

5-2012

Application of the Temporal Database Technology in the Development of Latvian Railway Information Systems

Eugene Kopytov

Transport and Telecommunication Institute, kopitov@tsi.lv

Natalia Petukhova

Transport and Telecommunication Institute, natalia@ldz.lv

Follow this and additional works at: <http://aisel.aisnet.org/confirm2012>

Recommended Citation

Kopytov, Eugene and Petukhova, Natalia, "Application of the Temporal Database Technology in the Development of Latvian Railway Information Systems" (2012). *CONF-IRM 2012 Proceedings*. 40.

<http://aisel.aisnet.org/confirm2012/40>

This material is brought to you by the International Conference on Information Resources Management (CONF-IRM) at AIS Electronic Library (AISEL). It has been accepted for inclusion in CONF-IRM 2012 Proceedings by an authorized administrator of AIS Electronic Library (AISEL). For more information, please contact elibrary@aisnet.org.

Application of the Temporal Database Technology in the Development of Latvian Railway Information Systems

Eugene Kopytov

Transport and Telecommunication Institute

kopitov@tsi.lv

Natalia Petukhova

State Joint-Stock Company "Latvian Railway"

Transport and Telecommunication Institute

natalia@ldz.lv

Abstract

The paper presents the research of temporal data usage in the information systems (IS) for railway transport. In the research the IS artifact building phase has been executed. The models and methods oriented at the operation with the temporal objects of the railway transport system have been designed. The work takes into consideration the problems of the temporal objects presentation in the relational databases, as well as the problems of provision of their integrity and the questions of interaction with them. The investigation results have been employed in the development of the interactive passenger train schedule IS for the Latvian Railway.

Keywords

Railway Transport, Information System, Artifact Building Phase, Temporal Database, Data Model

1. Introduction

The artifact building phase is considered to be a particular important phase of the design science in the information systems (IS) research (March & Smith 1995; Hevner & Chatterjee 2010). The artifact building phase produces "a viable artifact in the form of a construct, a model, a method, or an instantiation" (Hevner et al. 2004). Construction of the models of the stored data and development of the methods of their processing can be referred to the most complicated tasks of the IS artifact building phase. The data models must completely reflect the properties and the specifics of the IS objects and processes (Dey et al. 1999).

A specific feature of the data stored in the IS, which are used in transport, is their temporal character. The time is a fundamental index for the tasks of planning and forecasting of the transport company activity as well as for analytical and operative missions, demanding the usage of IS data measured in different moments of time. Partly, these data called the *temporal data* (Tansel et al. 1993; Date et al. 2002) are active only at the definite time span; they are connected with the definite dates. Accurate and proper organization of temporal data in databases and the access facilities to them determine the functioning efficiency of the transport IS. Each transportation mode has some additional specific peculiarities, which should be taken into consideration in the process of temporal database development. In particular, the railway transport is determined by the dynamic schedule, which depends on

many factors: transportation system physical limitation, interconnection between the schedules of passenger trains and freight trains, etc (Chiu et al. 2002; Kopytov et al. 2008). Certainly, the design of the databases, capable of characterizing the railway transport system condition at any moment of history, is exclusively a complicated task.

According to the authors' view, the situation is conditioned by the fact that the present-day IS continue using the classical relational databases technologies; and these technologies don't support the multi-versions of the storage objects. The undertaken study of the railway IS has shown that in many cases these systems use the temporal principles inefficiently. Thus, the majority of railway IS are characterized by the incompleteness of the stored temporal data, the non-trustworthiness of particular facts, etc. For provisional processing of temporal data, the users have to develop their own applications (Kopitovs et al. 2002).

For many years the main obstacle for the temporal databases technology development has been the requirement for huge resources of memory for multi-versions objects storing. There are other factors which restrict the temporal databases usage in such big enterprises as the railway; first of all: absence of the standards for the temporal data processing; lack of necessary skills for the operations with temporal data; lack of tools for supporting temporal data integrity, accessibility and privacy. The listed issues give evidence of current importance of the investigations directed at the efficiency enhancement of the temporal data usage in the railway IS.

Development and management of the IS temporal data in the railway transport is a complex task, the investigation of which is carried out in the following directions: IS development and functioning on the railway; temporal IS and databases designing; temporal data operative and analytical processing.

Investigations of the *IS development and functioning on the railway* demonstrate the complexity and peculiarities of the railway performance and propose some functioning models at the macro- and micro-levels (Ning & Brebbia 2010, Greiner & Volek 2010; Lecky et al. 2003). As a rule, researchers show the importance of time factor accounting but don't provide any peculiar solutions for the issues having in view mostly the classical approaches.

One of the most powerful investigations in the sphere of the *temporal databases* is the set of works by C. Jensen written from 1991 to 1999 in cooperation with different researchers. This collection of works was published in the form of thesis (Jensen 2000) and has more than 1300 pages. There are also fundamental works (Date et al. 2002) devoted to the *temporal technologies*, which cover the problems of storage and interaction with the temporal data within the framework of the relation model. However, the authors mostly concentrate on the problems of temporal databases generating and the Tutorial D language usage but don't go into the problems on periodicity and special calendar occurring in transportation IS. The immense contribution in the temporal databases development of the TimeCenter is worth mentioning. The center's researchers (Terenziani 2003; Terenziani & Luca 2006) gave a special consideration to processing the temporal constraints for the events reoccurring with the course of time, the so called "periodicity".

Implementation of the *operative and analytical temporal data processing* is found in the research works in the sphere of business intelligence as well as in the databases warehousing theory. B.Inmon, R.Kimball, and M.Ross have made an immense contribution to the formation of the fundamentals of the multidimensional data modeling and the classical approaches to the tracing of any changes in dimensions (Inmon 2002; Kimball & Ross 2002). Many researchers have investigated the different optimization problems on the railway (Bussiecky et al. 1997; Chiu et al. 2002; Ning & Brebbia 2010). But there are no researches which investigate the problem of temporal data using in railway IS, taking into consideration the specificity of the transport processes.

From the above, we can conclude that research on the application of temporal technologies in building IS for railway transport is really vital. The IS of the State Joint-Stock Company “Latvian Railway” have been chosen as the object of applying the results of the research. .

2. Usage of temporal data in the railway information systems

The Latvian Railway employs various IS developed in different times on the basis of different technologies. The interconnection scheme of the Latvian Railway IS with the objects exhibiting temporal characteristics as well as their affiliation to one of the temporal data types are shown in Figure 1. The scheme contains 17 different IS (ALSN, APFIS, etc), which have been functionally divided into five groups: cargo transportation, passenger transportation, economics and finances, infrastructure and reference IS. The types of temporal data considered by the authors are presented below.

