Medical Informatics Europe '99
P. Kokol et al. (Eds.)
IOS Press, 1999

Expanding DIOGENE with a Clinical Information System Based on a New Hospital-Wide Clinical Finding Dictionary

Claudine Bréant, François Borst, Daniel Campi, Vincent Griesser, and Huy Son Le
Division d'Informatique Médicale
University Hospital of Geneva, 1211 Geneva 14, Switzerland

Abstract. The aim of this project is to expand DIOGENE with a centralized and integrated patient clinical database system providing a standardized framework for the building of future clinical databases and for the integration of existing heterogeneous ones. The combined 'across time view' and 'across departments view' generated from the integrated clinical data will enable an evolutionary view of the patient state both across time and across medical specialties. For this purpose and to permit the communication and exchange of data, a new controlled vocabulary for representing clinical data has been created. The construction of this vocabulary is based on the international ICD classification, already being used in DIOGENE for encoding patient diagnosis and procedures. A new extension of the ICD is proposed for medical information that goes beyond diagnosis and procedures encoding. The building of a common clinical finding dictionary recording the definition of findings is based on this newly developed clinical vocabulary. This process is incremental, manual, and significant.

1. Introduction

At the University Hospital of Geneva (HUG), a recent inventory of about 50 clinical databases embedding many different medical fields has been performed. The inability of the DIOGENE [1] hospital information system to provide any framework for the management of clinical data coupled with the advent in the early 80's of personal computers prompted the uncontrolled development of many medical database systems. Indeed, these systems were independently designed and implemented by several medical groups precluding effective data communication and exchange. While fulfilling the immediate needs of each department, many long-term problems linked to the maintenance of such a large and growing heterogeneous environment have become increasingly important.

A four years project has begun in 1997 to expand DIOGENE with a repository of detailed, structured, and standardized clinical data and medical protocols. Detailed clinical data is needed for a complete understanding of quality of care and consumption of resources. Such data describes the precise state of the patient and its evolution beyond diagnosis and therapeutic codes. For instance, it includes items such as the measure of blood pressure, the description of a diabetic retinopathy, and the pulse rate.

The new clinical information system provides a generic and standardized framework to help for the building of future clinical databases and for the integration of existing heterogeneous databases with DIOGENE. This has required the development of a controlled vocabulary representing clinical data. The present work fits into the more general effort for building an integrated electronic patient record system at the HUG.

Motivations to build a centralized clinical information system integrated with DIOGENE can be expressed in two different ways:

a) In terms of electronic patient record functions:

- Reuse and share of the clinical data collected by many different departments mainly to avoid a redundant data capture.
- Automatic link with the DIOGENE hospital information system patient data (admission/discharge data, laboratory, medical records, patient administrative information, image database, etc.).
- Automatic production of textual documents from input data such as letters to referring physicians and patient summary reports.

b) In terms of prospective data processing:

- retrieval of patient data from various points of view:
  - The *'across departments view'*: The state of the patient is described by clinical parameters coming from different medical specialties such as for instance obstetrics, intensive care unit, surgery, etc.
  - The *'across time view'*: This will allow for a evolutionary view of selected patient clinical data through time.

  Both approaches can be combined in order to show for instance the evolution of neonates going from the obstetrics to the pediatrics department.
- Production of federal statistics, clinical studies and medical research queries: the system will enable the mass screening of a large amount of patient data matching certain user-defined criterions.
- Quality of care and consumption of resources: most existing grouping systems, such as Diagnosis-Related Groups (DRG) are based on a few sets of data. An accurate understanding of either the quality of care or the consumption of resources requires more detailed data about the patient. This data goes beyond the 'administrative' medical summary and relies on detailed clinical databases.
- Case-based reasoning or the process of retrieving similar cases [2].

2. Design of the integrated clinical information system

The general functional architecture of the proposed clinical information system is illustrated in figure 1 and characterized by the following features:

- A new common database repository for hospital-wide patient clinical data integrated into DIOGENE.
- A controlled clinical vocabulary built using a bottom-up approach driven by the development of new clinical applications.
- A common hospital-wide clinical finding dictionary defined from the controlled clinical vocabulary.
- A general framework and PC-based graphical tools to design medical questionnaires. The clinical finding dictionary drives the design of a medical questionnaire.

