Incorporating Family and Social Structures into Clinical Documentation and Retrieval

W. Gall, G. Duftschmid, P. Sachs

Department of Medical Computer Sciences

University of Vienna

Address:

Walter Gall
Department of Medical Computer Sciences
University of Vienna

General Hospital
Spitalgasse 23
A-1090 Vienna
Austria
Tel.: +43-1-40400 / 6694 (6698)
Fax.: +43-1-40400 / 6697
E-mail : walter.gall@akh-wien.ac.at

Incorporating Family and Social Structures into Clinical Documentation and Retrieval

W. Gall, G. Duftschmid, P. Sachs

Department of Medical Computer Sciences
University of Vienna

Abstract

Referencing familial and social relationships between patients supplies valuable information for the retrieval and interpretation of clinical data. We present a technique for the incorporation of patient relations into data retrieval that takes into account the specific properties of routinely collected clinical data. In most clinical databases, family relations are documented in a fragmentary manner at best. Furthermore, clinical retrieval systems do not support inter-patient queries in most cases. Our model is designed to formulate direct relations between patients and to identify patients as members of either temporary or persistent communities. In this way, the model supplies information on both genetic and social relations.

Key words

1 Introduction

Family relations are relevant in numerous areas of medical research. They are needed to understand genetically determined familial disease patterns, the effects of parents' diseases or social environmental factors on children, genetic risk factors in would-be parents, the effects of complications during pregnancy, to name only a few areas.

Yet medical information and retrieval systems usually have a patient-oriented design. That is to say, the data of a given patient are summarized in a patient record, and any queries referring to these data only pertain to that specific individual. However, as mentioned before, it is frequently desirable, both for therapeutic and scientific purposes, to look at relations across patient records, notably at family members. If a newborn shows manifestations of a specific disease, it would be very helpful to find out whether the parents' medical history points to a predisposition, thus extending the pathogenesis as it were into the past.

In the clinical routine, inter-patient relations are documented in a small number of specific contexts. Kinship relations are routinely captured in places like:

- *Birth clinics* – referencing the mother's and child's IDs as belonging together.

- *Genetic counselling clinics* – linking the data of would-be parents in specific ways to assess their chances of reproduction.

- *Hypertension clinics* – exploring inherited disposition as an important risk factor.

As an important spin-off, documenting kinship relations also facilitates the process of identifying and linking patient records [1]. Correct identification after multiple births is a major problem in this context [2], as it is difficult to distinguish between individuals when the key data attributes are identical. Given the case of automated record linkage being applied to the data of triplets whose only distinguishing mark would be their first names that are not yet known at the time of documentation, this might result in a single patient record being created. Integrating kinship relations therefore contributes to the overall data quality.

An obvious prerequisite for any scientific use of these kinship relations is that there exists an appropriate model. Family relations are a classical example for applying modeling and

visualization techniques [3]. In the old relational database systems it was a complicated task to retrieve hierarchical recursive relations like 'grandfather' as 'father's father'. Modern SQL does contain recursive search functions *([CONNECT BY...], [START WITH ...])* [4]. In addition, while previous models were also designed with a view to minimizing storage space [5], this is no longer a main issue today.

Rather, the main problem in linking related persons arise from the typical system's patient-oriented design and the lack of family-related information. Most systems only include simple relations like child>mother. Even these, however, are frequently irretrievable because they are 'hidden' in the patient records. Therefore, mechanisms are required for the management of incomplete information.

Furthermore, the model should be flexible, adjustable and expandable. The way in which kinship terms are used to classify relations varies with culture, language and time. In German, for instance, there have been a multitude of interpretations for 'cousin' throughout history [6]. In modern English, there exist around 20 kinship terms [7].

While genetic analyses [8] are easily the most important application for retrieving familial relationships, it is nevertheless also important to consider other relations like social contacts. For example, being able to retrieve spatial contacts within large hospitals can greatly enhance the effectiveness of infection control.

In chapter 2 we describe different types of inter-patient relations. In chapter 3 we present a model for the retrieval of relations between individuals and communities. In chapter 4 we show a possible integration in an existing retrieval system and conclude with a discussion of our approach in chapter 5.