Multi-version data reflect different object states as a set of versions. In the course of time, many objects participating in the organization of the railway transport functioning are subjected to changes: varying train routes, non-permanent currency rates, fluctuating railway service rates, volatile energy supply prices, railway network division into lines and areas, etc. Such object versions may refer not only to the present or past but also to the future. To manipulate versions, which are valid in different periods of time, each object state is associated with an active time period.

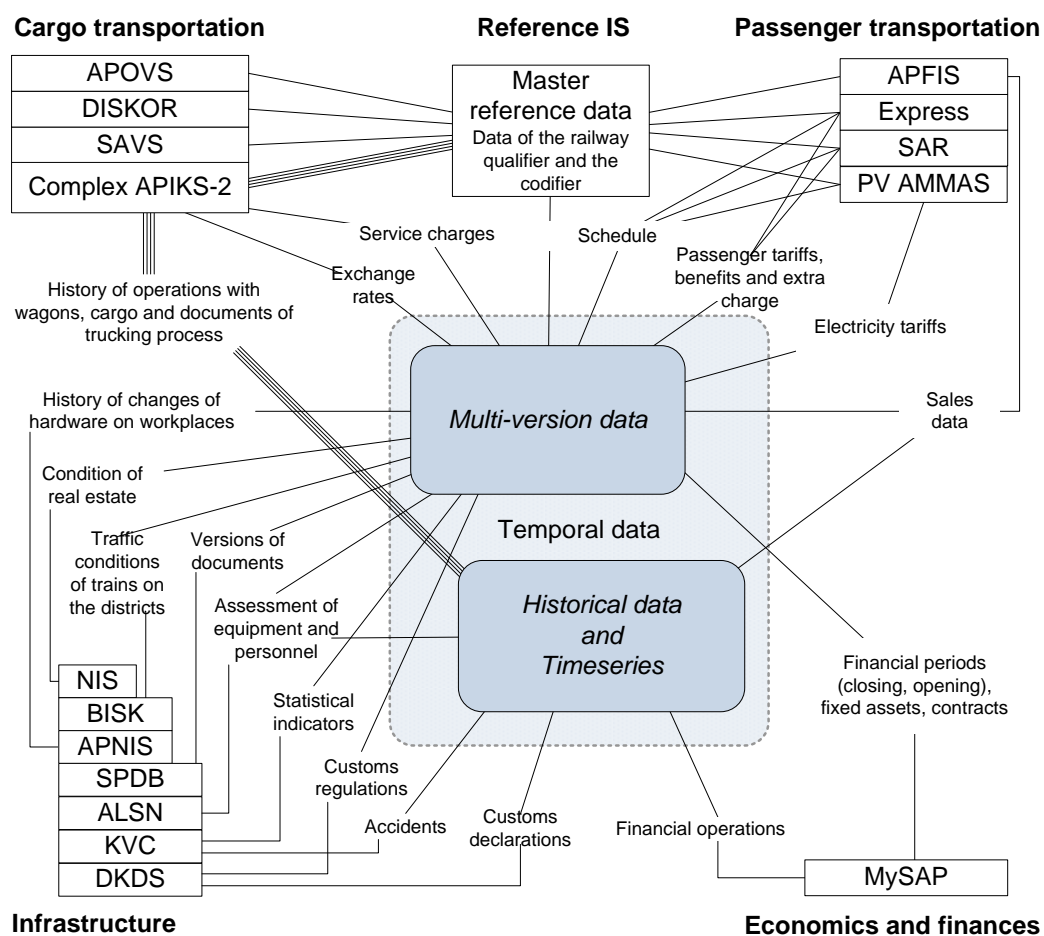


Figure 1: Temporal data in the Latvian Railway IS

The accumulated *historical data* are various operations, papers and other homogeneous data immanently occurring in the system and accumulated within it (ticket receipts, invoices, freight procedures, etc). The data of these types characterize functioning of the basic railway processes in time. Each issue of these data is inseparably associated with a particular date or time.

Time-series are the data collected in the process of the object observation. These observations are taken either in succession in time or with the uniform time intervals, or continuously, without intervals (for example, the carriage speed measurement results, their location coordinates). Observations of this type can be arranged chronologically.

Actually, in all systems under consideration there are used classical relation models which complicate dealing with temporal data. First of all, it concerns processing multi-version data. Transition to temporal data models both in newly built and in current railway IS will allow substantial simplification of the temporal data application. It is realization of this transition which is the main task of the given research. This task is being solved by the authors in regard to the artifact building phase, the suggested methods and models are presented in the next section.

3. Choice of models and methods for temporal data support

Let us consider the artifact building phase of the design of the interactive train schedule IS, which was created on Latvian Railway during the last years. Special models and methods were required for the temporal data support in the railway IS. Some of these methods and models were well-known and required the approbation and verification in the process of functioning, and several models and methods have been developed by the authors. The complex of these models and methods is presented in Figure 2. The double line in the scheme shows the models and methods groups under our investigation. The scheme demonstrates the details of the groups requiring development. A part of the methods (it is especially typical for the system level) are general methods, but some of them are the temporal data-oriented ones. The known models are indicated in grey colour and the models and methods created by the authors are shown in white. Development of particular models and methods was considered in the previous papers (Kopytov et al. 2004, 2010, 2011, etc).

The *aim of the given paper* is integration of the results of research in the sphere of temporal databases and presentation of the general concept of building models and methods in designing IS for railway.

All models and methods are classified according to the levels of the data processing. The system level includes databases management system facilities; the logical one comprises the components which allow describing the temporal model; and the application layer incorporates the models and time-dependent/oriented application systems methods, taking into account the data temporal characteristics. Main attention in the research is dedicated to the logical level, particularly, to the representation of the IS temporal objects in the database and the methods of the interaction with these temporal objects. Let us consider each level in more detail.

The system level has five groups of models and methods. The *data storing models*, the *methods of access to the physical storage objects*, and the *methods of increasing the data access speed* are connected with data allocation and organization within the files on disk systems, and application of various indexing methods. Similar tasks are solved by the *methods of storing and processing the bulk volume data*, but these methods are large-data-amount operation-oriented ones (Kopytov et al. 2001). The methods from the group *transaction mechanism* play an important role in the temporal model integrity implementation.

