COVER PAGE

Towards a Longitudinal Database Model for Improving Data Quality

Eladio Domínguez, Jorge Lloret, María A. Zapata

Departamento de Informática e Ingeniería de Sistemas

Universidad de Zaragoza

Edificio de Matemáticas. Pza. San Francisco, s/n

50009 Zaragoza – Spain

Ángel Luis Rubio

Departamento de Matemáticas y Computación

Universidad de La Rioja

Edificio Vives. C/ Luis de Ulloa s/n

26004 Logroño – La Rioja – Spain

Phone number: +34-941 299 449

Emails:

Eladio Domínguez: noesis@unizar.es

Jorge Lloret: jlloret@unizar.es

María A. Zapata: mazapata@unizar.es

Ángel Luis Rubio **(corresponding author)**: arubio@unirioja.es

Towards a Longitudinal Database Model for Improving Data Quality

Eladio Domínguez[a], **Jorge Lloret**[a], **Ángel Luis Rubio**[b], **María A. Zapata**[a]

(a) Departamento de Informática e Ingeniería de Sistemas. Universidad de Zaragoza
(b) Departamento de Matemáticas y Computación. Universidad de La Rioja

Abstract

Longitudinal studies have been used as a tool in applied research for decades. In turn, more recently the term 'longitudinal database' has appeared to denote the database that supports the management of data used in longitudinal studies. However, the majority of researchers are focused on the treatment of the data once obtained from the so-called longitudinal database and not on the design and management of the database itself. Taking into account that high quality of the managed data is demanded, we claim that the notion of longitudinal database must be investigated in itself, by means of the definition of structures and processes that promote the accuracy of the data involved in the longitudinal studies. In particular, in the present paper we briefly describe a new proposal of a longitudinal database model. Databases that are designed according to this model capture a level of knowledge that allows the user to obtain data with a higher degree of quality

Keywords

longitudinal database model, data quality assurance

1. Introduction

The term 'longitudinal' has been used in scientific research for decades. In the majority of cases the specific term used in the literature is 'longitudinal study' (Delgado and Llorca, 2004). A longitudinal study is a tool used primarily in applied research aimed at discovering trends or patterns in groups of individuals over time. Many scientific disciplines, such as medicine (Yeh et al., 2008), sociology (Orth et al., 2009), technology (Luse et al., 2013) and education (Crane, 1996) have used longitudinal studies.

It is much more difficult to find literature in which the term 'longitudinal' appears as the subject of basic research. A longitudinal study is a statistical study, and therefore in this context we can find some references of this other kind (Cepeda and Núñez, 2007). Moreover, a basic feature of longitudinal studies is that they manipulate a significantly large volume of data. Since the persistent storage (over time) of these data is critical for the development of longitudinal studies, the term 'longitudinal database' has received more and more attention (Clark, 2006; Ander-Peciva, 2005). However, based on its current use, a longitudinal database apparently is nothing more than a database that stores the data that is used in longitudinal studies.

The thesis of the present paper is that, taking into account that high quality of the manipulated data is demanded, the notion of longitudinal database must be investigated in itself. In this paper we defend this thesis making a review of the use of the terms longitudinal, data base and temporal database in the literature. The main conclusion of this review is that, in the existing approaches (Ferguson and Kreiter, 2004), the longitudinal aspects are focused on the treatment of the data once obtained from the database and not on the design and management of it. As a consequence of this, an accurate definition of longitudinal database is lacking. Furthermore, with the goal of obtaining a more enhanced approach, we also advocate the definition of a longitudinal database model that allows us to obtain information of higher quality from stored data (Data Quality Assurance) which can be used, for example, as evidence for auditing processes required in laws, norms, standards and best practice guides for information technology (Looso and Goeken, 2010).

As a step towards this goal, in this article we present a proposal of a longitudinal database model, whose main characteristic is that it allows the user to extract evidences from the database, in such a way that (s)he can be aware of the protocols that have been applied to generate such evidences, whether these protocols have changed over time and to know the states of objects from which those evidences were obtained. Databases that are designed according to this model capture a level of knowledge that allows the user to obtain data with a higher degree of quality.

The rest of the paper is structured as follows. In the following Section we briefly review the use of the term ‘longitudinal’ in applied research, and in Section 3 we study the present use of the concept ‘longitudinal database’. Section 4 includes a summary of basic ideas about temporal databases and in Section 5 longitudinal and temporal databases are compared. Section 6 is devoted to presenting our proposal of a longitudinal database model. We finish the paper with some concluding remarks.

