Data Representation for EMG Case Collection and Management*

Mária Bieliková, Pavol Návrat, and Mária Smolárová

Slovak University of Technology
Department of Computer Science and Engineering
Ilkovičova 3, 812 19 Bratislava, Slovakia
bielik@elf.stuba.sk, navrat@elf.stuba.sk,
smolarova@dcs.elf.stuba.sk

Abstract. In the paper we discuss different data representation formats used for EMG case collection and management. Data representation is described in the context of developed software tool called EMMA. EMMA is primarily based on relational data model. We describe the relational data model, which was designed according to binary ECCO format used for EMG cases exchange to date. Moreover, eXtensible Markup Language (XML) is used to exchange data between physicians, who use EMMA, or between applications working with EMG data. Interface of EMMA is designed with special care to fulfil requirements of a user – physician. Using this tool will ease not only storing EMG cases, but more importantly reviewing the cases during consensus exercises and making decisions about a diagnose.

1 Introduction

Electromyography (EMG) is a standard method for monitoring muscular activity at the level of bioelectrical signals. A physician has to deal with large amount of data. Computer based information processing supporting diagnostic process in neurophysiology may enhance the physician's ability to make appropriate decisions and find the right diagnosis. A software tool for collecting and managing EMG cases is useful because of its capability of:

- storing EMG cases, which enables retrieving previous examinations and comparing or evaluating them;
- exchanging and distributing EMG cases, which serves as a communication between physicians (during consensus exercises);
- tracking condition of a patient by effective retrieving of EMG examinations for a particular patient at different times.

* The work reported here was partially supported by project INCO Copernicus, No. 977069, European Research Network for Intelligent Support of Electromyographic Studies: EMG-Net.

During the ESTEEM project (European Standardised Telemetric tool to Evaluate knowledge-based EMG systems and Methods), the CASETOOL for EMG data collection was developed together with Communication Protocol EPC/ ECCO 3.2 for exchange of EMG data [5, 7]. The CASETOOL is written in Turbo Pascal. It uses binary ECCO format for data storage. ECCO format has been designed as an efficient format from the point of view of storage. The problem with ECCO format lies mainly in its inflexibility with respect to extensions, which can arise during the EMG case collection and evaluation (e.g., extensions specified in [6]). Binary format is also not very suitable for exchange using the Internet for obvious reasons.

In order to solve the above mentioned problems we designed a new relational format as a part of the EMG-net project [3]. Design of this format is based on the ECCO format, which defines all the necessary data and data types stored in a database. The relational format is advantageous because of its flexibility and extensibility. Created database can be easily used by different tools through standard interfaces to the relational database. What we should pay for this flexibility is less effective storing than the one that is achieved by binary format (such as ECCO).

In the paper we concentrate on the design of a database format for representation of EMG cases and a XML format for exchange of EMG data. The relational database format conforms with the requirements for an advanced software tool. Its main advantage is flexibility, possibility to create different views of data and its maintainability. The relational database is supported by existing database management systems on different platforms and can be shared between different applications. This increases the portability of the final solution.

We also report on the state of development of an advanced tool for collecting and managing EMG cases built on the top of designed relational EMG data model. Our aim was to develop a software tool for EMG case collection and management. This software tool could serve as a basis for exchange of expertise among a community of neurophysiologists.

The outline of the rest of the paper is as follows. In Section 2, we summarise concepts of the EMG case collection and management. In Section 3, data representation used by EMMA is described. Next, we concentrate on brief description of EMMA design and implementation, i.e. the software system architecture and user interface concepts. We conclude the paper with our conclusions and give some proposals for the future work.

2 Basic Concepts of Computer Supported EMG Case Collection and Management

EMG data can be collected at several levels: the local level, the national level, and the European level.

The local level is represented by a particular EMG workplace. The national and the European level is intended mainly for exchange of the EMG expertise accumulated during the years of practice. This would improve the quality of

an early diagnosis and prevention of neuromuscular diseases in each particular country, or a whole region. The national level of EMG data collection becomes also important when the aim of EMG data collection is development of normative EMG data [1].