The logical level. The choice of the *temporal data model* is a key point at the level of logical design of the database. The temporal model is considered similarly to the relational one, it consists of structural, manipulation and integral parts. The differences between the temporal

models in the structural part are expressed in the set of temporal attributes; in the integral part in the temporal integrity limitations, and in the manipulation part in the algebra of queries to temporal data. Accordingly, the methods presented in the scheme have their own peculiar implementation for every data temporal model.

The structural part The *temporal data model with the object abstract identifier* serves as a universal model of the temporal data representation in the relational environment (Section 4.1). Relations, describing the temporal object characteristics, may have various *forms*: Rollback, Historical-Classic, Bitemporal, and Historical-Overlapped. The first three forms are widely known, and the Historical-Overlapped is developed by the authors (Section 4.2). The temporal object models, possessing the periodicity and special calendar, are suggested for data presentation in the railway schedule (Section 4.4).

The manipulation part comprises the methods, describing correct interaction with the temporal data as well as a temporal framework, providing this interaction. Whereas the specialized temporal database management systems have not been employed in the railway IS, the developers have to generate the logic of the temporal objects manipulation by their own efforts and resources. Employment of the *temporal query language* as well as the *methods of the temporal queries optimization* increases the efficiency of the queries execution and allows avoiding the ponderous standard query language. The solution of many problems mentioned above can be found in the *method of providing the user's temporal environment* (Petukhova 2007). This method allows setting the user at the definite point in time and executes the non-temporal queries to the temporal data (Section 4.3). Every temporal form requires a peculiar solution for the task of the determination of the temporal object version, active at this moment of time. The authors suggest two alternative methods of the object active version determination: *the logical rules method* and *the temporal elements method* (Section 4.4).

Employment of the *interval data transformation method* can solve the problem of “clumsy” data presentation for computation by transformation the interval fact into the time sequence (Section 4.5). The known *discretionary access control* and the *mandatory access control methods* don't cover all the requirements of the temporal data access strategy. The developed methods of the *temporal data confidentiality* protect the temporal data from an unapproved access to reading and changing procedures at the relation tuple level.

The integral part is represented by the *integrity providing methods* comprising the methods of supporting the temporal normal form, the temporal attributes' values integrity and other integrity constraints. The integrity part also incorporates support of the temporal relation functionality of different forms (Kopitovs et al. 2002).

The application level has five groups of methods. *The methods of the temporal data completeness and validity enhancement* are in demand in both the analytical and the transactional systems. The temporal data analysis by virtue of any *analysis methods* including the *statistical methods* and *data mining methods*, contrary to the trivial methods, is extended by the skill of “understanding” the data temporal structure. The systems of the data bulk volume processing, such as the data warehouses, are inevitably connected with the temporal data (Section 5.1). *The processes of the data operating in the data warehouses* are connected with the temporal aggregates computation and with the time sequences activities.

The basic methods and models developed in the considered research are reviewed in detail in the next section.