2. Longitudinal in the sciences

The term ‘longitudinal’ is almost always linked to ‘longitudinal studies’. It shall here be demonstrated that the concept appears in relation to various scientific and humanistic disciplines, forming a ‘corpus’ from which, despite various nuances, an approximate definition of ‘longitudinal study’ can be extracted.

A longitudinal study is a research tool that implies the use of observations or measurements taken from the same group of individuals, normally over an extended period of time (Delgado and Llorca, 2004). Other related terms are tracking study (sometimes used synonymously) and cohort study (a longitudinal study is considered to be a particular type of cohort study[1]). Normally, a

[1] In a cohort study it is most important to determine the cohort, that is, the set of individuals in a population that, sharing the same experience of

longitudinal study is regarded as the opposite of a cross-sectional study, which is taken at a single point in time (Vikström et al., 2002).

Uses of the concept of ‘longitudinal study’ can be determined by whether they refer to applied or basic research. In applied research, a ‘longitudinal study’ is a tool used by researchers to obtain certain conclusions about the object of the study. In general, there are two large areas of learning in which this type of study can easily be found: health science and social science. The following specific examples illustrate the variety of disciplines in which it can be applied. For example, in oncology, it has been used to investigate the clinical factors associated with fatigue over time in paediatric oncology patients (Yeh et al., 2008). In neurobiology, to study the relation between Parkinson’s disease and dementia (Dubbelinka et al., 2013). In sports medicine, to link age to heart rate (Gellish et al., 2007). In epidemiology (a discipline that could be considered to straddle both health science and social science) it is a widely used tool. For this reason there is an extensive literature dedicated to it, spanning from the end of the 70s (Rosner, 1979) to books published more recently (Twisk, 2003). It is also used in economics – particularly social economics, (Yankow, 2009)– and in demographics (Clark, 2006). In the educational field it has been utilized as a means to investigate, for example, the effects of the social and familiar environment in educative achievements (Crane, 1996; McFarlin, 2007). As final examples, we must cite research made within psychology and sociology (Orth et al., 2009; Brooks and Redlin, 2009; Guttmannova et al., 2008). The latter referenced work leads us smoothly from examples in applied research to those in basic research, as some terms belonging to the sphere of statistics feature in its title (e.g. invariance). And undoubtedly, longitudinal studies are, in essence, statistical studies. It is therefore logical that there are also a large number of works on ‘basic statistics’ that define, improve, and refine the tool itself (Cepeda and Nuñez, 2007; Lee, 2006; Tavares and de Andrade, 2006).

A key point is that what is being manipulated is data, and in the majority of cases, a large quantity of data. And because of this, naturally, it is understood that these data must be handled in an automated manner, that is, that some type of computerized support must be used. As it has been established that longitudinal studies are statistical studies, it is common to use statistical packages (programs) to manage, control and analyze the data. But it must also be emphasized that the quality that distinguishes longitudinal studies is their time extension. For this reason, it is not just the handling of the data that is important but its ongoing storage, which is where the use of databases naturally enters the scene. And it is here that the key term whose precise meaning must be determined is found: ‘longitudinal database’.

To avoid a simple misunderstanding, it is first necessary to distinguish the ‘longitudinal studies databases’, that is, databases that act as a catalogue of different longitudinal studies that are carried out – see for example the repositories in the (National Institute of Aging (USA), 2010) or in the

---

an event, will be studied.

(Bureau of Labor Statistics (USA), 2010). These database catalogues are undoubtedly necessary in an applied research context as they are useful for contrasting results of related experiments. Once this point has been established, a general definition of 'longitudinal database' can be found.

3. Longitudinal databases

As has just been highlighted, the term 'longitudinal database' is used in applied research when one wants to refer to a computerized structure that supports all the data set used in a specific longitudinal study. As an example, (Ander-Peciva, 2005) states that "a broad definition of a longitudinal database could be a mass of computerized data that is available for consultation and provides information about individuals over time". In any case, it is a relatively new term, not widely used.