Figure 1 illustrates two sides of EMG data collection. Local side (left part of the Figure 1) corresponds to the computer support of local EMG data collection through a combination of an automated support of data acquisition from the EMG machine and a manual inputent (or modification) of EMG data by a physician. Manual input is inevitable at least in case of inferred data (physician's conclusions). Proposing conclusions at different level of diagnosis can be supported by a decision support tool.

The right part of the Figure 1 depicts a computer support of EMG data distribution through global computer network (e.g., Internet). EMG data are stored in a global database, which should be synchronised with local databases in order to contain correct data from all EMG laboratories. In fact, several global databases can exist: global national databases and global regional database. This hierarchy allows the use of a software tool on all three mentioned levels: local, national and European. Separating particular levels enables the use of the EMG case collection software tool in off-line regime, i.e. physician is allowed to enter data without the Internet connection.

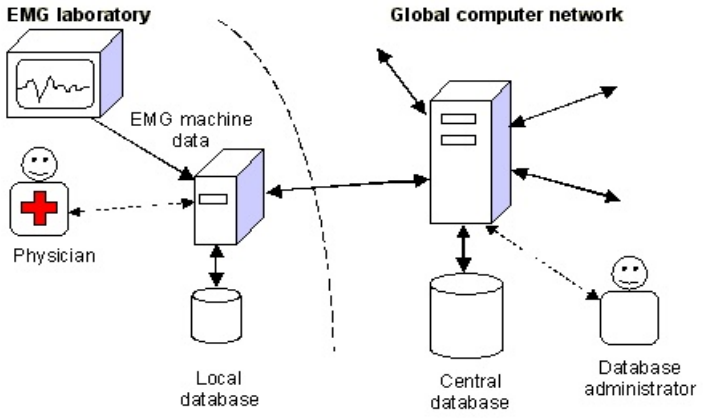

Fig. 1. Distributed EMG data collection.

EMG data collection should follow a diagnostic hierarchy introduced by Fuglsang-Frederiksen et al. [4]. The hierarchy comprises the following types of data (from the bottom of the hierarchy): symbolic parameter values, pathophysiological test conclusions, pathophysiological structure conclusions, EMG-diagnoses, clinical diagnoses (supported also by patient and clinical information).

Most patients are repeatedly examined. In order to be able to monitor patient condition, a software tool for EMG data collection should keep relationship

between different examinations of the same patient and store basic patient information in one place.

3 Data representation

In order to preserve continuity in the EMG-net project as well as to support future developments, we use several data representation formats, which range from

- binary format (Communication Protocol EPC/ECCO 3.2) through
- relational format (represented by a relational database) to
- text format (represented in the eXtensible Markup Language).

The proposed new case collection and management tool should support all the three formats. The ECCO binary format is designed as an efficient format from the point of view of storage (developed during ESTEEM project). Several existing tools use it (e.g., KANDID [4]). Moreover, EMG laboratories, which form EMG-net consortium [3] have collected in the last years together more than 1000 EMG cases stored in this format. There exists so called *gold EMG data collection*, which stores more than 200 cases resulted from the consensus of partners. ECCO is thus retained in the new case tool in the form of export/import capabilities.

The relational format is advantageous because of its flexibility, extensibility and portability. It is used for storing EMG cases. Relational format presents also a base for data mining. Finally, XML format is suitable for distribution of EMG cases between partners (users of the system). EMG cases exchange is important in the process of consensus exercises, where a group of physicians discusses particular EMG cases in order to reach consensus regarding an EMG diagnose as well as a clinical diagnose.

Detailed description of ECCO format can be found in [5]. In the following sections we provide information about proposed relational model for EMG data storage and the use of XML in order to support EMG data exchange.

3.1 EMG Data Relational Model

We designed a relational data model as a result of the analytical model of the ESTEEM Communication Protocol (ECCO version 3.2). The proposed model comprises all the attributes with their corresponding values from the ECCO-format, which are necessary for EMG examination. Moreover, we proposed additions, which are inevitable for efficient data management and extensions proposed in User-5 specification [6]). In order to maintain readability and compatibility with the ECCO format, the names of entities and attributes in the proposed relational model correspond to the ECCO-format names (whenever it is possible). The relational data model is further capable to store normative EMG data.