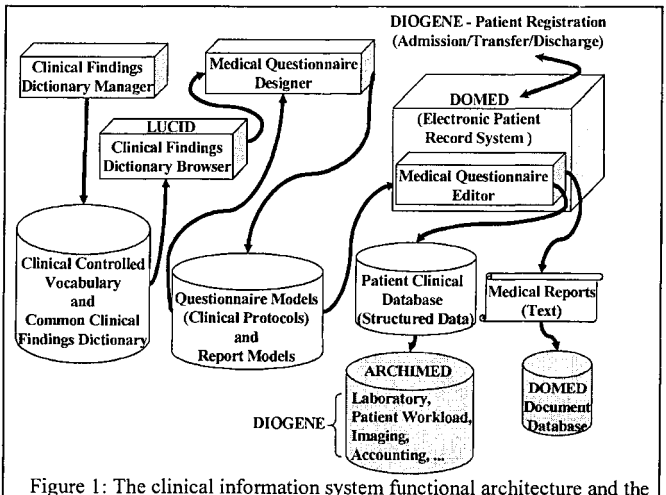

Figure 1: The clinical information system functional architecture and the integration with DIOGENE.

- A centralized and sharable database repository for medical questionnaire models representing medical knowledge about on-going hospital-wide medical protocols and clinical studies. This database includes report models for the automatic generation of textual medical reports, follow-up letters, etc.
- An integration of the questionnaire data entry interface within the framework of the DOMED [3] electronic patient record system currently being implemented at the University Hospital of Geneva.
- Storage and indexation into the DOMED database of automatically generated documents from the clinical database.

The Clinical Findings Dictionary Manager is a tool to facilitate the tasks of adding and/or maintaining the controlled clinical vocabulary as well as of defining new clinical findings.

The Medical Questionnaire Designer is a graphical tool to build questionnaire models detailing a clinical study and/or a medical protocol. The process of establishing a new model is driven by the Common Clinical Findings Dictionary and by the Questionnaire Models database. The Common Clinical Findings Dictionary is accessed through the browser LUCID [4]. The Questionnaire Models database allows previous models of questionnaires to be reused by new applications. Report Models for generating textual documents are built along with a questionnaire model. These allow an automatic production of textual medical documents from data retrieved from the Patient Clinical Database. When a questionnaire model and its report models are published, they can be accessed through DOMED and then through the Medical Questionnaire Editor. An instantiation of a questionnaire is given in figure 2. Signing out a questionnaire triggers the storage of input data into the Patient Clinical Database. Clinical data is subsequently archived into ARCHIMED, the patient information archive gathering data sent or retrieved from DIOGENE remote applications. At the same time, automatically generated medical reports from clinical data are indexed and stored into the DOMED database.

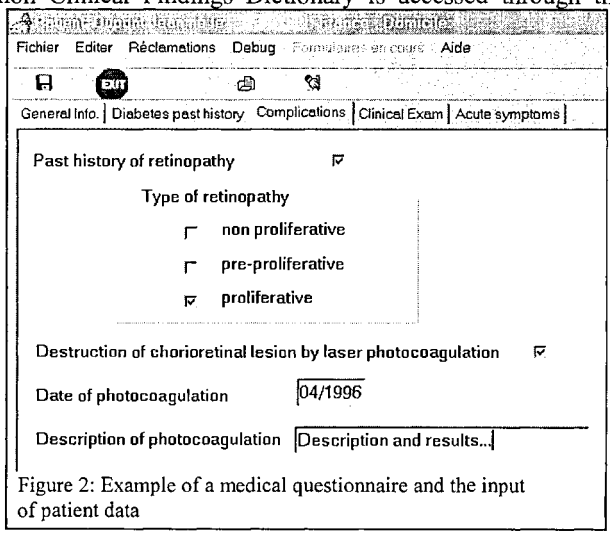

Figure 2: Example of a medical questionnaire and the input of patient data

3. The Hospital-Wide Dictionary for Clinical Findings

Increased interest in decision support, clinical guidelines and protocols requires more detailed, understandable, and sharable clinical vocabulary terms. Much work has been done in order to look for a common and comprehensive clinical controlled vocabulary set. However, the current systems are unable to meet the requirements for clinical data representation [5][6][7]. A general-purpose controlled vocabulary such as the Systematized Nomenclature of Human and Veterinary Medicine (SNOMED International) is multi-axial and cannot be easily used. As for the version 3 of the READ [8] codes classification, it is today limited to a utilization in the UK only. Therefore, it has been necessary to create a centralized controlled vocabulary to support any type of clinical applications developed at the HUG.