2. Family relations

2.1 Consanguinity

Kinship relations play an important role in sociology and ethnology, to name only two disciplines. Relevant relations for medical purposes mainly pertain to inheritance. The first findings on the importance of genes in medical research date back to Gregor Mendel (1822-84) [9]. In medical practice there are hundreds of genetic disorders, representing examples of mendelian inheritance [10]. Mendel's laws of inheritance can be used to construct mathematically sound and meaningful models of prognosis. Furthermore, as the human genome is about to be decoded in its entirety, the issue of including kinship relations in clinical contexts is gaining new dimensions. Retrieval of kinship relations therefore mainly aims at consanguinity, a subset of family relations. Clinically, links to parents are of greater interest than links to spouses, the mother-child link being the most commonly referenced family relation.

2.2 Temporary family and social relations

While inter-patient relationships by blood are clinically highly relevant, other types of links are also worth looking at. Clinical interest centers on parents and children living together in the family context, where the entire spectrum of social factors exert their impact. Aside from this *nuclear family* with its biological-social double nature, there also exist family relations in a broader sense, as between parents and adopted children, an extremely close kind of social relationship.

From the clinical perspective, interesting social relationships include 'household communities' but also other types of communities in which a number of individuals share the same working place or occupy the same hospital room, etc. It is therefore desirable to include comprehensive information on the person's social environment in a model.

Therefore it should be possible to document and retrieve arbitrary 'clusters' such as *member(s) of a household* in addition to *member(s) of a family* only. By extension, social and spatial relations should also be implemented in a model.

One important aspect in investigating transmission paths of infectious diseases is to look at the spatial environment capturing personal contacts. In large hospitals such as the Vienna General Hospital with one million outpatient visits per year and 2200 beds, where infection control is a high-priority issue, spatial relations between people in clinics and wards are a highly relevant issue. For example, documentation and retrieval of contacts within a ward can be used in analyzing bacterial strains and antibiotics resistance [11]. Studies of this type become possible by automatically capturing admission and discharge dates, so that the fact that any two patients inhabited a specific ward at the same time is made plain.

Additionally, the *time factor* plays an essential role in data documentation because family configurations and household communities may change over time. All 'memberships' must include a time reference in order to illuminate its temporal dimension in relevant queries (e.g. 'Search for all patients living in the same household *for more than five years*', 'Search for all patients occupying the same hospital room *at the same time*').

2.3 Characteristics of clinical routine documentation

Kinship relations are numerous. Family assessments will consider not only genetic relationships but, depending on the application and cultural factors, different sets of additional relations like marriage, divorce, remarriage, polygamy or adoption. There exist various models to minimize the multitude of kinship relations, replacing one term by stringing others together [7] [12]. Most of these approaches emphasize patrilineal descent, as in substituting 'fathers' wife' for 'mother'. Unlike in social applications, however, clinical models have to draw an accurate line between social and genetic relationships.

As opposed to theoretical models, in a practical clinical application database criteria have to be considered in addition to the specific requirements of medical documentation: As many family relations can be derived, one would usually generate these derivables in a dynamic manner when needed rather than storing all relations in the database.

A central problem in medical documentation, that needs to be addressed as well, is incomplete information. We cannot take for granted that all direct inter-patient family relations are reflected in the database. Therefore, if the direct path is not documented, there must be alternative paths to execute queries (e.g. 'sister' through 'daughter of both parents'). Minimizing kinship terms, as proposed in theoretical models, is not an option in clinical documentation. In a practical system, replacing 'grandfather' by 'father – father' will fail if the father is not documented.

Analyzing clinical legacy data for attributes and relations relevant to data retrieval and interpretation, we obtain the following results:

Attributes – The attributes 'sex' and 'date of birth' are well documented in most legacy databases. In data retrieval, dates of birth are mainly used to track the sequence of siblings.

Relations – The central relation is the mother-child link. It is the only relation that is well documented in most systems, sometimes supplemented by the father-child relation. Relations to siblings are rarely documented and are only derivable if both, mother-child and father-child relations are documented. Therefore, in general vertical relations to direct ancestors and descendants (e.g. grandfather based on father-father) are more readily accessible than horizontal ones (e.g. uncle based on father-brother).

As an example for the implementation of inter-patient relations, the Vienna General Hospital's information system KIS provides only two primitive concepts for their representation:

1. Relation variable – Relations are established not externally between patient IDs but are documented as a normal variable in one patient. For a mother-child relation, the mother is stored as a normal item in the child's record. This requires that all patient records be screened for relations.