If we analyze the situation from the point of view of research within computer science, particularly into information systems, and more specifically still into databases, a certain disparity in the use of the term database from its usual use can be seen in this other field. The key to this disparity provides us once again with the distinction between basic and applied research. Essentially, most of the works where the term 'longitudinal database' appears in the literature, it is in terms of applied research, when the (note the importance of the definite article the) longitudinal database is a specific one (one example, one instance), one specifically designed as support for carrying out the longitudinal study in question. There are few exceptions to this situation, to our knowledge, only (Vikström et al., 2002) and (Williams et al., 2010) discuss modelling issues for a longitudinal database. In any case, for basic research into databases, a general definition of the term 'longitudinal database' is needed, at the same general level as 'relational database', 'object–oriented database', 'deductive database', 'documental database' or 'multidimensional database'. Put another way, for basic research within databases it is necessary to define the type and the category of 'longitudinal databases'. It is therefore another example (yet one more) of the gap between basic and applied research. When a demographer, a historian or an epidemiologist is asked to define the concept of a 'longitudinal database' they do so based on the background knowledge they possess, and so it undoubtedly becomes difficult, if not impossible, for them to give a definition that would be accepted by the other fields. This is with the exception that the answer is, as in the definition at the beginning of this section, a notion that is practically indistinguishable from any 'database'.

In effect, the most basic, general and broadly accepted definition of 'database' is 'a structured collection of related data'. If this is compared with the definition given at the beginning, only one essential difference can be noted: the appearance of the reference to the aspect of 'time'. As it is a definition of 'longitudinal database' that is needed, if the reference to 'time' is crucial, it is evident that analysis and comparison of the subject under study with the notion of 'temporal database' must be made.

4. Temporal databases

The most broadly accepted definition is that a temporal database is a database in which the 'time' aspect is in some way relevant –see the best effort to clarify temporal database related concepts: 'The Consensus Glossary of Temporal Database Concepts' (C.S. Jensen and et al., 1993). Therefore, if the definition cited above for longitudinal database is accepted, it is found that a longitudinal database is nothing more than a particular kind of temporal database. This raises an apparent contradiction: if a longitudinal database is a particular kind of temporal database, has there not been sufficient research into temporal databases? A brief review of the literature clarifies this.

Temporal databases have been studied since the beginning of the '80s. It is significant that one of the first publications that directly relates 'time' to 'databases' is the 1981 article 'Displaying Clinical Data from a Time-Oriented Database' (Blum, 1981), that is, an applied research article, specifically medical research. But from the first references that can be considered to be basic research into the subject (two articles, one from 1979 (DeAntonellis et al., 1979) and the other from 1981 (Klopprogge, 1981), that raise the 'time' aspect in the entity-relationship model, and a doctoral thesis from 1982 entitled 'The Time Relational Model') until the middle of the '90s, approximately 900 references can be found –documented through various surveys and books, such as (AbdullahTansel and et al., 1993) and (Ozsoyoglu and Snodgrass, 1995)- that cover relative aspects of temporal databases or one of their 'subtypes', such as 'real time databases'. This large volume of publications is a clear indication of the interest aroused by 'temporal databases' in this period. However, the lack of industrial development transferring this enormous quantity of research into the commercial sphere is significant.

It is particularly surprising that in this period not a single producer of Database Management Systems (DBMS) decided to support any kind of temporal database model. There are two reasons that can, at least partially, explain this situation. One of them is the statement by Edgar Codd, inventor of the Relational Model of databases, in the preface of the book introducing Version 2 of his model, where he literally says that "in many cases [DBMS providers] failed to understand the first version of the Relational Model" (Codd, 1990). Codd's words suggest that if the DBMS did not faithfully take to a model as straightforward as the Relational Model, it would be more difficult to tackle more complex models. The second reason that conspired against the use of temporal DB tools was the lack of consensus during the standardization process. In spite of having a complete specification proposal for a language for temporal databases –TSQL2 (Snodgrass and et al., 1995)-, the initiative to integrate this specification as part of the SQL3 standard (by the inclusion of a new section within the standard, under the name 'Temporal/SQL') hung in limbo around 2001 due to the lack of agreement by the ISO representatives, as Richard T. Snodgrass, one of the principle promoters of the initiative (and a world expert in temporal databases), graphically describes in (Snodgrass, 2009). This anomalous situation has been highlighted once again very recently by (Johnston and Weis, 2010): in the preface of this book they state that 'in spite of several decades of work on temporal data [...] little has been done to help IT Professionals manage temporal data in real-world databases'.

