance with standards of the OpenGIS consortium (OPEN GEOSPATIAL CONSORTIUM, 2013). The graphic interface is formulated by means of the CAD MicroStation software (or the PowerMap Bentley Systems); the core of the system is the well scalable LIDS data server which supports long-term transactions and replications of the database data (BENTLEY SYSTEMS, 2013). The interconnection with users and third parties' products is provided by the intranet LIDS application server (LAS). The LIDS system is designed as a general system and its specific behaviour is controlled by the data model in the form of system configuration tables unique for the actual application (ASSECO BERIT, 2013). In this case, the application was based on the newly designed data model for KIDSOK public transport geodata management.

The disadvantage of the GIS solution of LIDS is the pre-defined configuration for a specific data model which is difficult to adapt or extend; the great advantage is the effective management of code lists (set of values may be acquired by the attributes) by means of so-called Codelist Manager (Fig. 3). It is a user-friendly interface where all code lists and their values, which occur in the data model, may be managed. Apart from erasing, other values may be added or the existing values may be modified. The LIDS Codelist Manager for the geodata management data model of the PT network comprises nearly 40 different code lists. Some were adopted from CIS JŘ, others were adopted and modified from the NeTEx standard and some were newly created for the needs of KIDSOK upon their new requirements or the existing sources. The names of the code lists have strict rules according to their content (C_name_of_codelist); most of them are of text data type with the scope up to 50 characters. The key codelist is the list of complete names of all stops which comprises three codelists from the CIS JŘ – the codelist of town names, the codelist of town parts and local designations and the codelist with stop IDs according to CIS JŘ which makes it easy to keep the up-to-date combination of codelists to identify the stop.

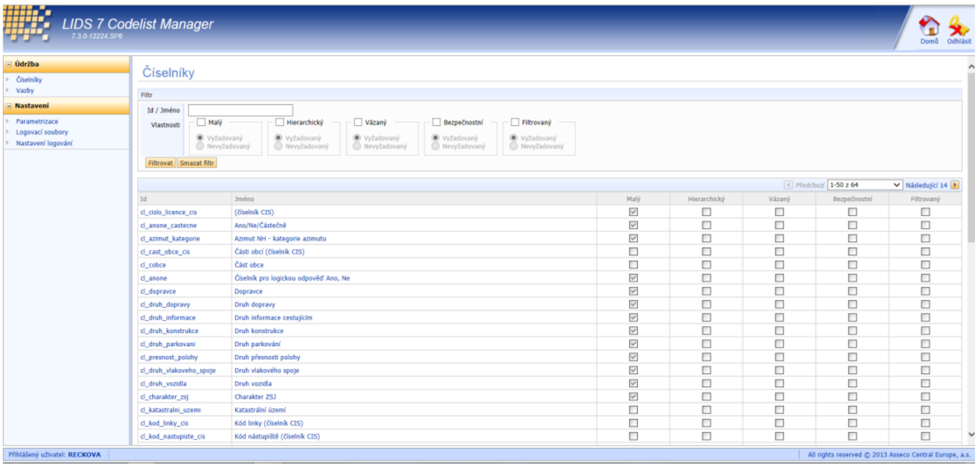

Fig. 3. The Codelist Manager for the management of data model codelists (Source: internal)

THE APPLICATION FOR DATA ACQUISITION

In the course of the data model development and the data warehouse implementation, an outline of suitable method for PT network data collection has been compiled. The network geoobjects are rather stable compared to the operation; the problem is, however, the number of registered entities and their attributes. That was the reason why for the static network elements (stops, boarding edges, IDs, stop lanes, equipment and spaces and objects related to the boarding edge or the stop) the most effective way of data collection was looked for. Upon the established data model, specification and KIDSOK requirements, Asseco developed the MAPDD mobile application for field public transport geodata documentation. There were many discussions and analyses whether it is more favourable to develop a field application for a Smartphone or for a tablet. In-depth analysis came to the conclusion that a tablet is definitely a better option for field surveys. The main reason was the size of the screen on which it is much faster, clearer and more comfortable to work with forms. The great advantage of the selected technology is the in-built GPS and navigation, the possibility to connect to WiFi or to take pictures. All these options are used when collecting the data in the PT network. Another plus is the long-life of the battery which can moreover be recharged in the car when moving to the next position.