The basic part of our model is the interconnection between *Examination*, *Structure* and *Test* entities. It is based on the current situation in the EMG

examination process, where a single EMG-examination includes examination of several structures and each structure can be tested by several tests.

The central entity of our model is *Examination* entity. It contains general informations acquired from one examination. This entity originates from "PatientInfo...NonEMGInfo" entity from the ECCO-format, which evolved into two entities: *Examination* and *Patient*.

The entity *Patient* contains only those attributes, which are permanent for all examinations of a patient. At the present it contains only attributes obtained from the ECCO-format and two new attributes - *Status* and *Consensus* (both for marking purposes in EMMA software tool). However, we assume some other attributes will evolve. We also expect the EMG-examiners would like to add some attributes for their internal purposes (e.g., address, comment, phone number). The relation between Patient and Examination is 1:N, so each patient can be examined more than once.

There exist many entities with $Type$ attribute, which affect contents and meaning of attributes in their corresponding entities. This approach is used to simulate generalisation relationship between particular entities.

A lot of attributes in the proposed data model have symbolic values, which are stored as numeric values. The values are stored in $CEnum$ entity, which has three attributes: $Type$, $Value$ and $Text$. The two latter mentioned represent numeric value or text meaning of the attribute identified by $Type$ respectively.

We identified some parts of the data model, which were designed with the likely future changes of requirements in mind. The identified entities contain their symbolic values in separate tables. Each of such tables contains $CodeTableVer$ attribute, which determines version of the given code table (at the present the values are equivalent with ECCO, v. 3.2 values). This allows possible changes and support of different symbolic value assignment versions by a case collection tool without change of the software tool. As an example we can mention flexible representation of clinical diagnosis names and their numeric representations; EMG conclusion names and their numeric representations, or anatomical structures and their codes. Described approach complicates the data model, so we had to carefully decide between simple data model and its flexibility.

A complete report of the proposed relational model is available at the Slovak EMG-net local Web page: `http://www.dcs.elf.stuba.sk/emg/`.

3.2 XML and EMG data

XML is a metamarkup language that allows authors to define their own sets of markup elements that are most appropriate to the specific problem they are dealing with. Structure of the XML based document type can be described by the document grammars (DTD, Document Type Definition) that provide rules on the use of tags, their inclusion and attributes.

The proposed markup language for EMG data exchange (EMG-ML) is supposed to create XML data-centric documents [2]. EMG-ML is based on the proposed relational model. EMG-ML document contains information about all the

patient's examinations. Translation of relational database content to the EMG-ML document is fairly starightforward. Data are represented as tag attributes or tag values depending on data character.

Bellow is depicted structure of EMG-ML document (part of DTD):

```
<?xml version="1.0" encoding="windows-1250"?>
<!DOCTYPE emd [
<!ELEMENT EMD (PATIENT, EXAMINATION+)>
    <!ELEMENT PATIENT ((NAME| PATIENTID), BIRTHDATE,
                                          RACE, SEX)>
    ....
    <!ELEMENT EXAMINATION (HEAD, NONEMG, EMGDIAG?,
                        FINALDIAG?, STRUCTURES, TESTS)>
    ....   ]>
```

Using XML files as the exchange format has several advantages:

- the content of the document is dynamic, depending on the actual anatomical structures that are examined and used examination techniques,
- the content of the document is self-describing because of the text-based nature of XML files,
- the EMG data transmission is highly error-free because its non-binary character, and
- the transmitted data are "open" for third party software processing.

Main disadvantage of such data format is the length of the document and inefficient manipulation of the data. However, its combination with a relational database solves these problems.

4 Design and Implementation of EMG Case Collection Tool

A software tool for case collection and management is being developed in Visual C language in the Windows operating system [8]. The user interface comprises several windows and a set of forms for filling and maintenance EMG case data. Relational database is built in Microsoft Access.