Since the middle of the ’90s until the present day, and probably due to the unstoppable growth of everything related to the web and the internet, interest in temporal databases has noticeably diminished (it is in fact somewhat difficult to find 21st century publications on the subject). Of course this does not mean that the subject lacks importance, rather that the focus has been redirected towards other related matters, for example, spatiotemporal databases, a key piece in geographical computer systems (digital cartography services, GPS positioning, etc). In fact, it is strange that in the very latest versions of one of the most powerful DBMS (Oracle9i, 10g and 11g) tools that manage properties of temporal databases have been included (Flash-Back Queries (Weiss and Oracle Corporation, 2003), Workspace Manager (Oracle Corporation, 2009)). Finally, other initiatives that in some way keep an interest in temporal databases alive should be mentioned, for example TimeDB (Time Consult, 2005) implements a temporal DBMS through the use of a bridge to a standard DBMS; or TimeCenter (Jensen et al., 2010), an international centre for the application of temporal databases in traditional and emerging DBMS technology.

5. Longitudinal: more than just temporal

As has just been demonstrated, research into temporal databases has been intense and extensive for over 30 years. Moreover, longitudinal studies have been undertaken from even earlier times. The fact that the two have certain aspects in common is undeniable. And yet, as has been shown, a definition of ‘longitudinal database’ that can be used by the database community does not exist. It therefore seems appropriate to explore the reasons for this discrepancy.

Let’s return to the basic definition of ‘database’ as outlined above: ‘a structured collection of related data’. On adding the word ‘longitudinal’, it is natural to wonder in which part of the basic definition the ‘longitudinal’ aspect should fit. In other words, what is, (or what should be, always from a database point of view) ‘longitudinal? Is it the structure (the structured collection) that should have longitudinal characteristics? Or is it the data (and more importantly, their relationships) that are longitudinal? In view of the literature, it seems that it is the second concept that is correct, as the term ‘longitudinal data’ is a very common one. However, closer observation shows that this expression practically always appears with the expression ‘longitudinal data analysis’. It is here that, in the authors’ opinion, one of the keys to the aforementioned disparity can be found. In longitudinal studies, the most fundamental and relevant point is the data processing. It is from this processing that relationships between the data are obtained (e.g. that births in a particular year in a particular area of population have a greater probability of being under particular circumstances). In some sense, in longitudinal studies, how this data is stored, under what data model (in the computer sense of the word) is not so important: what is essential is for the data to be available and to work from them. In the authors’ opinion, specific data modelling techniques always reflect a biased reality representation, so taking into account that high quality of the manipulated data is demanded, the notion of longitudinal database must be researched in itself, trying to define structures and processes that promote the accuracy of the data involved in the longitudinal studies. The idea is that increasing the richness of the database model that is used as an ongoing support for data likely to be treated longitudinally, more effective data processing would be achieved. The objective is to improve the data quality in order to facilitate longitudinal studies, and for this it is essential to establish a specific data model. Temporal databases, as they are, allows the realization of cross-sectional studies, but little temporal database research is done with specific focus on longitudinal studies. A special emphasis on the longitudinal aspect of the related data is necessary in order to support longitudinal studies with maximum accuracy, , which is why the creation of a new model is necessary, as shall be discussed in the following section.

6. A longitudinal data management proposal

In order to give a brief overall description of our approach, highlighting its main features and achievements, there are four aspects that deserve to be explained: (1) our longitudinal approach, (2) our conceptual longitudinal database model, (3) a methodology for the realization of such a model and (4) a specific real-world longitudinal application.

Longitudinal related aspects. A relevant aspect of every longitudinal study is 'time'. For current databases, time is just one more piece of data, on the same level as numbers or text, and for this reason DBMS are limited to giving time ‘data types’. Temporal database models involve increasing the importance of time in such a way that it is not treated as just another kind of data, but receives a specialized treatment. But even so, in these models, time is considered as a dimension (as linear, in the vast majority of cases) in which the most important characteristic is therefore the order (one piece of time data is greater or lesser than another; or, in more colloquial language, one moment comes before or after another). In our approach, taking into account that the objective is longitudinal knowledge, being necessary to follow the event history of a cohort, time is considered not just order, but fundamentally, change.

We consider that time is perceived through the change observed in the state of an object. If we did not observe change in the objects around us, we would not be capable of perceiving time. Moreover, we consider that there are three fundamental aspects related with change: (1) what causes the change, (2) the object affected by the change and (3) what has happened. We, the authors, understand that this form of interpreting time is innovative, and in fact have only located one ‘position paper’ sharing this idea (Letondal and Tabard, 2009).