2. *Cases* – Cases can be thought of as ID sets. These IDs may originate in one (patient case) or several (group case) patients. Most *patient cases* are comprised of a set of documents, e.g. all documents pertaining to a pregnancy. Inter-patient relations can be established by combining patient cases to form a *group case*.

3 Model

3.1 Requirements

In the design of the data model and the retrieval functionality the specific properties of medical documentation have to be considered. To offer a flexible tool to the users, the following requirements should be satisfied:

- Since the available relations are usually incomplete, there need to be alternative routes to find related individuals. Additional knowledge on family relations (what is a cousin?) must be available.

- Frequently used relations (e.g. siblings) should be available 'ready made', i.e. statically present. Less frequently used ones (e.g. uncle) should be generated dynamically during execution of the query.

- The temporal aspect, which is of interest in many clinical issues, should be included (e.g. 'symptoms in the first child').

- Sex-specific differentiation should be possible (e.g. 'all female ancestors').

- The closeness of kinship should be definable (e.g. 'all male descendants to the second generation')

- The model should allow for cultural and social diversity (e.g. polygamy). Also, it needs to be adjustable to reflect new developments both in society and in medicine.

3.2 Static Layer

Fig. 1 shows an object-oriented data model in UML notation [13] that allows the expression of different kinds of relations between persons. This model creates the basis to document and retrieve/interpret inter-patient relations.

Figure 1

Links may either be created directly between two persons or indirectly between several persons over a community. These links are implemented by two abstract relation classes: The general *Association* class enables groupings of two or more instances of class *Person* by binding them to a common instance of class *Community*. The *Directed Association* class specializes the *Association* class by providing separate links to the two connected instances of class *Person*. This functionality is needed for expressing directed one to one relations between two persons.

The *Association* class references the *Relation Type* class to distinguish between different kinds of relations. Each concrete subclass uses its own dictionary of possible relation terms, shown in the corresponding notes in the class diagram. The dictionaries for the *Consanguinity* and *Temporary Kinship* classes are defined in a gender-independent way but may also be extended without having to change the model itself. If wanted, also relations of higher degree (e.g. Great grandparent) may be added to the list of class *Consanguinity*. In some cases (Sibling, Cousin, Marriage) an undirected link between two persons would actually be sufficient. We added these relation types to the dictionaries of the *Directed Association* subclasses to reduce the number of classes in the model.

Kinship relations are realized as directed links to capture the roles of the two persons in the relation itself. We distinguish between the persistent *Consanguinity* relation and the *Temporary Kinship* relation, where the latter inherits its time attributes from the abstract *Episode* class.

Groupings of several persons can be expressed by means of the *Temporary Membership* class. Hereby, two or more persons are associated with a common instance of class *Community*. Each instance of class *Person* may be member of any number of communities, and for each membership a specific time interval may be specified.

We are not interested in the possible communities as such, but only in groupings of persons. Therefore, as expressed by the cardinality of the association between the classes *Association* and *Community*, each instance of class *Community* must be connected to at least two instances of class *Person*. 'Empty' communities are irrelevant for us.

The *Geographical Community* subclass is provided to define groupings of persons, based on a certain location. Hereby, the corresponding persons are linked to one instance of

Geographical Community by means of *Temporary Membership* instances. The latter are annotated with the type of the relation to the community, which may for example be the person's common residence, working place or a hospital room they occupied at the same time.

3.3 Dynamic Layer

Different services are offered in the dynamic layer of the model for purposes such as enabling retrieval of particular instances or checking newly inserted relations for consistency with existing ones. In the object-oriented domain, these services are usually implemented by methods that define the behavior of the corresponding classes.

These methods are supported by additional knowledge such as:

- Inverse relations: For each relation there exists a corresponding inverse relation that may be of the same type (e.g. Sibling $\leftrightarrow$ Sibling) or different type (e.g. Parent $\leftrightarrow$ Child). These inverse relations have to be considered in the reasoning process.

- Degree of kinship relations: For some applications it may be necessary to refer to some measure for the degree of kinship relations. As an example, particular inherited properties reappear every second generation according to Mendel. Therefore, knowledge must be provided how to calculate the difference of generations between a person and one of her ancestors.