The software tool has been developed in a process that involved prototyping. After analysis and preliminary design we developed a prototype system called KATE. This prototype served as a preliminary specification of the case collection and management tool and facilitated generating a feedback from our medical partners (this prototype was presented at the EMG meeting in Grenoble in February 2000). KATE functionality was restricted to the import of EMG data from ECCO files and manipulation with them. The user could also enter data. However, the relational model was not completed at that time, so KATE did not offer *save* functionality. In fact, existence of a database was not necessary for validation of the user requirements.

A feedback from the presentation of the KATE tool served as a basis for a detailed design of a new case collection tool. The KATE tool was completely replaced by the EMMA tool, which uses a relational model as a basis for storing data and information about EMG cases (import from ECCO was reused from KATE).

The functionality of EMMA is influenced by the process of EMG examination. During the examination a physician works with three groups of data [6]:

1. *General data:* patient information, examination information, clinical information.
2. *Examination data:* test conditions, parameters.
3. *Inferred data:* symbolic parameter values, pathophysiological test conclusions, pathophysiological structure conclusions, EMG diagnoses, clinical diagnoses.

The software tool should store and enable manipulation of all the above mentioned data. Moreover, we also incorporated EMG normative data into the design, which serve for accurate examination interpretation. EMMA supports inputting and modification of normative data in all the required formats (e.g., in table format, functional format) and calculating symbolic parameter values according these data in an actual test. Collection of normative data is a complex process, which requires EMG examination of many healthy persons. Advanced information technology can help in this process significantly [1].

Figure 2 depicts basic modules of EMMA tool. This tool fulfils the role of local EMG data collection and manipulation.

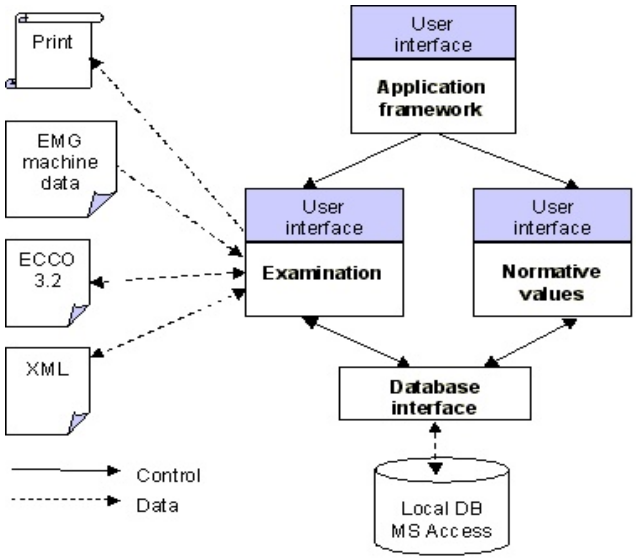

Fig. 2. Architecture of EMMA software tool.

The software tool is designed in such way that a patient is considered a central entity in the system. A physician can work with patient data (general and clinical information) and create several EMG examinations for the patient. This approach enables observing changes in subsequent examination of a particular patient (see Figure 3).

The system allows the user to work with several EMG cases (related to the same patient or to different patients) at the time. Forms for inferred data are logically grouped, which will ease the decision process. A physician can see on one screen both Structure and Structure conclusions, EMG diagnoses and Structure conclusions, and Clinical diagnoses together with already stated EMG diagnoses.