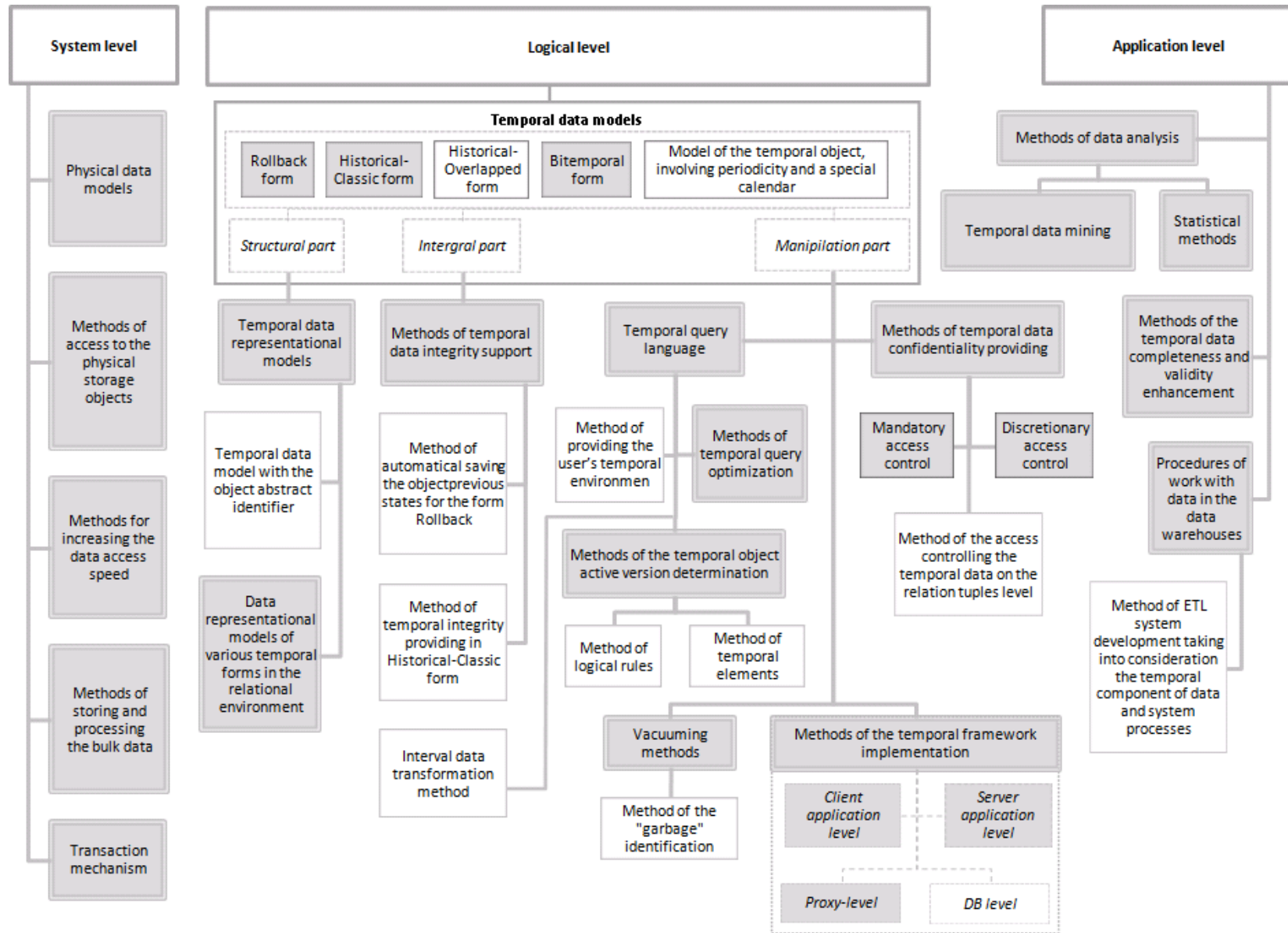


Figure 2: Models and methods used in temporal IS

4. Temporal models and methods of development and implementation

4.1. Temporal data model with the Object Abstract Identifier

Classical approaches to the database design are not enough for the representation of temporal data in the relation environment which is largely used in the railway IS. The multi-version nature of temporal objects, maintenance of their integrity at the structural level, consideration of their variety (periodicity, versions' overlap) require special approach to the database design. The authors suggest the following pattern: *the temporal data model with the object abstract identifier* (AOID) serves as a universal model of the temporal data representation in the relational environment (Kopitovs et al. 2002). It allows presenting the temporal object in the form of a relation set. The object life section in this model is described with the lifespans of all its characteristics determined in different relations and having temporal attributes: the time t_s of lifespan beginning and the time t_e of lifespan ending. The mathematical description of the model AOID is presented below.

It is assumed that the object O of the temporal system is characterized by the unique identifier OID and the set of properties A and can be represented in the terms of:

$$O = (OID, A).$$

The set of the object properties $A = \{a_1, a_2, \dots, a_m\}$ is divided into two subsets: the set of the static properties $A^s = \{a_1^s, a_2^s, \dots, a_k^s\}$, permanent in time, and the set of the dynamic properties $A^d = \{a_1^d, a_2^d, \dots, a_n^d\}$, changing in time, moreover:

$$A = A^s \cup A^d; A^s \cap A^d = \emptyset, \quad (1)$$

where k and n are the amount of the static and dynamic properties respectively; $m = n + k$ is the general amount of the object O properties.

Since the attribute values in the relation databases have to be atomic, it is not enough to introduce only one relation for presenting the numerous static and dynamic properties. The object O is presented as the interconnected relation scheme:

$$O = (R^{OID}, R_1^s, R_2^s, \dots, R_l^s, R_1^d, R_2^d, \dots, R_q^d),$$

where $R^{OID} = R^{OID}(OID, A_0^s)$ is the parent relation, describing the AOID and comprising the aggregate of the object static atomic attributes A_0^s , $A_0^s \subseteq A^s$; OID is the primary key of the relation R^{OID} ;

$R_1^s, R_2^s, \dots, R_l^s$ are the child relations, describing the object A^s static properties ($l \leq k$), and presented in terms: $R_i^s = R_i^s(OID, A_i^s)$, $i = 1, 2, \dots, l$, where A_i^s is the subset of the object static properties, logically aggregated in one relation, $A_i^s \subseteq A^s$. If the relation R_i^s describes the object atomic property, OID is the key, and if there are multi-valued properties, the key comprises the additional attributes;

$R_1^d, R_2^d, \dots, R_q^d$ are the child relations, describing the discrete ranging in time dynamic attributes

A^d , and presented in terms: $R_i^d = R_i^d(OID, A_i^d, T_i)$, $i = 1, 2, \dots, q$; $q \leq n$, where A_i^d is the subset of the discrete dynamic attributes, logically and on the basis of the simultaneous values changing aggregated in one relation, $A_i^d \subseteq A^d$. By employing such attributes separating for q relations the temporal normal form is achieved; T_i is the time attributes vector, which can present valid time (VT) or transaction time (TT), or both of them simultaneously, depending on the relation temporal form: $T_i \subseteq T$, where $T = \{t_s^{VT}, t_e^{VT}, t_s^{TT}, t_e^{TT}\}$ is the set of time dimension attributes. In case the relation R_i^d describes the object atomic properties, the relation key can be in the form of (OID, t_s^{TT}) or (OID, t_s^{VT}) , depending on the temporal form.

4.2. Temporal relation model with the overlapping life spans

One of the vital problems occurring in the railway ISs is connected with the overlapping periods of the object life cycles. As opposed to the classical notion of the life cycle and the principle of the temporal objects versions change, the object version life cycle might be disaggregated into several lifespans, meaning the possibility for the object version to enter the “shadow zone” and to get out of it at several time moments. The system of train schedule can serve an example of the system with the listed peculiarities (Kopytov et al. 2008). Usage of the classical temporal relations forms Snapshot, Historical, Rollback and Bitemporal is not efficient for the description of the temporal objects with the mentioned peculiarities; it is seen not only through the wasteful disk space employment but also through the excessive risk of violating the integrity of the temporal data in case of their changing. Under the condition of the steady continuous increase of the data volumes and the advanced requirements for the data integrity and availability in the transport systems the listed shortcomings become crucial (Petukhova 2007).

The temporal relation scheme Historical-Overlapped is performed in the form:

$$R(OID, A, VT, t_s^{TT}),$$

where t_s^{TT} is the time of the transaction fixation. The active version of the object OID is computed on the basis of this attribute values.

The existence of two or more versions with the overlapping time periods doesn't contradict the fact of the relation Historical-Overlapped integrity. It means that the relation R can comprise two tuples r and p with the following condition:

$$(r.OID = p.OID) \wedge (r.VT \cap p.VT \neq \emptyset). \quad (2)$$

The Historical-Overlapped model stipulates the operations over the objects, which period of life is discovered not only in the present or in the past, but also in the future.

4.3. Temporal logic implementation

For the temporal logic development the authors have suggested the enhancement of the manipulative and integral parts of the relation model (Kopitovs et al. 2002). The principal tasks of this enhancement are the following:

- support of the time-based multi-version and temporal integrity provision;
- executing such commands as DELETE, UPDATE, INSERT and SELECT with the account of the temporal properties of the object;

- support of the temporal upward compatibility with the existing non-temporal applications;
- providing access to all the previous and all the following object versions;
- ensuring of the existence of the single active version of the object at any moment of time.

As an example of the suggested approach, let us consider the SELECT-command execution in non-temporal and temporal databases shown in Figure 3. The scheme shows that the application generates the same non-temporal queries; however, besides the non-temporal interface holding, the new temporal database is capable of processing the queries in different time slices. The example considers processing the queries of two independent users (User1 and User2) with different determined environment variables (time slices at 11.05.2008 and 15.06.2008 respectively). It is obvious that the users receive different values since the responses to the same query and these values are active at the moment of time, pointedly specified in the user's environment variable.

