

INTEGRATION OF DATA ON THE LONG-TERM CARE FOR ELDERLY PEOPLE FROM HETEROGENEOUS SOURCES FOR RESEARCH PURPOSE

Christine Pelletier¹, Thierry Chausalet¹, Nick Szirbik², Christos Vasilakis³

¹ Department of Mathematics, Cavendish School of Computer Science, University of Westminster, UK

² Information Systems Cluster, Management and Organisation faculty, University of Groningen, NL

³ Harrow School of Computer Science, University of Westminster, UK

Abstract: The rapid increase in the demand for long-term care of elderly people leads to a need for a better understanding of resource use, therefore a better knowledge of the survival patterns. The data necessary to model survival and resource use in long-term care are available, but distributed among different institutions – purchasers of care, e.g. Local Authorities in the UK, and providers of care in nursing and residential homes and in the community. Building models of survival and resource use requires the availability of historical data over a long period of time. An added difficulty is the anonymous and confidential nature of the data. In this paper, we propose a methodology for the design of a database that integrates anonymous data from such heterogeneous sources. A procedure allowing for systematic update of the database is also presented.

Introduction

The placement of elderly people in long-term care (LTC) is often controlled and monitored by local authorities, as is the case in the UK. Local authorities are responsible for the placement of each person, referred to as clients, in appropriate care facilities within the constraints caused by the scarcity of resources, e.g. availability of beds and budget [1]. There exist mainly three kinds of LTC services provided. LTC clients may be kept in their own home with some form of community support, or placed in institutional LTC, either in a residential or a nursing home. By definition, nursing home clients are sick and dependent whereas residential home clients are frail and less dependent and the persons supplied with services at home are less dependent. There is an obligation for a local authority to supply the care needed by each person it is responsible for, whether in the person's own home or in an institution.

Between 1982 and 1993, the number of older people in Britain in residential and nursing home care increased from 16,000 to 281,200. In the same period, the cost met by the social security system increased from £9 million to £2,575 million [2]. As the UK population ages, it is believed there will be an unsustainable increase in the need for LTC [3], and its corresponding costs. However decrease in expenditure of the Danish home and community-based LTC system [4] suggests that costs could be controlled with better management of the available financial and human resources. In this context, there is a growing need to tightly manage elderly LTC resources. A first step in the improvement of the management of elderly LTC resources is to understand and model survival

within the elderly LTC system, as the associated resource consumption and costs [5].

However this approach is hindered by the lack of structured historical data. According to [6], we face a case of semi-structured data, different sources, with the same central entity as central object to focus on: the aged LTC client. The data concerning the elderly people supplied with LTC are recorded in different sources. Medical data of the elderly people at home, with or without services, are stored in the general practitioner register. Data linked to a hospital sojourn are stored in the hospital register. The death register gathers data related to death (date of death, reason of death...). And the data concerning institutional LTC are present in two main registers, one owned by the local authority and one by the care provider, i.e. the organisation which owns the home where the client is placed. The access to all these registers is strongly regulated due to the data protection act. This protection leads to great difficulties to compare the different files in order to build historical data. Even without this protection, the data collected in each of the registers are seldom structured towards an historical perspective; they are mainly flat yearly files.

The main goal of this paper is to present the design of the database capable to integrate data from heterogeneous sources, thus to support the research on survival and the development of corresponding survival and costing models. We also present a procedure allowing the updating of the database. Before presenting the technical details, we describe the characteristics of the data available. Then the structure of the database is explained and a way to update the database is presented. We then discuss the advantages of this approach and possible further improvements to the updating of the database.

Data characteristics

The data of interest concern transactions about elderly people in LTC over a long period, in some cases more than 10 years, and are collected from different sources, as mentioned in the introduction. We focus on the life-cycle of elderly clients of LTC. A life-cycle corresponds to a succession of events in time. The first record for an object corresponds to the first time the client enters LTC – either Home + service, Residential home, or Nursing home (see Figure 1). Any LTC service can be accessed from home or from hospital. Then, a succession of discharge to hospital or home and admission to one of the LTC state (service) can occur until death. The

events experienced by a LTC client when s/he is outside the LTC system are not usually recorded in any LTC database.

In Figure 1, the trajectory of a LTC client is modelled via a Finite State Automaton (FSA). The nodes represent the different states a client can be in, and the arrows, the possible transitions, transfer, events a LTC client can go through, that are recorded in LTC data sets.