The MAPDD application is designed for a mobile device (tablet) which meets at least the following technical parameters:

- LCD touch-screen with the resolution min. 1024x600,
- 1 GHz processor, 8 GB internal memory,
- camera, WiFi and GPS,
- OS Google Android 4.0 and above.

The MAPDD application communicates with the data warehouse on the basis of the server part of the application in LIDS which contains – apart from the environment required for the LIDS data warehouse management – also the application servers. The basic application environment for the mobile application comprises three server services which must be running when the mobile device communicates with the data warehouse. The first service – OracleServiceXE and OracleXETNSListener – is the operation of the Oracle XE database and allows for connection to the involved database. LIDS, which is a visualization of the Oracle database, is the data source for the mobile application and concurrently the recipient of the change data from MAPDD. The second service represents the application server for LIDS, so-called JBoss Application Server 4.2.3, which provides authentication of the service for the client mobile application. The third service is the domain1 GlassFish Server which represents the application server for the client mobile operation and provides various application services.

The MAPDD application mostly operates off-line; on-line operation is only used when transferring all the data between the data warehouse and the mobile device during the synchronization at the beginning of the field survey. At the end of the survey, only the data exchange from the mobile device is transferred by synchronization to central data warehouse. The process is provided by a module which recognizes and makes versions of the newly edited data. The name of the field worker passing the batch, date and time of the data change is stored together with each downloaded batch (for process scheme of the application operation see Fig. 4). The user account with the assigned role and therefore also the offered functionality is defined for each MAPDD user. All changes made in the data warehouse within the GIS interface of the data warehouse are stored in the data modification history.

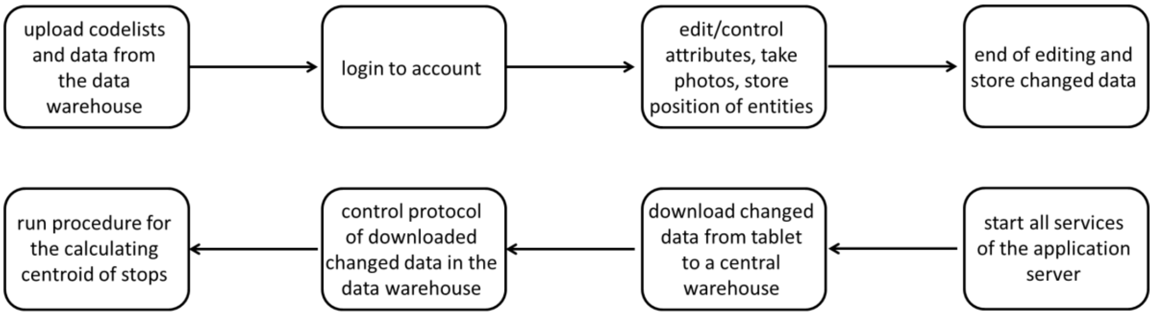

Fig. 4. The process scheme of the application operation (Source: internal)

When the field survey is carried out in the Olomouc Region, MAPDD enables to choose from the list of all existing stops in alphabetic order the actual stop for data acquiring or updating manually from the list, through filters by using a partial string from the name or by means of automatic search for stops within the radius of two kilometres from current position of the tablet. When the particular stop is selected, a tree of object types which belong to the actual stop opens in the left-hand part of the basic form. Basic data on the edited stop are shown on the top of the form; the right-hand side includes the attributes of the edited entity (Fig. 5). Each type of entity (boarding edge, ID, stop lane, access space or equipment) can be edited, erased or created as a new one.