Every system is dynamic and, therefore, it is subject to changes. However, we restrict our attention to reactive systems (they are continuously reacting to external inputs) and besides we only consider reactive systems whose external inputs are always produced by the execution of a protocol. In this way, the protocols act on one or more objects of the system which react changing their state. Within this particular context, the perception of change is restricted to the outcomes of the execution of a protocol on an object that are shown as a change of the object state. In this way, all the event history that happens to an object through pre-established protocols is registered, so that the cohort can be considered to be constantly under observation, which is a main requirement of longitudinal studies (Vikström et al., 2002). More formally, the change can be represented as a set of three elements comprising (1) the execution of a protocol at a given moment in time, (2) the object on which it has been executed and (3) the data related to the resulting object state.

Conceptual Longitudinal fact base Model. Starting from this longitudinal approach, we define a conceptual model of longitudinal fact base. According to our proposed model, a longitudinal fact base is a structured system of storage of facts which captures occurrences perceived along the time during the execution of a protocol. The term fact base has been used in the literature to refer to storage structures in which the time was not explicitly considered –see, for example, (Ramirez and de Antonio, 2007). For this reason, in order to distinguish our notion, we have considered it appropriate to incorporate the term 'longitudinal' in its name.

The facts are represented as three-tuples (p, o, d), being p a protocol that has been applied at time t to an object o, giving rise to data d which are associated to the resulting state of o. By means of this structure, the longitudinal fact base stores the states which the objects are going through as a consequence of the executed protocols, both the protocols that are currently in force and the protocols that are not valid at present but were applied in the past to the objects of the system. The three-tuples are stored progressively in the longitudinal fact base, it not being possible to delete them with the goal of keeping a complete track of the system occurrences, which is very useful for data provenance and auditing purposes.

The longitudinal database model also defines several longitudinal objects which are oriented to the representation of facts and data along the time. The main longitudinal object is the lifeline. Every lifeline is associated to an object showing the occurrences registered in the system about the object and allowing the dynamic access to the traces of the possible objects related with it. The lifelines, together with the objects and data traces, allow the extraction of evidences from the fact base, so the user can be aware of the protocols that have been applied to generate such evidences, whether these protocols have changed over time and to know the states of objects from which those evidences were obtained. In this way, the proposed longitudinal fact base model makes it feasible to assure the accuracy of data this being a great contribution towards data quality assurance.

Methodology for the construction of longitudinal fact bases. A methodology for the construction of longitudinal fact bases has been developed following the Model-Driven Architecture (MDA) approach (Kleppe et al., 2003). In particular, within this development method, we propose to use UML for the specification of the platform independent models (PIM), the relational model for the specification of the platform specific models (PSM) and the SQL of a DBMS for source code specification (we refer to (Domínguez et al., 2013) for a more complete description). We have chosen UML due to it provides extension mechanisms by means of UML profiles and stereotypes so that UML models can be adapted to specific contexts. In particular, the most noteworthy aspect of this methodology is that we have defined the Occurrence UML profile which embraces the basic concepts of our proposal and, on the other hand, the Change State pattern which, by means of a UML class diagram, describe the basic relations among such concepts at a generic level. Both artifacts (profile and pattern) gather, in its architecture, the three longitudinal fact base dimensions we propose (protocol, object, data), the persistence of its elements as well as aspects directed to increasing the effectiveness of the system to answer queries.

These two artifacts are used as base for constructing the longitudinal fact base following a MDA approach. More specifically, such artifacts are used to establish the longitudinal PIM model.