- Classification according to the cardinality of relations (e.g. a search for 'father' can be stopped after one hit, while a search for 'cousin' cannot).

The central services provided are *Retrieval of Relatives* and *Deduction of missing kinship relations.*

3.3.1 Retrieval of relatives

Identifying all persons that are in a certain kinship relation to a given individual seems to be trivial at the first glance, as we only have to go through the corresponding *Directed Association* subclass instances. However, as already mentioned before, we cannot expect that all existing relations of a person are actually defined. This means that even in the case of all

relevant persons being present in our database, the specific relation we are looking for may not have been registered. To handle this problem, we can make use of the fact that kinship relations are usually derivable. In other words, a particular relative of a person may be reached on alternative routes, if the direct relation is missing in the database. Hereby, the corresponding method creates several sets of persons, one for each alternative route, and unifies them, recognizing duplicates.

3.3.2 Deduction of missing kinship relations

For each kinship relation, knowledge must be provided how it may be derived from existing paths within a network of instances. As an example, the class *Person* may contain a method *unclePaths* that stores alternative routes for this particular relation: The first, direct route leads to all male persons that are in relation *SiblingOfParent* to the current person. The next, indirect route may go to all male *siblings* of the current person's *parents*. For these two relations corresponding methods *siblingPaths* and *parentPaths* are again provided, where the first contains the route to all persons that are linked to all common children of both of the current person's parents over a chain of direct *Sibling* relations (e.g. persons p8, p11 and p12 are siblings of person p7 in figure 2). The latter returns the routes to direct parents (relation *Parent*) and to persons defined as parents of any sibling. A third route within method *unclePaths* may lead to all male parents of the current person's cousins (relation *Cousin*), and another route may be defined by all male persons that are in relation *SiblingOfParent* to any of the current person's siblings. Figure 2 shows a set of persons that are in relation *uncle* to person *ego*. For reasons of simplicity the relations are represented as arcs here.

Figure 2

In the same way, methods can be provided for each relation that may be defined through multiple paths.

The deduction of missing relations may be done dynamically whenever required within a query. For faster retrieval it may also be performed in advance and the derived relations may be statically inserted into the network. For queries that only test whether a particular person is in a certain relation to another person, the corresponding routes may be ordered according to

the probability that the relations involved are actually defined. As an example, the relations *parent* and *sibling* may be recorded more frequently in the database than the relation *uncle*. When testing a person for being the uncle of another person, one might therefore want to check the indirect route 'parent – sibling' first.

4 Results

4.1 Integration in existing retrieval systems

Our retrieval systems WAREL [14] and ArchiMed [15] use logical queries both for *cohort formation* and for the *selection and linkage of variables* pertaining to the members of a patient cohort (formation of a statistical matrix). These systems offer a high functionality, as in documented text searches. Normally these queries are confined to intra-patient data. 'AND in the same patient' is a condition implicit in all operators used to select and link variables and intermediate results.

Queries are rendered much more powerful by including family relations into the system, in conjunction with temporal relations where available. This is achieved by expanding the functionality of operators and the structure of the results matrix:

1. Queries: By removing the implicit intra-patient condition and adding family relations to the operators, relations can also be created between individuals, enabling queries such as 'All patients suffering from pneumonia, whose mother shows the risk factor smoker'.

2. Results structure: The LINKAGE format [16] is a consensus structure for genetic studies. It has an interlinked design with the attributes Pedigree ID, Individual's ID, father, mother, sex, generic data (1-n). Based on this structure, the result tables of the procedures *cohort formation* and *variables selection* were expanded by two attributes, one containing the patient ID to which a specific link exists, the other one containing the type of relation. The relation attribute can contain various relation types (e.g. biological, social) that are maintained in a system table. The results include additional attributes not listed in Table 1 (e.g. age, document ID).

4.2 Example

4.2.1 Situation

Figure 3 represents a family of seven with two households (dotted line). One child (ego) does not live with her parents.

Figure 3

4.2.2 Instances

The diagram as such does not indicate which relations are actually documented. As an example, the 'uncle' relation might be referenced directly in ego's own record, or it might be reached through 'father > brother'. If the father relation of ego is not documented, another alternative path might be 'brother > father > brother'. Figure 4 shows an example how the situation of Figure 3 could be instantiated using our model.