The construction of the new clinical controlled vocabulary relies on the ICD classification already being used in DIOGENE to encode patient diagnosis and procedures. More specifically, ICD10 [9] applies for the diagnosis and ICD9CM [10] for the procedures

and they both constitute the Swiss standard. ICD has been completed with two extensions that are local to the HUG: the Diagnosis and Procedures Extension (ICD-DPE) and the Clinical Extension (ICD-CE). The process of inserting a new element in the clinical controlled vocabulary is incremental and is performed each time the current version of the vocabulary fails to fulfill the needs of a medical department.

The ICD-DPE was started in 1979 when the hospital diagnosis and procedures encoding process begun [11]. It is concerned with local extensions for diagnosis and procedures only. These are still periodically added to the classification in order to meet specific needs of the medical staff for a more accurate coding. The codes of these extensions end by the character '-' followed by a three digit number. It represents an additional level in the nomenclature. For instance, the ICD10 diagnostic: *H36.0 diabetic retinopathy* has been extended by three local extensions in order to define the more specific following diagnosis:

H36.0X-001 non proliferative diabetic retinopathy
H36.0X-002 pre-proliferative diabetic retinopathy
H36.0X-003 proliferative diabetic retinopathy

The same process is applied to extend a procedure code. For instance, the ICD9CM item: 68.3 *subtotal abdominal hysterectomy,* has been extended by the two following local extensions:

68.3X-001 open subtotal abdominal hysterectomy
68.3X-002 laparoscopic subtotal hysterectomy

The ICD-CE has newly been created to expand the controlled vocabulary beyond diagnosis and procedures proposed by the existing ICD and ICD-DPE classifications. Therefore, the existing approach already in place for the ICD-DPE is generalized and permits the organization of the vocabulary for clinical items like:

type of diabetes
date of destruction of chorioretinal lesion by laser photocoagulation
measure of the body weight

The codes are created in the ICD-CE with a similar syntax to the codes in the ICD-DPE. They end by the character '=' followed by a three digit number. Codes are included in the corresponding chapter of the ICD structure as a matter of priority. When this is not possible, codes are inserted in the "U" (user-defined) chapter. For instance, the following codes have been created in the ICD-CE structure:

E14.9X=004 type of diabetes
14.24=001 date of destruction of chorioretinal lesion by laser photocoagulation
U50.0X=005 measure of the body weight (Kg)

The existing programs for managing the ICD-DPE have been extended for the purpose of managing the ICD-CE and are part of the Clinical Findings Dictionary Manager toolbox. These tools primarily help for checking the non-redundancy and non-ambiguity when some vocabulary is added to the classification.

The Clinical Findings Dictionary records the definition for all the clinical findings. A clinical finding is composed of several attributes including a link to the controlled vocabulary. Clinical findings in the proposed clinical information system are defined from the vocabulary set including the standard ICD and its two local extensions (the ICD-DPE and the ICD-CE). A clinical finding is fully defined by the following attributes:

- The name of the finding, that needs to be self-explanatory and non-ambiguous.
- The context of the finding such as past history, family history, and complaint. It helps to further qualify a finding to avoid tremendous increasing of the list of codes.
- The type of the value: integer, float, text, date/time, enumeration, or boolean.

An enumeration is a set of codes selected among the ICD, ICD-DPE, or the ICD-CE classifications. For instance, the following set of codes defines the type for the clinical finding *type of retinopathy*:

H36.0X-001 non proliferative diabetic retinopathy
H36.0X-002 pre-proliferative diabetic retinopathy
H36.0X-003 proliferative diabetic retinopathy

- The unit, if the type attribute is numeric.
- The link with the corresponding item in ICD, ICD-DPE, or ICD-CE.