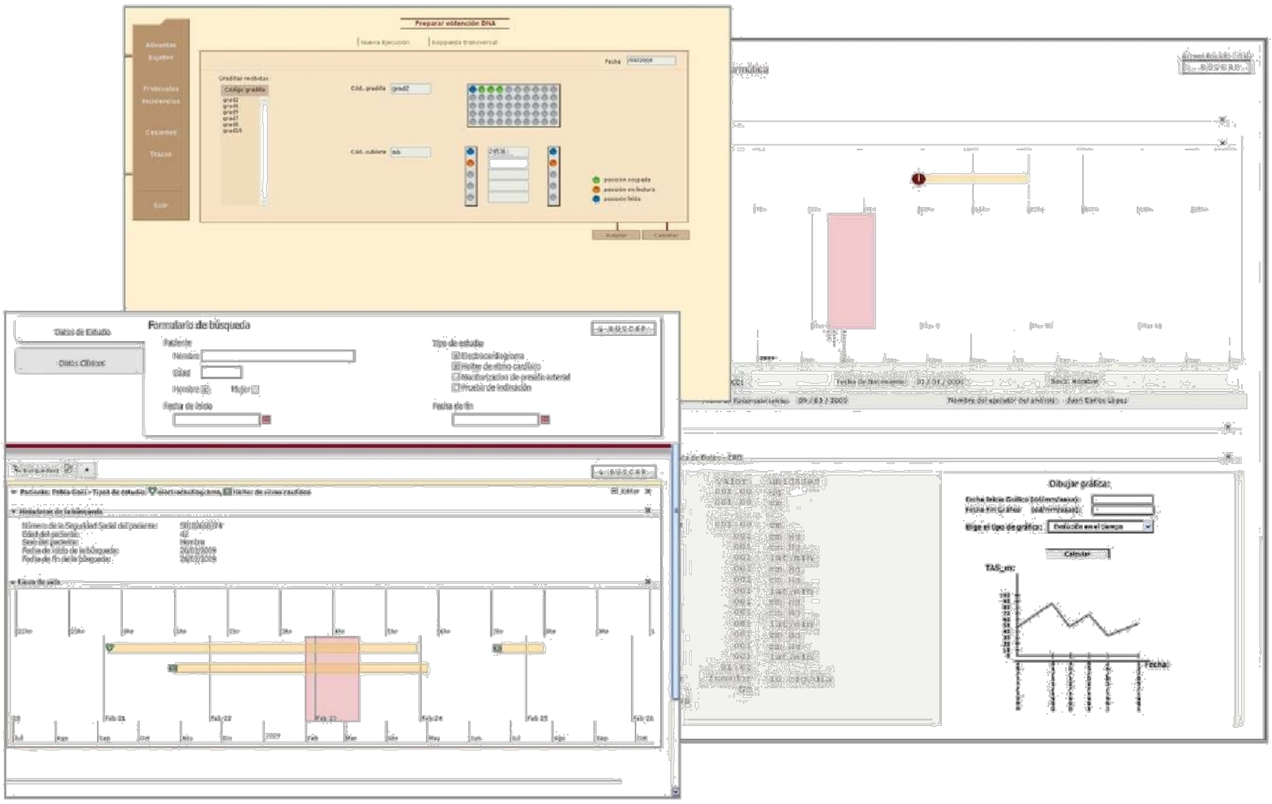

Figure 1. Screens of our biobank management tool

Starting from the PIM model, the method also provides specific transformations to transform this model to a PSM model, by means of which new tables are included in the PSM model which store the information related to the profile stereotypes. These transformations are defined embracing conditions and quality aspects of construction of a longitudinal fact base.

This strategy, although being a particular one, presents theoretical concepts susceptible of being applied using other different methodological and/or technological approaches, which gives outstanding value to our proposal.

Real-world longitudinal application. As a real-world implementation case, our methodology has been verified and validated through the construction, among others, of a biobank management tool (see Figure 1). The tool is being tested in the open exploitation stage as computer support in the ‘Aragon Workers Health Study’ project (Casasnovas et al., 2012) which is funded by the National Center of Cardiovascular Research Carlos III and the Autonomous Region of Aragon. This project aims to study health conditions over time of a cohort of over 5,000 workers in the General Motors Spain production plant, located in Figueruelas (Zaragoza). The tool, which enables the management of information related to the longitudinal history of a biological samples bank, is being used successfully by users (laboratory technicians and physicians) as it guarantees the dynamic building of evidences of the traceability of both the objects (such as racks and aliquots) and the data.

7. Conclusions and further work

There are two main conclusions that can be obtained from the ideas presented in this paper. The first conclusion is that it is clear that longitudinal studies (as tools for applied research) are paramount in a multitude of scientific disciplines. In addition, the volume of data which is handled in such studies is very large, and it is a clear requirement that such data need to be sustained over time with high quality. Automated support is required to ensure properly structured information, to maintain consistency, to visualize the evolution of data and so on.

Therefore, and as a second conclusion, the authors advocate that a proper and sound notion of what should be a longitudinal database is required since, as has been shown, such a definition does not exist in the literature. In this paper we have outlined very briefly our proposal for a longitudinal fact base model. Our approach is being validated in a real context in which the Model Driven Architecture (MDA) approach is being used.