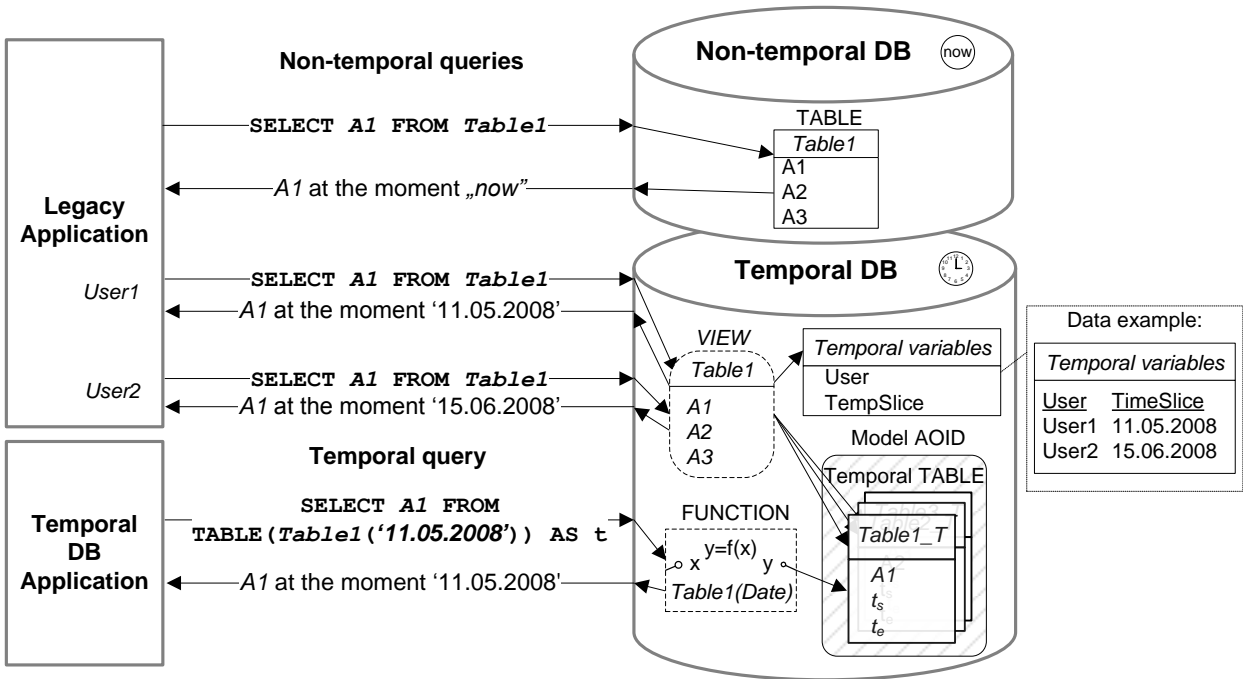


Figure 3: The temporal data reading

With the help of the considered method it is possible to manipulate efficiently the time slice where the user has to work without changing the application code and without employing any additional conditions in the queries.

4.4. Periodicity models in the railway scheduling tasks

The principal issues of the train schedule storage in the databases are connected with the existence of the multitude of its versions conditioned by the seasonal cycles of schedule changes, the days of week, the systematic and unplanned repair operations, the transfers of the working and festive days, etc. The authors have investigated three possible models of presenting temporal data of the train schedule: the duplicating model, the model based on logical rules, and the model

based on temporal elements. These differ in the building principles, the capabilities, and the resources requirements (Kopytov et al. 2011). Let us consider these models in detail.

In the suggested models, two basic sets are introduced for the train schedule description: the set of all stations of the railways $S = \{s_1, s_2, \dots, s_l\}$, and the set of all trains $N = \{n_1, n_2, \dots, n_m\}$, where $l = |S|$ and $m = |N|$.

Duplicating model (point form). The given model is based on making a calendar of each train traffic for every day t_i . Then, the i -th schedule version $v_i^{(n)}$ for the train with number $n \in N$ will be determined by the tuple:

$$v_i^{(n)} = \{n, t_i, \langle s_1^{(n)}, T_1 \rangle, \langle s_2^{(n)}, T_2 \rangle, \dots, \langle s_h^{(n)}, T_h \rangle\}, \quad i = 1, 2, \dots, k \quad (3)$$

where the pair $\langle s_j^{(n)}, T_j \rangle$ determines the j -th stop of n -th train, and $s_j^{(n)} \in S$ is the station, T_j is the departure time; h is the number of the train stops in the timetable $v_i^{(n)}$; k is the number of days on which the n -th train schedule is stored in database.

It should be noted that the number of the stored records stating on what day, at what time, at what station every train stops will be equal to the product of multiplication of the number of trains by the average number of stops and by the average number of days for which the schedule is stored. Though, the main problem of the above approach lies not in the volume of the stored data but in the complexity of providing the reliability of data and the immediacy of the schedule changes.

Model LR – the model based on logical rules (Interval form). The main point of the considered model is calculation of an active schedule version on the logical rules basis, which takes into account all the specificities of the schedule: multi-versioning caused by the operative changes, multi-variance in case of a periodical schedule, etc. For the specific schedule version identification of the train with number $n \in N$, the tuple $\langle n, C, t_f, t_s, t_e \rangle$ is employed. The periodicity property C presents a logical expression consisting of one or several elementary characteristics of the periodicity p_m connected by the logical operations \vee, \wedge and \neg (Kopytov et al. 2010). Then, the i -th schedule version $v_i^{(n)}$ for the train with number $n \in N$ will be

$$v_i^{(n)} = \{\langle n, C, t_f, t_s, t_e \rangle, \langle s_1^{(n)}, T_1 \rangle, \langle s_2^{(n)}, T_2 \rangle, \dots, \langle s_h^{(n)}, T_h \rangle\}, \quad i = 1, 2, \dots, q, \quad (4)$$

where q is the number of n -th train schedule versions, it is significant that $k \gg q$.

The model works with the schedule's temporal database and accounts all changes in the schedule as soon as they appear in the system. The central parameter of the access method is the day for which we need to find an active train schedule version.

Model TE – the model based on temporal elements (Interval form with reduction to point form) (Terenziani 2003). The peculiarity of the suggested model of determining the active version of the temporal object lies in the preliminary reckoning of the dates of activity for every schedule version and saving them in the temporal elements for the posterior employment. The timetable version $v_i^{(n)}$ of the train with number $n \in N$ can be determined by the temporal element TE_i :

$$v_i^{(n)} = \{\langle n, TE_i \rangle \langle s_1^{(n)}, T_1 \rangle, \langle s_2^{(n)}, T_2 \rangle, \dots, \langle s_h^{(n)}, T_h \rangle\}, i = 1, 2, \dots, q. \quad (5)$$

The temporal element (TE) represents the calculable set of dates when the exact version of the train schedule becomes active. The TE comprises all information on the periodicity of the timetable with the account of all operative changes and special calendars (Kopytov et al. 2010). One of the problems of the use of TE is maintenance of the TE in the consistent condition because when the schedule changes, the versions temporal elements might become out-dated and need to be recalculated.