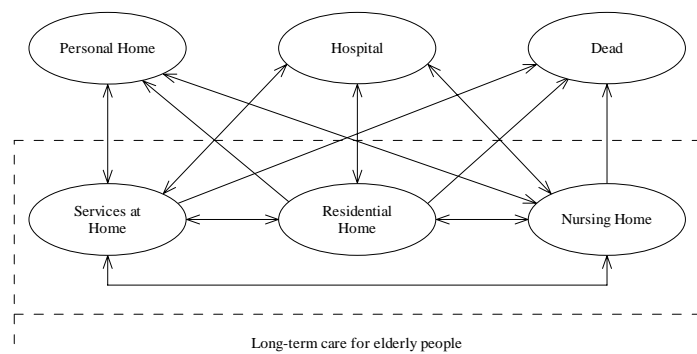


Figure 1: Data object life-cycle for LTC for elderly people.

In our research, administrative records represent the main source of raw data. The major disadvantage of this kind of information is that most of the data is recorded on paper or in the best case in very old legacy systems that still use flat files. We managed to obtain the data in electronic formats from all the organizations involved in the project (Local Authorities and Home Cares). However in some situations, the transfer of data from paper to the files implied serious costs, both in human resources and needed infrastructure. A recurring problem is that this transfer introduces supplementary errors.

Fortunately our data collection task is facilitated by the historical nature of the data, and their very slow dynamics. This is the case of both the database content (records are updated and inserted with a low frequency), and the database models.

Information contained in different sources can overlap. For example, records of clients who are placed by social services in a home for elderly people will appear in both the social service database and in the database of the home where s/he is placed. These two files are not identical, therefore do not completely overlap because some clients of the home can be either private clients or sent by a hospital. Moreover the home may have other beds allocated to different social services.

Database structure

Using the life-cycle of a client, we identified two main entities in our universe of discourse: Client and Sojourn. These two entities are related to each other, as in figure 2, which depicts the conceptual data model as an Entity-Relationship (E-R) diagram.

Using this conceptual model (fig. 2), we built an implementation (detailed design) model for the database, using the UML modelling language (fig. 3).

The choice of type of care induces the choice of the home type, whether the care is institutional or not. Also, the same choice starts the negotiations for the weekly cost, and the payment schema. These two latter factors can evolve during

the sojourn of the client. A sojourn ends when the client changes home or type of care.

Figure 2: E-R based conceptual data model for LTC

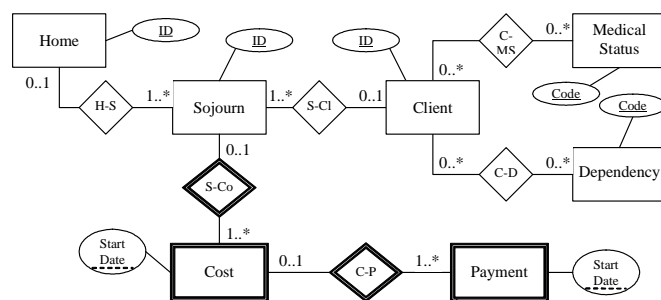


Table 1: Relational Database Schema

Relation	Primary Key	Attributes
Client	ID	Gender BirthDate
Sojourn	ID	ClientID HomeID Admission_date Discharge_date Adm_originated_at Discharge_reason Care_type Location Type Economic_status Economic_type
Home	ID	EndDate Weekly_amount Negotiation_date Cancellation_reason
Cost	SojournID StartDate	Reason_for_change HA_amount LA_amount SocSec_amount SelfFund_amount Description
Payment	SojournID CostStartDate StartDate	Description
Standard_Medical	Code	EndDate
Standard_Dependency	Code	EndDate
Medical_Status	ClientID MedicalCode StartDate	EndDate
Dependency_Status	ClientID DependencyCode StartDate	EndDate

Updating method

Our goal is to extract from separate and partially available data sets historical data corresponding to the events described in Figure 1. To retain this database for future projects, data must be completely anonymised. We have to devise a strategy to enable systematic update of the database when additional