Figure 4

4.2.3 Query

Let us assume that we search for 'all children 10 years or under, having asthma, and with one person living in the same household being a heavy smoker'. Asthma is documented by the variable *diagnosis*, smoker by *risk factor nicotine* grades 0 to 3. In the ArchiMed system, the logical query for *variable selection* might comprise the following operands and operators:

'diagnosis = asthma' AND 'Age <= 10 years'
    JOIN 'same household'
'Risk factor nicotine = 3'

4.2.4 Results structure

There should be an opportunity for the user to select the structure of the results matrix depending on the nature of the planned statistical analysis. A row can basically contain variables referring to one or to several persons. If one person is referenced per line, then the observation contains the values of the variables and the connection to the family members (similar to the LINKAGE format). In ArchiMed, variables to be analyzed for several persons can be combined into one line using a flexible JOIN operation. A result generated by 'JOIN family, exclude missing values' would be structured as can be seen from Table 1. Ego is not included in the result set because there is no smoker in her household.

Table 1

5 Discussion

In this paper, we have presented a practical model to formulate inter-patient relations. We have presented a 'community' concept incorporating social relations and intra-hospital contacts in addition to kinship relations. Despite the flexibility of our approach, it is obviously impossible to cover all cultural and social facets of kinship relations. Some cultures may distinguish between 'father's brother' and 'mother's brother', while others use a catchall term like 'uncle'.

The opportunity to include family relations in clinical data retrieval opens the door for clinical research to be based on extended medical histories. Most documentations, however, pursue other objectives. Routine documentation of clinical data is centered on individuals rather than on relations between individuals. Patients are identified by a unique ID, which serves as the main reference point in registering any kind of clinical findings and other pertinent information.

Frequently, the derivability of kinship relations is an inadvertent spin-off that was never envisaged in the original design. In transplantation surgery, for example, kinship relations are

documented because it is evident to use organs of consanguineous relatives for reasons of histocompatibility. Thus the documented donor-recipient relation and other information such as last name or address serve as a basis to derive the fact, and sometimes even the closeness, of kinship.

While hospitals accommodate large data volumes across departments and patients, the scope of most retrieval systems is still confined to individual departments and patients. One of the few applications that support inter-patient data retrieval is Banhard & Klaeren's object-oriented graphical Query Generator [17]. This application enables users to formulate kinship relations as a recursive function.

Obviously, the kinship-related information offered by clinical databases is not up to the standards of family histories collected in the general practitioner's office over extended periods of time, which is precisely where the most immediate need for this type of information arises [18]. The benefits of hospitals and private practitioners exchanging clinical data are mutual [19], as the hospitals can verify and improve the kinship information contained in their clinical databases, while the practitioners have a strong interest in research on family health.

In any country, there exist a number of databases (e.g. national registries) containing information on family relations. While some of these databases would add a lot of valuable information to family relations, it is of course mandatory to comply with data confidentiality requirements. There are high standards attached to the quality of family data. Errors such as false reminders due to outdated entries need to be prevented at all costs. In addition to the technical problems of integration and confidentiality requirements, there frequently exist 'political' obstacles to be overcome before data can be retrieved across hospital departments.

Acknowledgments

References

1. Newcombe H. B., 1993, Distinguishing Individual Linkages of Personal Records from Family Linkages, *Methods of Information in Medicine*, **32,** 358-364.

2. Sachs, P., Gall, W., Marksteiner, A. and Dorda, W., 2000, Unambiguous identification of hospital patients, *International Journal Of Medical Informatics,* **57**(2-3), 165-179.

3. Kamada, T. and Kawai, S., 1990, A general framework for visualizing abstract objects and relations. *ACM Transactions on Graphics*, **10**(1), 1-39.

4. Oracle Corporation, 1998, *Orcle8 Enterprise Edition Documentation Library*, (Redwood City, CA).

5. Mavaddat, F. and Parhami, B., 1979, A data structure for family records, *The Computer Journal*, **22**(2), 110-114.

6. Jones, W. J., 1990, *German Kinship Terms (750-1500)*, (Berlin; Walter de Gruyter),

7. White, H. C., 1963, *An Anatomy of Kinship*, (Prentics Hall).

8. Ott, J., 1991, *Analysis of Human Genetic Linkage*, (London; The Johns Hopkins University Press).

9. Mendel, G., 1970, *Versuche über Pflanzenhybriden*, edited by Weiling, F. (Braunschweig; Vieweg).

10. Dewald, G., 1977, Gregor Johann Mendel and the Beginning of Genetics, *Mayo Clinic Proceedings*, **52**(8), 513-518.

11. Chizzali-Bonfadin, C., Adlassnig, K. P., Koller, W. and Rotter, M. L., 1996, MONI: Monitoring of Nosocomial Infections. *Proceedings of the Sixth National and Fourth International Conference on Information Technology and Community Health, ITCH'96,* Victoria, Canada, pp. 281-285.