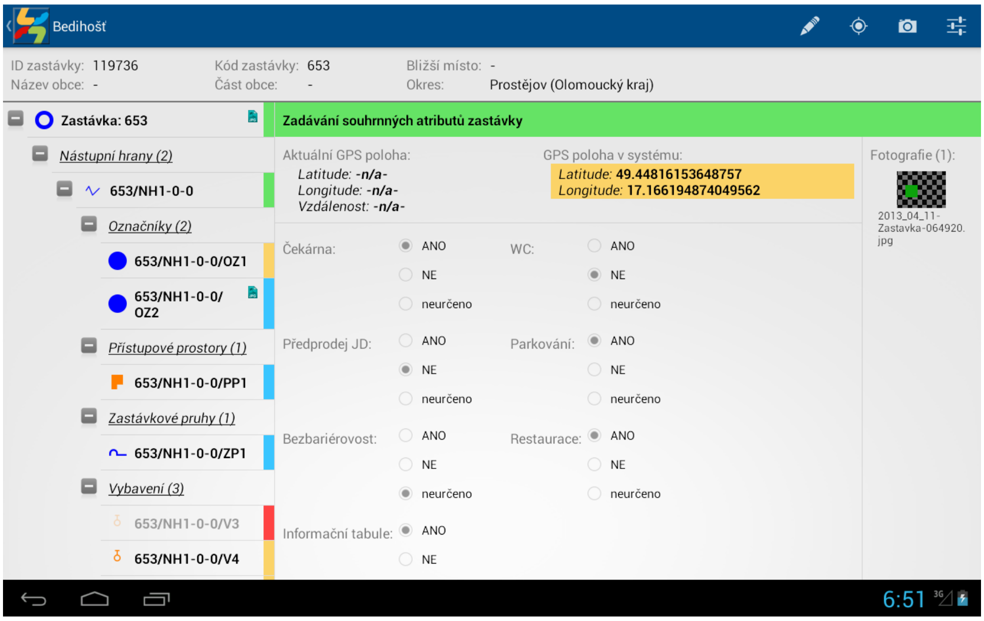

Fig. 5. Example of the MAPDD application form for editing basic data on the Bedihošť stop

MAPDD have the functions to register the position of spatial elements (the position of the beginning and end of the boarding edge, stop lanes, ID position and the position of spaces and objects related to the boarding edges or to the stop in general), take pictures and enter describing attributes to parts of the stop or to the stop in general. Entering of most of the attributes is dealt with by means of codelists in the form of check boxes where only one option can usually be selected. In the field data collecting application, attribute values may be selected from more than 25 codelists; several other codelists work only to allow moves within the entity tree. Entering from the keyboard is minimised to decrease the risk of non-uniform or incorrect entries in the sense of entering various strings for the same aspect. The key codelist is the stop ID codelist from CIS JŘ and derived complete names of all stops comprising codelists for the names of towns, parts of the towns and specific locations which the field worker cannot modify (in other codelists it is possible to add extension of the codelist in comments or select the “other” option). Thanks to this measure, editing of a non-existing stop is avoided. Besides the elimination of logic mistakes, entering of incorrect values resulting from misunderstanding or poor knowledge is dealt with. Help is attached to some attributes or codelists in case the field worker is not sure of the variant of the codelist value or they do not understand the sense of the entered attribute. Apart from texts, the help often contains pictures for clearness.

USER ASPECTS OF MAPDD

The data will prevailingly be used by the officials who manage the PT network, potentially change the routes or names of the stops, create synoptic maps and provide information reports and notices for general public. So far the KIDSOK staff only worked with a point layer where each record represented a stop with all its compounds in one point. Upon these data, they elaborated synoptic maps for passengers and source materials for internal purposes (negotiations on integrating routes or lines etc.). As a part of the requirement to increase the attraction and competitiveness of public transport, the demand was expressed to form an interactive transport plan and analyse transport services in the area, define attraction districts of stops and many others. That was the moment when KIDSOK staff realized that the existing data are insufficient and the effort and financial means will have to be invested to acquire correct, up-to-date and first of all complete data. The data will be used by the organiser's officials, later also by passengers and carriers' controller systems because the data model contains attributes important for PT organizers (focused on passenger needs and public transport management and control) but also attributes related to operation. While the passenger is interested in information about wheelchair access, vehicles moving within the network, delays, possible connections between two points and the appearance of the boarding edge (Fig. 6), the carrier needs to know the data concerning the height of the boarding edge, the length and depth of the bay, lighting and other technical parameters of the stop or its parts.