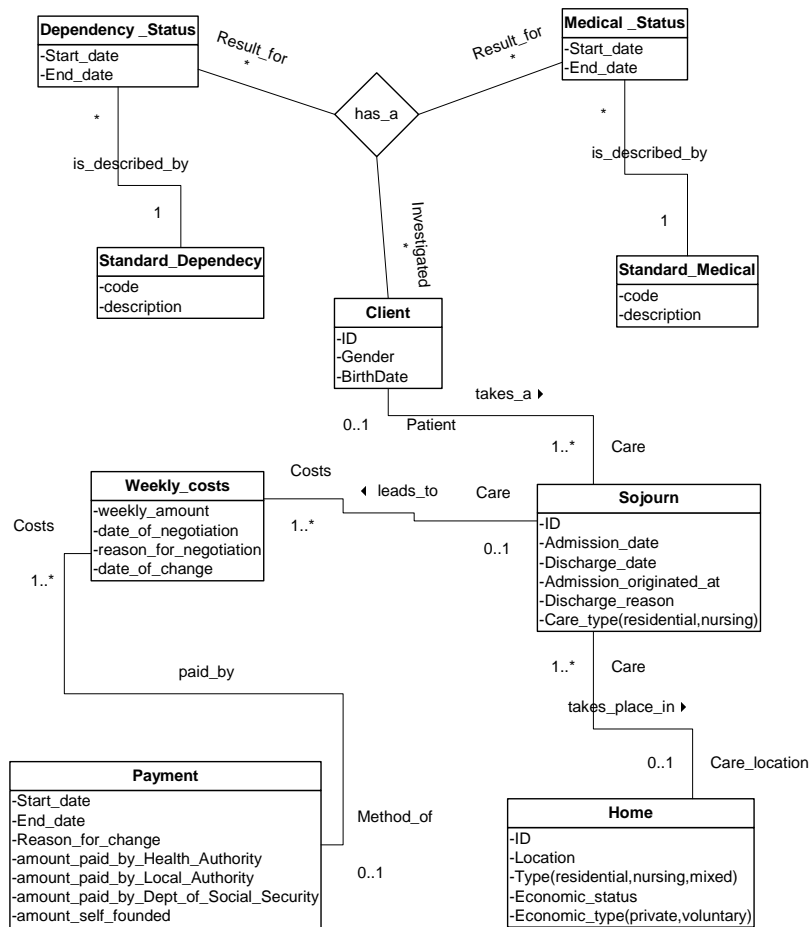


Figure 3: UML detail design model of the database

data. The method we propose can cope with relatively large numbers of records.

The updating process is done in two phases. The first phase consists in sorting the records related to new LTC clients for the database and clients that have already experienced a LTC stay and were previously registered in the database (Figure 4). Then the database is updated during a second phase (Figure 5).

Because data come from different sources, the same client usually has a different id in each data set. It is then necessary to keep in an id file a record of the different ids used for the same client, and of the source of these ids. Figure 4 shows the FSA associated with the sorting process. In our case, we consider two kinds of sources: Local Authority and Homes. We assume that a client returning to the same home will keep the same id in the database. Moreover because it rarely occurs, we assume that it is impossible for a client to migrate from one Local Authority to another.

If the id of a new record, which already existed in the database, appears in the id file, then the new record is labelled as “LTC client”. Otherwise, the process checks if the client concerned with the new record can be considered as a new input.

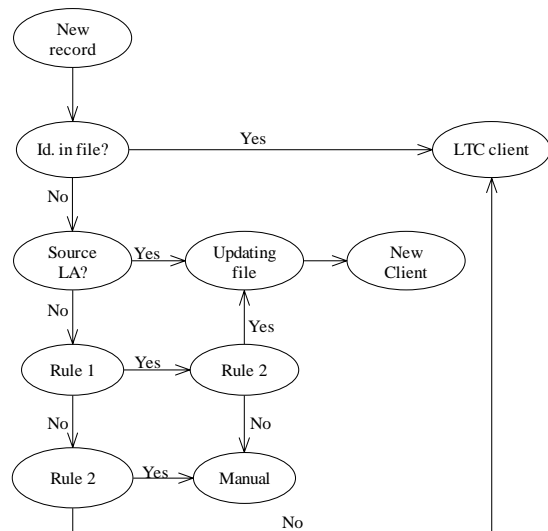


Figure 4: FSA of the sorting process.

Using our assumption concerning Local Authority clients, if the source is Local Authority, the record can be labelled as “new client”. If the source is a Care Home, then one checks if the new record matches a record already in the database (Rule

1: Does exist in the database no record with the same date of birth and gender? Rule 2: Are the date of admission in the two records different?). If it is the case, the record is considered as that of a “LTC client”. If the record has nothing in common with any record existing in the database, then it is associated with a “new client”. A case of uncertain match has to be solved manually.