As further work, our research group has started to analyse how to store, within our conceptual longitudinal fact base model, non-protocolized occurrences along the longitudinal study. This aspect is important in order to being able to store any type of evidence along the study and also due to its relevance for security management and incident management.

Acknowledgements

This work has been partially supported by the Spanish Ministry of Science and Innovation (projects TIN2009-13584 and SMOTY, IPT-2011-1328-390000), the Centre for Industrial Technological Development (CDTI) (project THOFU, CEN20101019), and the University of La Rioja, project API12/12

References

AbdullahTansel and et al. (1993). *Temporal databases: theory, design, and implementation.* Benjamin-Cummings Publishing.

Ander-Peciva, S. (2005). The construction of longitudinal databases. In International Commission for Historical Demography, editor, *New sources for historical demographic research, Session in World History Conference*, Sydney.

Blum, R. (1981). Displaying clinical data from a time-oriented database. *Comput Biol Med*, 11(4):197–210.

Brooks, W. T. and Redlin, M. (2009). Occupational aspirations, rural to urban migration, and intersectionality: A comparison of white, black, and hispanic male and female group chances for leaving. *Southern Rural Sociology*, 24(1):130–152.

Bureau of Labor Statistics (USA) (1979–2010). *National longitudinal surveys*. Last visited, February 19, 2013. http://www.bls.gov/nls/.

Casasnovas, J. A., Alcalde, V., Civeira, F., Guallar, E., Ibanez, B., Jimenez-Borreguero, J., Laclaustra, M., Leon, M., Ordovas, J. M., Pocovi, M.; Sanz, M., Fuster, V. (2012) Aragon workers' health study – design and cohort description. *BMC Cardiovascular Disorders* 12 (45)

Cepeda, E. and Núñez, V. A. (2007). Bayesian joint modelling of the mean and covariance structures for normal longitudinal data. *Sort: Statistics and Operations Research Transactions*, 31(2):181–199.

Clark, S. J. (2006). A general temporal data model and the structured population event history register. *Demographic Research*, 15:181–252.

Codd, E. F. (1990). *The relational model for database management: version 2*. Addison-Wesley.

Crane, J. (1996). Effects of home environment, ses, and maternal test scores on mathematics achievement. *Journal of Educational Research*, 89(5):305–314.

C.S. Jensen and et al. (1993). *A consensus glosary of temporal database concepts*. Technical report, Last visited February 19, 2013. http://people.cs.aau.dk/~csj/Glossary/.

DeAntonellis, V., Degli, A., Mauri, G., and Zonta, B. (1979). Extending the entity-relationship approach to take into account historical aspects of systems. In Chen, P. P., editor, *Proceedings of the International Conference on the E-R Approach to Systems Analysis and Design*, pages 231–236. North-Holland.

Delgado, M. and Llorca, J. (2004). Estudios longitudinales: concepto y particularidades. *Revista española de salud pública*, 78(2):141–148.

Domínguez, E., Pérez, B., Rubio, Á.L., Zapata, M. A., Lavilla, J., Allué, A. (2013). Occurrence-Oriented Design Strategy for Intelligent and Quality Systems Development. Submitted to IEEE Transactions on Knowledge and Data Engineering.

Dubbelinka, K.T.E.O., Stoffersa, D., Deijenb, J.B., Twiskc, J.W.R., Stamd, C.J., Berendsea, H. W. (2013). Cognitive decline in Parkinson's disease is associated with slowing of resting-state brain activity: a longitudinal study. *Neurobiology of Aging*, 34: 408–418.

Ferguson, K. J. and Kreiter, C. D. (2004). Using a longitudinal database to assess the validity of preceptors' ratings of clerkship performance. *Advances in Health Sciences Education*, 9(1):39–46.

Gellish, R. L., Goslin, B. R., and Olson, R. E. (2007). Longitudinal modeling of the relationship between age and maximal heart rate. *Medicine & Science in Sports & exercise: Official Journal of the American College of Sports Medicine*, 39(5):822–829.

Guttmannova, K., Szanyi, J. M., and Cali, P. W. (2008). Internalizing and externalizing behavior problem scores: Cross-ethnic and longitudinal measurement invariance of the behavior problem index. *Educational and Psychological Measurement*, 68(4):676–694.

Jensen, C. S., Snodgrass, R. T., and et al. (2000–2010). *Timecenter*. Last visited February 19, 2013.

http://timecenter.cs. aau.dk/.