Presence of various models of the data presentation is caused by the existence of the set of different IS using the train schedule, which are characterized by their functionality, resource requirements and traditions of development. For evaluating the efficiency of suggested data presentation models the Analytic Hierarchy Process method has been applied (see Kopytov et al. 2011).

4.5. Interval data transformation method

The majority of the data analysis methods work with the point form of data representation where the fact is associated with the time moments. But usually the multi-version data are stored in IS databases in the interval form which is inconvenient for the analytical calculations. However, the mathematical apparatus of the relation databases management systems does not provide the ready-to-use facilities for the interval form transformation into the point form and a suitable method has been developed by the authors).

We consider two basic relations $R1$ and $R2$:

- $R1$ is the relation of the interval form describing the set of the certain object versions and having the scheme $R1(A, t_s, t_e, C)$, where $A = \{a_1, a_2 \dots a_k\}$ is the set of the attributes, describing the object, k is the number of attributes. Every version has a functioning period from the moment t_s to the moment t_e and possesses a feature of periodicity and a special calendar Z (for example, the transfer of the weekends schedule on Tuesday 01.05.2011);
- $R2$ is the point form temporal relation with the scheme $R2(A, d)$, where d is dates when fact A is active.

It is required to execute $R1(A, t_s, t_e, C)$ transformation into $R2(A, t_n)$ and the data set of $R2$ has to be restricted by the fixed time period $T^{(q)} = [t_s^{(q)}, t_e^{(q)}]$, where $\forall d \in T^{(q)}$. The transformation formula describing the transformation procedure from interval form to point form has the following view:|

$$R2 = \sigma_{d \in T^{(q)}} R1 \times \tau(t_s, t_e, C), \quad (6)$$

where $\tau(t_s, t_e, C)$ is the function distributing the range $[t_s, t_e]$ on the d dates set, responding the requirement of the specified periodicity C ;

$$\tau(t_s, t_e, C) = \{d \mid d \in [t_s, t_e], p(d, C) = true\}; \quad (7)$$

$p(d, C)$ is the periodicity predicate, which is true in the case if date d has the property C , it takes into consideration the special calendar Z ; the selection $\sigma_{d \in T^{(q)}}(\dots)$ restricts the result by the analysis area $T^{(q)}$. The capacity of the resulting set can reach $|RI| \cdot |\tau(T^{(q)})|$.

5. Research results implementation in the practical tasks solution

5.1. Temporal data employment in data warehouses systems

In the research there has been developed a conceptual approach of the corporate data storage and the analysis in the Latvian Railway IS. This approach is based on the data warehouses technology implemented in practice (Kopytov et al. 2004). The schema of the technological process of data preparation for the Latvian Railway information-analytical system of the passengers' transportation is shown in Figure 4. In the technological process of the system functioning, there have been exposed at least four critical procedures:

- partial data deletion in the data warehouse;
- data correction in the data warehouse;
- correction of the database of Reference IS;
- partial data deletion in the data marts.

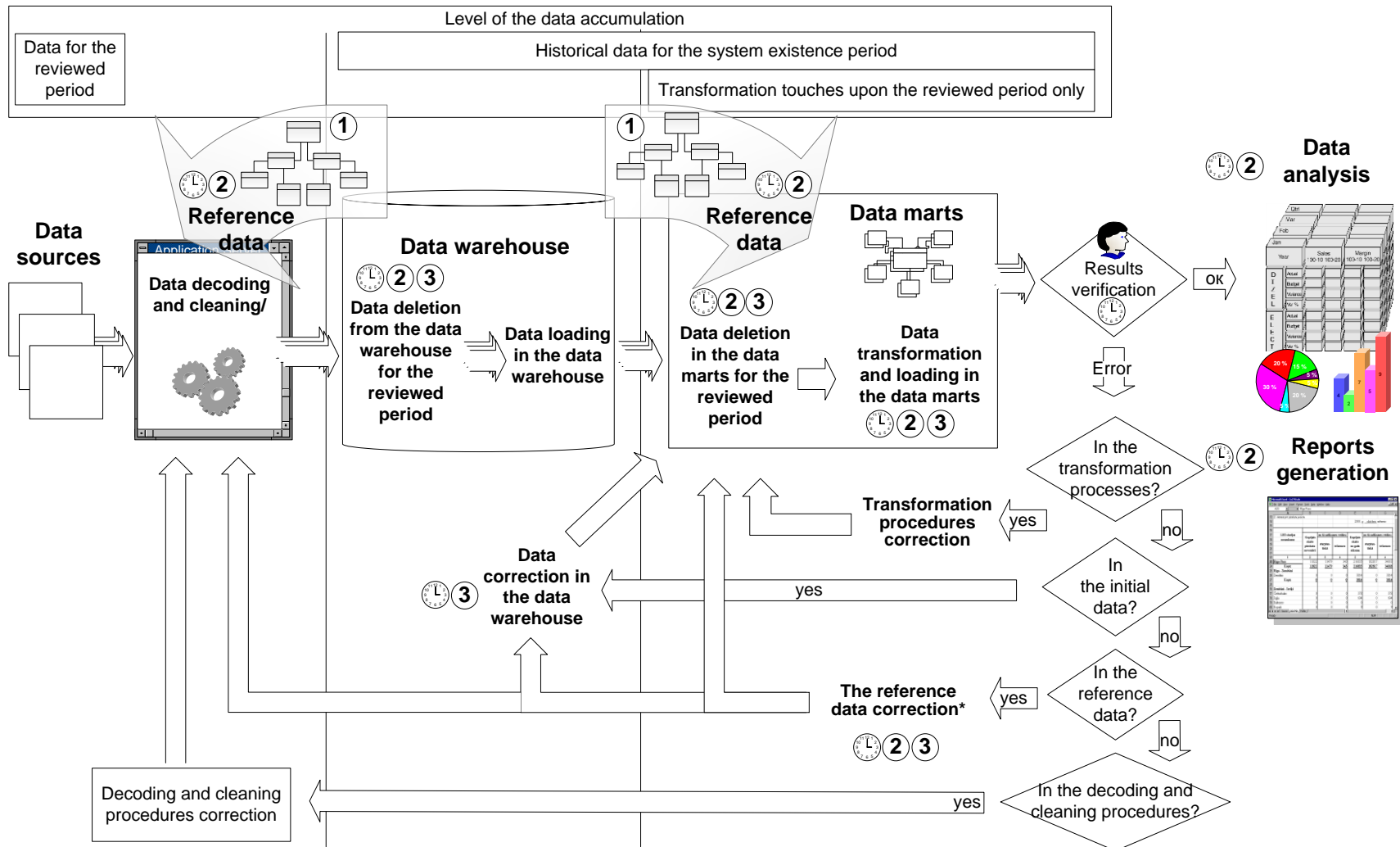
Implementation of the methods and models developed in the research permits to decrease the risk of the data warehouse integrity breaking, to simplify the data transformation and preparation for the analysis for any time period and to increase the immediacy and safety of these operations.