Table 1 describes the clinical findings *weight, past history of retinopathy*, and *type of retinopathy*. Once defined, these items become available to the medical staff through the Medical Questionnaire Designer tool as shown in figure 2.

Table 1: Clinical findings *weight, past history of retinopathy*, and *type of retinopathy* as registered in the common clinical finding dictionary

Clinical finding	*weight*	*Past history of retinopathy*	*type of retinopathy*
Type	*float*	*Boolean (y/n)*	enumeration: *H36.0X-001 non proliferative diabetic retinopathy* *H36.0X-002 pre-proliferative diabetic retinopathy* *H36.0X-003 proliferative diabetic retinopathy*
Unit	*Kg*	*n/a*	*n/a*
Context	*n/a*	*Past history*	*n/a*
ICD link	*U50.0X=005 measure of the body weight*	*H36.0 retinopathy*	*H36.0X=002 type of retinopathy*

4. Results and perspectives

Clinical questionnaires in the field of diabetes have been designed at the HUG using this common clinical finding dictionary. Integrating existing heterogeneous clinic data implies a system integration (at the level of patient admission/transfer/discharge data), a data structure integration (with the medical archive system ARCHIMED), and a vocabulary integration (description of items in a common clinical dictionary). A hospital-wide clinical finding dictionary based on extensions of ICD10 and ICD9CM has been developed. This is essential for sharing information between departmental clinical systems. However, in the absence of internationally recognized clinical vocabulary, the work necessary to create such a clinical dictionary is significant. The main tasks have included the setting of manual procedures and algorithms as well as computerized tools such as the Clinical Findings Dictionary Manager and LUCID. Sophisticated vocabulary maintenance tools have yet to be designed and implemented to automate some of the work in adding new terms in the ICD two extensions. This would help to add vocabulary more rapidly with a higher quality than it would be possible manually. Finally, the development of an internationally recognized vocabulary has yet to be done. ICD was not designed for organizing clinical vocabulary. However, its structure makes it possible to simply determine clinical extension codes for element related to the ICD10 and the ICD9CM chapter structures.

References

[1] Borst F., Lovis C., Thurler G., Maricot P., Rossier P., Revillard C., Scherrer J.R. Happy birthday DIOGENE: a Hospital Information System born 20 years ago. *Medinfo 98*. B. Cesnik et al. (editors).

[2] Jaulent M.C., Le Bozec C., Zapletal E., and degoulet P. Case Based Diagnosis in Histopathology of Breast Tumours. *Medinfo 98*. B. Cesnik et al. (editors).

[3] Baud R., Lovis C., Scherrer J.R., Nouvelles perspectives en matière de dossiers patients en réseau. *Journées Francophones d'Informatique Médicale à Lièges, Belgique*, Avril 98.

[4] Michel PA, Lovis C, Baud R. LUCID : a semi-automated ICD-9 encoding system. *Medinfo* 1995.

[5] Hammond WE. Call for a Standard Clinical Vocabulary. *J Am Med Inform Assoc.* 1997 ;4 :254-255.

[6] Campbell JR, Carpenter P, Sneidman C et al. Phase II Evaluation of Clinical coding schemes : completeness, taxonomy, mapping, definitions, and clarity. *J Am Med Inform Assoc.* 1997 ;4 :238-251.

[7] Cimino J.J., Johnson S.B., Hripcsak G., Hill C.L., and Clayton P.D. Managing Vocabulary for a Centralized Clinical System. *Medinfo 95 Proceedings*. R.A. Greenes et al. (editors).

[8] Buckland R. The Language of Health. *British Medical Journal* (1993) 306:287-8.

[9] International Statistical Classification of Diseases and Related Health Problems. *World Health organization*, Geneva, 1992; volume 1 and 3.

[10] International Classification of Diseases Nineth Revision. *MedIndex Publications*, Salt Lake City, USA, 1994.

[11] Borst F., Griesser V., RossierP., Bourdillou R., and Scherrer JR. Fifteen Years of Medical Encoding in the Diogene HIS. *Medinfo (Canada)*, 1995.