12. Parkin, R., 1997, *Kinship - An Introduction to the Basic Concepts*, (Oxford; Blackwell).

13. Booch, G., Rumbaugh, J. and Jacobson, I., 1998, *The Unified Modeling Language User Guide*, (Reading, Mass.; Addison Wesley).

14. Dorda, W., 1989, WAREL: a system for retrieval of clinical data considering the course of diseases. *Methods of Information in Medicine*, **28**, 133-141.

15. Gall, W., Sachs, P., Duftschmid, G. and Dorda W., 1999, A retrieval system for the selection and statistical analysis of clinical data, *Medical Informatics and the Internet in Medicine*, **24**, 201-212.

16. Terwilliger, J. D. and Ott, J., 1994, *Handbook of Human Genetic Linkage*, (London; The Johns Hopkins University Press).

17. Banhart, F. and Klaeren, H., 1995, A graphical query generator for clinical research databases. *Methods of Information in Medicine*, **34,** 328-339.

18. Hogg, W.E., 1992, The family-oriented computerized medical record. *MD Computing*, **9**(6), 343-345.

19. Van der Kam, W. J., Moorman, P. W. and Koppejan-Mulder, M. J., 2000, Effects of electronic communication in general practice, *International Journal of Medical Informatics*, **60**(1), 59-70.
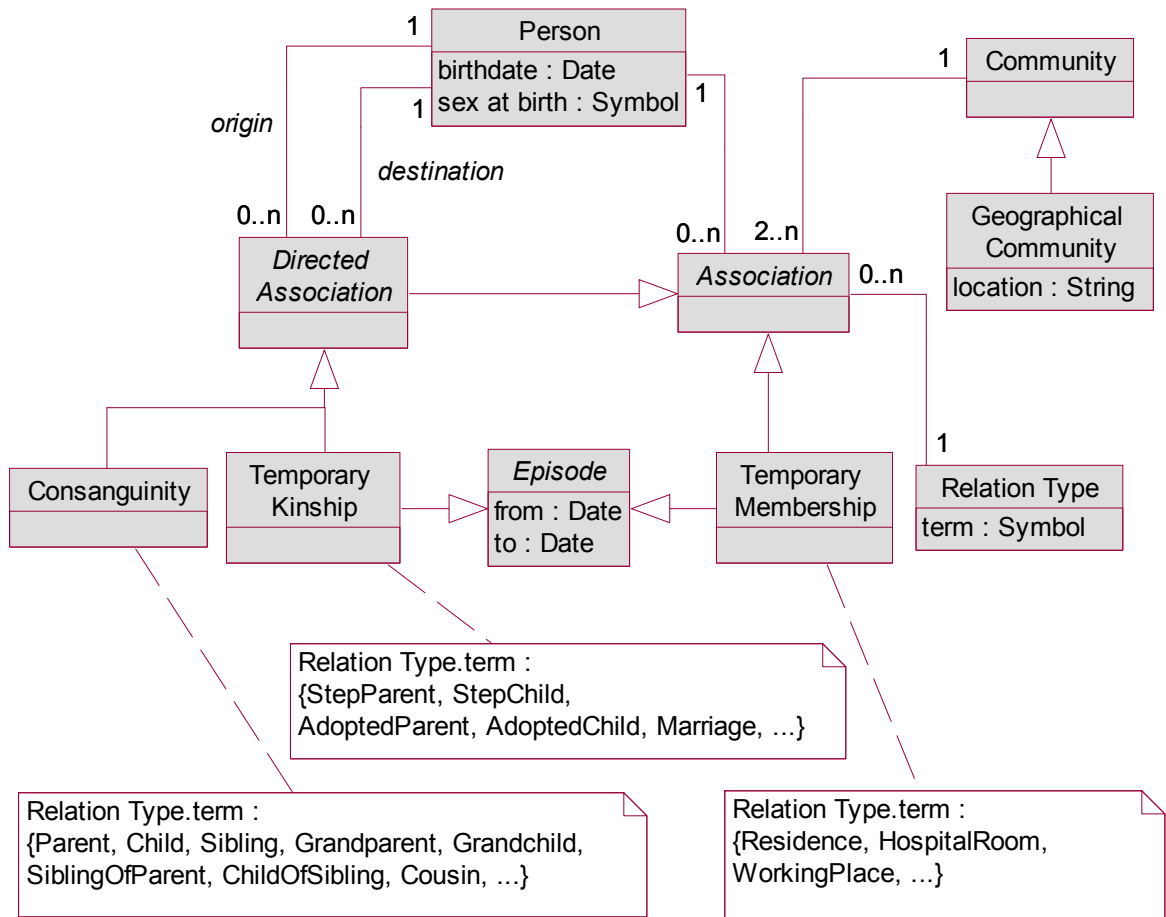
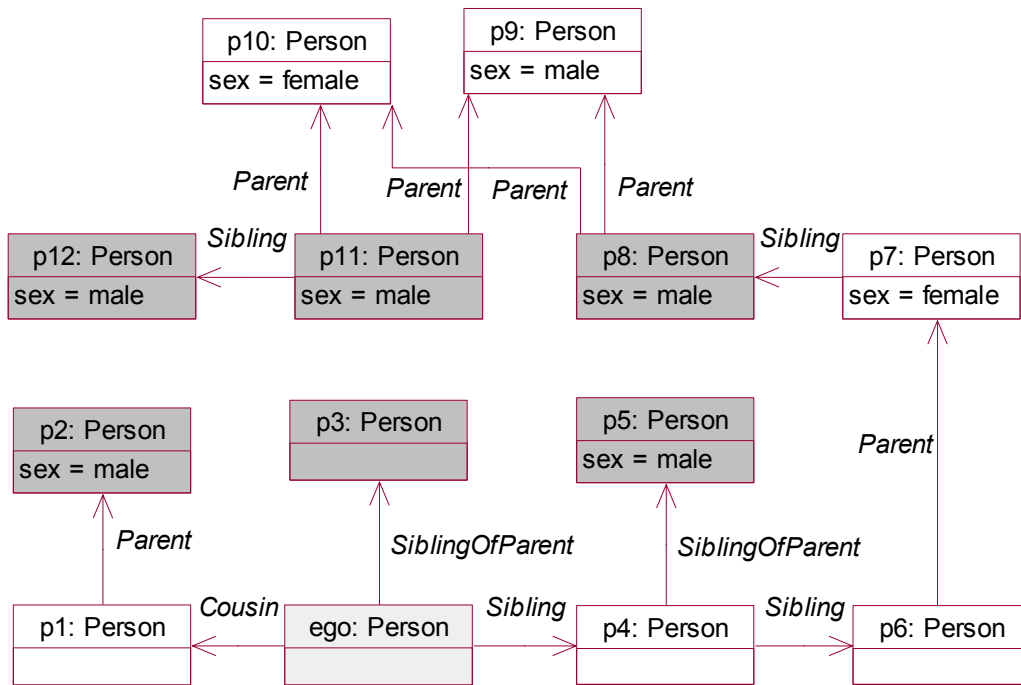Fiure 1.   Class diagram of a model to represent familial and social structures.