Fig. 6. Picture of the boarding edge in-situ (Source: internal)

The benefit of the newly established data warehouse of the expertly created data model is the evidence of complete data on PT including all existing attributes, the existence of the expertly elaborated codelists preventing unprofessional interpretation, great positional and topological accuracy of objects, ensuring complete data collection by means of modern technologies (GPS, tablet, GIS) and easy updating of the data.

SUSTAINABILITY OF DATA CURRENCY

Within the draft of the data model, elaboration of MAPDD and pilot filling of the data warehouse, the concept of PT network data updating was also formulated as it is complicated due to the large number of registered entities and their attributes. Within the KIDSOK operated area, there are more than 2500 unique stop names and more than 6000 signposts; similar numbers were reached also with boarding edges and other stop elements which need to be registered. Moreover, many attributes and aspects need to be monitored with respect to providing information for the handicapped. This is the most time consuming part of the work on the data warehouse. The newly created stops, parts of stops and any spatial and organization changes will further be monitored with the participation of several subjects (Fig. 7). At the moment, 20 % of the whole network in the Olomouc Region have undergone MAPDD pilot testing and debugging. The time frame necessary to document one latest, fully equipped stop (including all its parts) is 20 minutes on average. The rest of the network will be documented in spring 2014; the data warehouse shall be complete by autumn 2014.

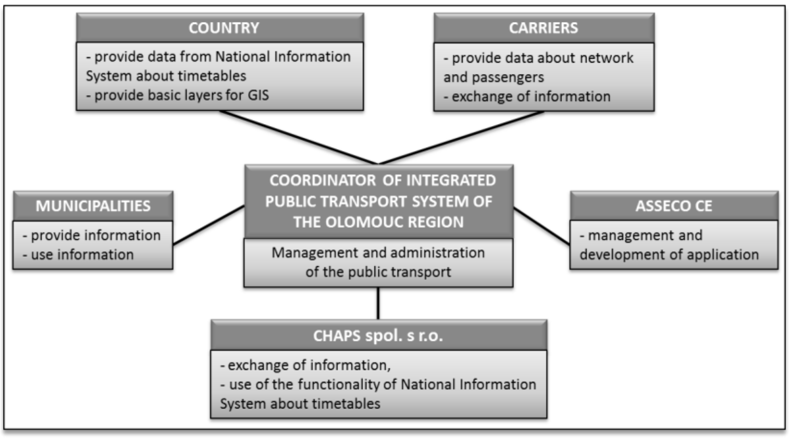

Fig. 7. The organization of the public transport network management

Updating of the data from CIS JŘ (codelists of stop names, stop IDs etc.) are contracted with CHAPS Ltd. with the period of one month, other up-to-date information is acquired from the Transport Authority of the Olomouc Region (ensured by the Department of Transport and Road Management of the Olomouc Region) or is consulted directly with carriers. The last source of information about transport network is represented by the towns which are usually owners or administrators of the stops and which often provide primary information. MAPDD elaboration, maintenance and technical support is provided by Asseco.

CONCLUSION

Within the draft of the data model, elaboration of MAPDD and pilot filling of the data warehouse, the concept of PT network data updating was also formulated as it is complicated due to the large number of registered entities and their attributes. Regarding time perspective, filling of the data warehouse is the most demanding work. The newly created stops, parts of stops and any spatial and organization changes will further be monitored with the participation of several subjects. In future, the same problem as KIDSOK will be dealt with by most of other PT organizers who can draw inspiration from this solution. Asseco is considering the possibility to offer MAPDD and LIDS as a commercial product for PT network geodata documentation and management in the Czech and Slovak Republics.