Following the sorting process, the database is updated according to the process depicted by Figure 5. Depending on the type of client (new or not), the database is either directly updated - creation of new records in each table – or the fields of the new record are tested before deciding if the update can be done automatically or need to be done manually. We choose not to update everything automatically to avoid populating our database with suspicious data susceptible to deteriorate the quality of the database.

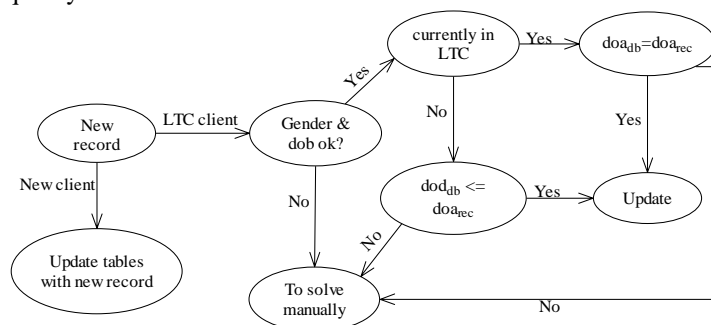


Figure 5: FSA showing the database updating

Discussion

The database resulting from the approach presented above is dedicated to the estimates of survival models and estimation of costs related to groups of clients.

Considering that data have been collected from various sources, a problem that need to be addressed is that database models (schemas) can differ significantly. A semantic alignment is in this case necessary for the integration of these data: Fields apparently similar may have identical names but different meanings, or conversely different names and same meanings. Our search for the description of the word “long-term care (LTC) for the elderly” (and merging the different points of view presented by different sources) leads us to the definition of the “common shared ontology” that the database structure (Figure 2) illustrates a conceptual point of view. This ontology acts as a reference describing the whole system currently under study. Each time a new file (set of data) is added to the repository, each field (entity type and entity attribute) is linked to the common shared ontology. The builder of the repository should ask people in the data source organisation the exact meaning of each field in the provided file and map that name over a term in the ontology (if none, a new term has to be introduced there, together with explanatory relations to existing terms). By using XML technologies, this can be done automatically. Because the dimensions considered within the study of LTC can increase, this ontology may evolve too. From the database point of view, the translation of this evolution will be the addition of tables or fields within an existing

table. Sometimes the source data organizations migrate their files to a newer format, with new fields. This may cause a problem when newly introduced data is collected again. If the new data model drops some of the previous fields, the integration method used previously must be changed manually.

Acknowledgments

We thank Mr. Peter Crowther and Ms. Teresa Temple from the London Borough of Merton Social Services Department for their collaboration on this project, in particular data provision. We also thank Professor Peter Millard for his advice on clinical issues. This work was supported by the Engineering and Physical Sciences Research Council (Grant GR/R86430/01).

References

- [1] K. BAY, P. LAETT and S. STINSON (1982): ‘A patient-classification system for long-term care’, *Medical Care* **20** pp. 468-488
- [2] H. LEWIS, G. WISTOW, S. ABBOTT and L. COTTERILL (1999): ‘Continuing health care: the local development of policies and eligibility criteria’, *Health Soc Care Community* **7** pp. 455-463
- [3] J.A. DARVEY (1999): ‘Exploring shared options in funding long-term care for older people’, *Health Soc Care Community* **6** pp. 151-157
- [4] M. STUART and M. WEINRICH (2001): ‘Home- and Community-Based Long-Term Care: Lessons From Denmark’, *Gerontologist* **41** pp. 474-480
- [5] C.M. PELLETIER, T. CHAUSSALET and H. XIE (2004): ‘Forecasting the costs of institutional long-term care: a local authority perspective’, *Journal of Operational Research Society*, submitted
- [6] J. PARSONS and Y. WAND (2002): ‘Property-Based Semantic Reconciliation of Heterogeneous Information Sources’, in S. SPACCAPIETRA, S.T. MARCH and Y. KAMBAYASHI (Eds.): ‘Conceptual Modeling - ER 2002: 21st International Conference on Conceptual Modeling’, (Springer, Heidelberg), pp. 351-364

Address for correspondence

Dr. Thierry Chausalet, Department of Mathematics, Cavendish School of Computer Science, University of Westminster, 9-18 Euston Centre, London NW1 3ET, UK