Fiure 2. Instance diagram 'Uncles': Grayed instances p2, p3, p5, p8, p11 and p12 are uncles of instance ego.
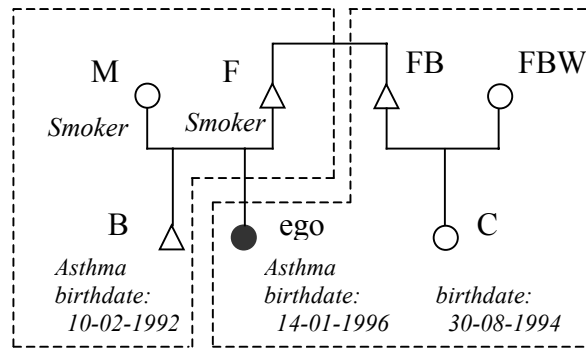
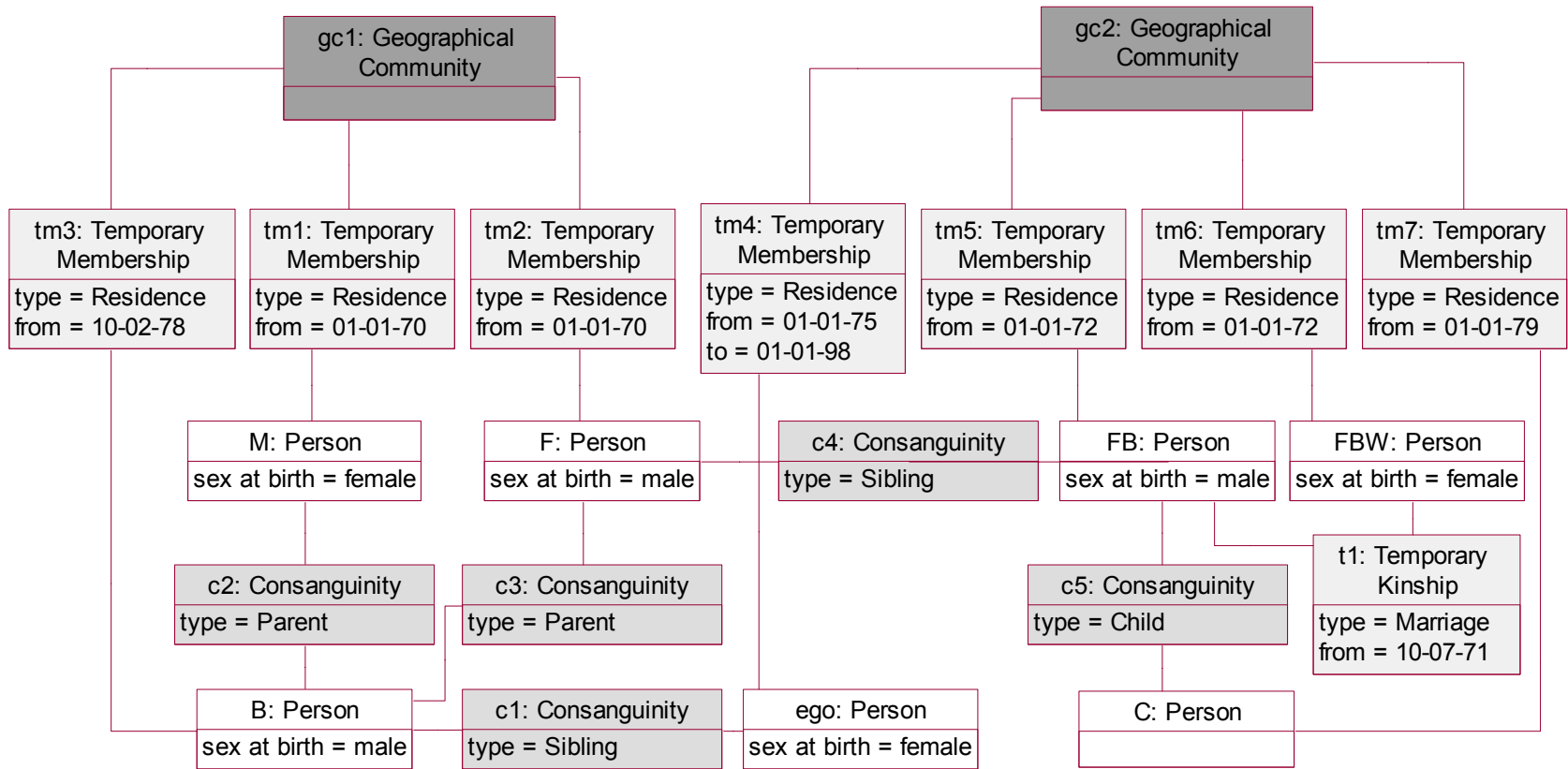Fiure 3.   Two households (dotted lines) with seven persons, ego living with her uncle (modified from [12], page 17).

Figure 4. Instance diagram of possible representation of figure 3.

Table 1. Data matrix after linkage through related person.

Pat_ID	Sex	History	Relation	Pat_ID-R	Sex-R	Nicotine risk
P5	M	Chronic asthma	mother	P1	F	3
P5	M	Chronic asthma	father	P2	M	3