Marks ⌚ and ①, ②, ③ in Figure 4 perform the implementation areas of the developed models and methods in the process of the temporal data processing, and among them there are: the queries to the temporal reference data, the temporal sampling and aggregation, the data deletion of the specified time period, the access delimitation to the data referring to the unprocessed time range.

5.2. Temporal methods and models employment in the system of the Interactive Passenger Train Schedule Information System

The system of the interactive passenger trains schedule SAR is one of the principle objects of the research. It has the maximum amount of the temporal peculiarities (the multi-version, the periodicity, and the peculiar calendar, versions overlapping, etc) in comparison with the other Latvian Railway IS (Kopytov et al. 2011).

The system of the passenger trains schedule has the WEB interface. The system permits to take into account any prompt changes in the existing schedule and to project the timetable for the long run. The system manipulates with the data thematically, referential to three fields: the schedule, the fares, and the classifiers. The users interacting with this system data can be divided into three categories: the Internet users, the operators of the scheduling and fares systems introduction, and the operators of some other systems, including the inherited systems not capable of operating with the temporal data. Each group of users is characterized by their own subset of queries to the system.



The reference data correction* – depending on the fact what data have been corrected or replenished, the technological process recommences from the different stages. It is sometimes convenient to correct the data in the warehouse, but not to decode and to reset them again

- 🕒 - Time dimension support in the queries to the data
- ① - Model AOID
- ② - Method of providing the user's temporal environment
- ③ - Method of the access control to the temporal data on the relation tuples level

Figure 4: Technological processes in the data warehouse

The diagram of the developed methods and models employment in SAR system is performed in Figure 5. The known models and methods are indicated in white colour, and the models and methods generated by the authors are shown in grey. The methods and the models, in accordance with the presented functions, are divided into four groups: the temporal data representational model, the temporal data integrity, the access to the temporal data and the data physical arrangement.

The employment of the developed models and methods in the SAR system has permitted:

- to simplify significantly the IS development;
- to organize the data of the schedule in the optimal way taking into consideration all the revealed temporal peculiarities of the data (the model with *AOID*, *Historical-Overlapped form*, *Historical-Classic form*, *the Periodicity model* and *the Specific calendar*);
- to remove the limitations on the storage time of the schedule (for example, the Express tickets sale system limits storage of the schedule by the frame of 45 days);
- to transform the segments of the centralized database of the classifiers into the temporal ones, providing at the same time the temporal up-ward compatibility with other applications developed before (*Rollback form*, *Historical-Classic form*, *Method of the automatic saving the objects previous states for the form Rollback*, *User's temporal environment providing method*);
- to provide correct access to the active data versions (*Logical rules method*);
- to provide the data temporal integrity (*Method of the temporal data integrity providing for the form Historical-Classic*);
- to provide the interference immunity for the history of the schedule data, tariffs and classifiers (*Method of the access control to the temporal data on the relation tuples level*);
- to achieve the demanded speed of performance by employing *different indexation methods* including the *clusterization procedure*.

6. Conclusions

In the presented research there have been defined the problems of temporal data employment in the Latvian Railway IS. The specificity of the temporal IS on the railway transport has been investigated; the main tasks of the IS connected with the temporal data accounting and processing have been formulated; the problems originated in the development and implementation of these systems have been determined, artifact building phase has been executed. The fundamentals of the temporal databases designing in the relation environment through the employment of the temporal data model with the object abstract identifier and the temporal logic have been suggested. The unique models and methods, oriented at the operation with the temporal objects with the account of the railway transport tasks peculiarities, have been designed.

The proposed models and methods have been applied in the train schedule interactive IS in the Latvian Railway. The results obtained confirm the effectiveness of the models and the applicability of the methods. Application of created models and methods has removed the limitations for the period of data storing in the system, has simplified dealing with multi-version objects, has simplified queries to temporal data, etc.

The research results have been approved in the Latvian Railway IS. However, they can be implemented in the IS of other railway companies. Moreover, in many cases the obtained results have a universal nature and can be applied in different areas.