At the moment, 20 % of the whole network in the Olomouc Region have undergone MAPDD pilot testing and debugging. The data warehouse shall be complete by autumn 2014. This phase will be followed by elaboration of an interactive transport plan based on the collected data, the possibility to find a connection with a visualization in the map and with the option to provide information concerning on characteristics of the stops (especially the wheelchair access for general public). Concurrently, the managed area will undergo many spatial analyses based on these data (transport services, attraction districts of stops, number of connections and many others). These data will also comprise a base for a controller system.

ACKNOWLEDGMENT

The author gratefully acknowledges the support by the Operational Program Education for Competitiveness - European Social Fund (project CZ.1.07/2.3.00/20.0170 of the Ministry of Education, Youth and Sports of the Czech Republic). The author also acknowledges the support of the Internal Grant Agency of Palacky University in Olomouc (Project No. PrF 2013 024).

REFERENCES

ASSECO BERIT (2012). Solutions for Demanding Business. *ASSECO Berit* [online]. 2012 [cit. 2013-11-04]. Available from: http://www.asseco-berit.de/asseco-berit/ueber-asseco-berit/

ASSECO BERIT (2013) Solutions for Demanding Business: LIDS 7 Geografisches Informationssystem (GIS) und Betriebsmittelverwaltung (BIS) aus einem Guss [online]. [cit. 2013-10-07]. Available from: http://www.asseco-berit.de/produkte/lids/.

BENTLEY SYSTEMS (2013) Bentley PowerMap [online]. [cit. 2013-10-07]. Available from: http://www.bentley.com/cs-CZ/Products/Bentley+Map/PowerMap-Product.htm.

Centrum dopravního výzkumu, v. v. i. (2012) Jednotný systém dat ve veřejné dopravě [online]. [cit. 2013-03-04]. Available from: http://jsdv.cdvinfo.cz/.

CEN/TS 15531:2011 (2011) Service Interface for Real Time Information. Brusel: European Committee for Standar-dization. Available from: http://www.kizoom.com/standards/siri.

DRDLA, P. (2008) IDS in The Czech Republic – comparisons and strangeness (IDS v České republice - srovnání a zvláštnosti). Perner´s Contacts. vol. 3, num. 5. pp. 69-74 Available from: http://pernerscontacts.upce.cz/12_2008/drdla1.pdf.

EN 12896:2006 (2006) Road transport and traffic telematics - Public transport - Reference data model. Brusel: European Committee for Standardization. Available from: http://www.transmodel.org/en/cadre1.html.

EN 28701:2012 (2012) Intelligent transport systems - Public transport - Identification of Fixed Objects in Public Transport (IFOPT). Brusel: European Committee for Standardization. Available from: http://www.dft.gov.uk/naptan/ifopt/ifoptV1.0-36/CENTC278WG3SG6_IFOPT_20081110_36.pdf.

CHAPS spol. s r. o. (2013) Celostátní informační systém o jízdních řádech: Popis systému. CHAPS: Celostátní informační systém o jízdních řádech [online]. [cit. 2013-03-16]. Available from: http://www.chaps.cz/cs/products/CIS.

CHAPS spol. s r. o. (2005) Popis formátu a struktury dat pro elektronické zpracování drážních jízdních řádů (verze 1.1) CHAPS: Celostátní informační systém o jízdních řádech [online]. [cit. 2013-03-18]. Available from: http://www.chaps.cz/files/cis/jdf-1.1-v.pdf.

OPEN GEOSPATIAL CONSORTIUM (2013) OGC Standards and Supporting Documents [online]. [cit. 2013-10-07]. Available from: http://www.opengeospatial.org/standards.

PRE/CEN TC 278 WG9 (2012) Network Exchange. Brusel: European Committee for Standardization. Available from: http://www.kizoom.com/standards/netex/schema/index.htm.

Zákon č. 111/1994 Sb., o silniční dopravě, v platném znění (1994).