Johnston, T. and Weis, R. (2010). *Managing Time in Relational Databases. How to design, update and query temporal data*. Morgan Kauffman.

Kleppe, A., Warmer, J., and Bast, W. (2003). *MDA explained. The Model Driven Architecture: Practice and Promise*. Addison-Wesley.

Klopprogge, M. R. (1981). Term: An approach to include time dimension in the entity-relationship model. In Chen, P. P., editor, *Proceedings of the Second International Confer-ence on the Entity- Relationship Approach to Information Modeling and Analysis*, pages 473–508. North-Holland.

Lee, S. K. (2006). On classification and regression trees for multiple responses and its application. *Journal of classification*, 23(1):123–141.

Letondal, C. and Tabard, A. (2009). Temporal data and data temporality: Time is change, not only order. In Mackay, W., Kleek, M. V., and Tabard, A., editors, *CHI 2009 workshop Interacting with temporal data*. http://www.lii-enac.fr/~letondal/Papers/LetondalTabard-temporal-CHI09-v10x.pdf.

Looso, S., Goeken, M. (2010). Application of best-practice reference models of it governance. Proceedings of the 18th European Conference on Information Systems.

Luse, A., Mennecke, B., Triplett, J. (2013). The changing nature of user attitudes toward virtual world technology: A longitudinal study. *Computers in Human Behavior*, 29: 1122–1132.

McFarlin, I. (2007). Do school teacher parents make a difference? *Economics of Education Review*, 26(5):615–628.

National Institute of Aging (USA) (2003-2013). Population studies database. Last visited February 19, 2013. http://nihlibrary.ors.nih.gov/nia/ps/niadb.asp.

Oracle Corporation (2009). Oracle workspace manager. Last visited February 19, 2013. http://www.oracle.com/technology/products/database/workspace_manager/.

Orth, U., Robins, R. W., and Meier, L. L. (2009). Disentangling the effects of low self-esteem and stressful events on depression: Findings from three longitudinal studies. *Journal of Personality and Social Psychology*, 97(2):307–321

Ozsoyoglu, G. and Snodgrass, R. T. (1995). Temporal and real-time databases: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 7(4):513–532.

Ramirez, J. and de Antonio, A. (2007). Checking the consistency of a hybrid knowledge base system. *Knowledge-Based Systems*, 20(3):225–237.

Rosner, B. (1979). The analisis of longitudinal data in epidemiologic studies. *J Chron Dis*, 32:163–179.

Snodgrass, R. T. (2009). *Tsql2 and sql3 interactions*. Last visited February 19, 2013.

http://www.cs.arizona.edu/people/rts/sql3.html.

Snodgrass, R. T. and et al. (1995). *The TSQL2 temporal query language*. Kluwer.

Tavares, H. R. and de Andrade, D. F. (2006). Item response theory for longitudinal data: Item and population ability parameters estimation. *Test*, 15(1):97–123.

Time Consult (2005). TimeDB - a temporal relational DBMS. Last visited February 19, 2013. http://www.timeconsult.com/Software/Software.html.

Twisk, J. (2003). *Applied longitudinal data analysis for epidemiology. A practical guide*. Cambridge University Press.

Vikström, P., Edvinsson, S. and Brändström, A. (2002), Longitudinal databases - sources for analyzing the life-course: Characteristics, Difficulties and Possibilities, *History and Computing*, 14(1-2): 109-128.

Weiss, R. and Oracle Corporation (2003). How Oracle database 10g revolutionizes availability and enables the grid. Last visited February 19, 2013. http://www.oracle.com/technetwork/oem/grid-control/10gavailability-133799.pdf

Williams, B., J. Dowell, G. Humphris, M. Themessl-Huber, R. Rushmer, I. Ricketts, P. Boyle and F. Sullivan (2010). Developing a longitudinal database of routinely recorded primary care consultations linked to service use and outcome data. *Social Science & Medicine*, 70(3): 473-478.

Yankow, J. J. (2009). Some empirical evidence of the efficacy of job matching in urban labor markets. *International Advances in Economic Research*, 15(2):233–244.

Yeh, C., Chiang, Y., Lin, L., Yang, C., Chien, L., Weaver, M., and Chuang, H. (2008). Clinical factors associated with fatigue over time in paediatric oncology patients receiving chemotherapy. *Br J Cancer*, 99(1):23–29.