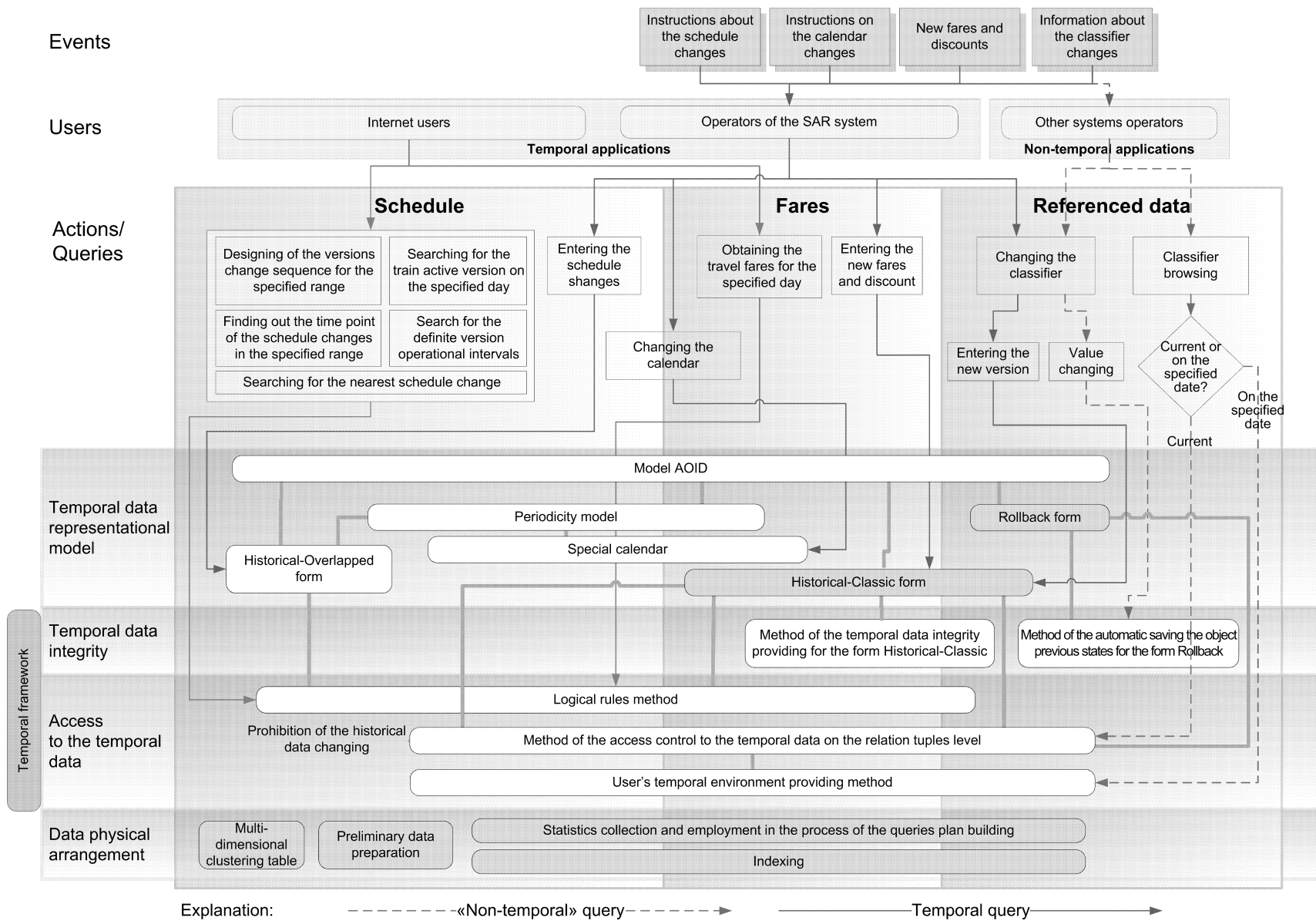


Figure 5: Temporal models and methods employment in SAR system

References

- Bussiecky, M.R., T. Winter, U.T. Zimmermann (1997) "Discrete optimization in public rail transport", *Mathematical Programming*, 79(3), pp. 415-444.
- Chiu C.K. et al. (2002) "A Constraint-Based Interactive Train Rescheduling Tool", *Constraints*, 7(2), pp. 167-198.
- Dey, D., V.C. Storey, T. M. Barron (1999) "Improving Database Design through the Analysis of Relationships", *ACM Transactions on Database Systems*, 24(4), pp. 453-486.
- Greiner, K., and J. Volek (2010) "Information System of Railway Undertakings' train Track Requirements". *Proceedings of the International Conference on Applied Computer Science*. Atény: WSEAS Press, pp. 254-258.
- Date, C.J., H. Darwen, N.A. Lorentzos (2002) *Temporal Data and the Relational Model*. Morgan Kaufmann.
- Hevner, A.R., S.T. March, J Park, and S. Ram (2004), "Design science in information systems research". *MIS Quarterly*, 28, 75-105.
- Hevner, A.R., and S. Chatterjee (2010) *Design Research in Information Systems*, Integrated Series in Information Systems, 22, Springer + Business Media, LCC.
- Inmon, W.H. (2002) *Building the Data Warehouse*, 3rd ed. NJ: Wiley.
- Jensen, C. (2000) *Temporal Database Management*. Dr. techn. thesis, Available at: <http://www.cs.auc.dk/~csj/Thesis/>
- Kimball, R., and M. Ross (2002) *The Data Warehouse Tool Kit*, 2nd ed. NJ: Wiley.
- Kopitovs, J., V. Demidovs, N. Petukhova (2002) "Method of Temporal Databases design Using Relational Environment", *Scientific proceedings of Riga Technical University*, Series #5(13), Riga: RTU, pp. 236-246.
- Kopytov, E., V. Demidovs, N. Petukhova (2004) "Principles of Creating Data Warehouses in Decision Support Systems of Railway Transport", *AIP Conference Proceedings. The American Institute of Physics*, 718, Melville, NJ: pp. 497-507.
- Kopytov, E., V. Demidovs, N. Petukhova (2008) "Application of Temporal Elements in the Railway Schedule Systems", *Transport and Telecommunication*, 9(2), pp. 14-23.
- Kopytov, E., N. Petoukhova, V. Demidov (2001) "Methodology of Huge Data Volume Processing System Development for Analysis of Latvian Railway Passengers Transportation", *Proceedings of VI International Conference "TransBaltica 2001"*, Riga, pp. 201-208.
- Kopytov, E., N. Petukhova, V. Demidovs (2010) "Methods for Railway Schedule Periodicity Support in Temporal Databases", *Proceeding of 9th Joint Conference on Knowledge-based Software Engineering*, Kaunas, Tehnologija, pp. 178-191.
- Kopytov, E., V. Demidovs, N. Petukhova (2011) "Application of the Analytic Hierarchy Process in Development of Train Schedule Information Systems". *International Journal of the Analytic Hierarchy Process*, Vol. 3(2), 2011, pp.130-149.
- Lecky E.K. et al. (2003) *Designing information systems on railway transport*. Moscow: Route.
- March, S. and G. Smith (1995). "Design and Natural Science Research on Information Technology". *Decision Support Systems*, 15, pp. 251 - 266.
- Ning, B. and C.A. Brebbia (eds.) (2010) *Computers in Railways XII: Computer System Design and Operation in Railways and other Transit Systems*, WIT Press.
- Petukhova, N. (2007) "Investigation of Temporal Data Models in Information Systems on the Railway", *Proceedings of the 7th International Conference "Reliability and Statistics in Transportation and Communication"*, Riga: TTI, pp.168-177.
- Tansel A.U., J. Clifford, S.Gadia (1993) *Temporal Databases: Theory, Design, and Implementation*. Redwood City, CA: The Benjamin/Cummings Publishing Company.
- Terenziani, P. (2003) "Symbolic User-Defined Periodicity in Temporal Relational Databases", *IEEE Transactions on Knowledge and Data Engineering*, 15(2), pp. 489-509.
- Terenziani, P. and A. Luca (2006) "Temporal reasoning about composite and/or periodic events", *Journal of Experimental & Theoretical Artificial Intelligence (Taylor & Francis)*, 18(1), pp. 87-115.