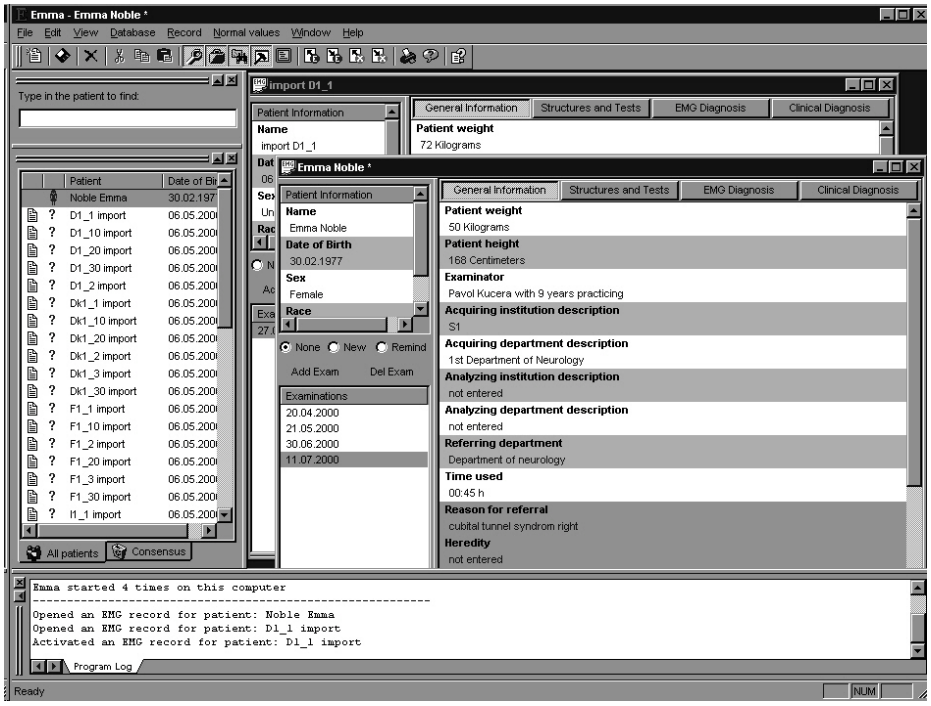

Fig. 3. Example of EMMA interface.

5 Conclusions

The work presented in this paper is aimed towards EMG standardisation by design and realisation of EMG data collection tool. We concentrated in our work on data representation, which is crucial task in such data intensive area as is the EMG examination. The proposed data model is implemented in the MS Access database. In order to use this data model for data mining, EMG

cases created by physicians in the ESTEEM and EMG-net projects should be imported into the database.

The designed data representation allows future extensions of EMG data management features:

- support of tests and techniques selection during EMG examination;
- support of conclusion statements determination (on all levels, i.e. structure conclusion, EMG diagnose conclusion and clinical diagnose conclusion);
- support of retrieving cases based on their similarity;
- support of monitoring patient's course of illness.

The listed features can be provided by the designed relational model. For communication purposes between different agents performing the mentioned activities, the proposed EMG Markup Language could be used. The problem with current state of data representation is still a large amount of non-structured information, which is entered into the database as free text. The structured clinical information from MYOSYS [9] can be used to enhance possibility of data mining over created databases.

The new case collection and management tool will improve management of EMG cases. The relational format is advantageous mainly because its flexibility and extensibility.

References

1. Bieliková, M., Návrat, P., Smolárová, M., Kučera, P., Záhon, P. and Traubner, P.: Computer supported collection of normative data for surface electromyographic studies. In EMG Workshop 2000. Lučenec, May 9-10, 2000.
2. Bourret, R.: XML and Databases. Technical Report. Technical University of Darmstadt, September 1999.
3. EMG-Net: European Research Network for Intelligent Support of Electromyographic Studies, INCO project No. 977069.
   $http://www.inrialpes.fr/sherpa/emgnet/emg_index1.html$.
4. Fuglsang-Frederiksen, A., Ronager, J. and Vingtoft, S.: PC-KANDID: an expert system for electromyography. Artificial Intelligence Medicine, 1989, 1: 117-124.
5. Jakobsen, L.S., Fog, B. and Talbot, A. ESTEEM communication protocol, ver. 3.2. ESTEEM report Advanced Informatics in Medicine (A2010), 1993,1-87.
6. Johnsen, B. and Fuglsang-Frederiksen, A. USER-5: EMG dataset specification version 3. ESTEEM report Advanced Informatics in Medicine (A2010), 1994,1-44.
7. Jonhsen, B., Vingtoft, S., Fuglsang-Frederiksen, A., Barahona, P., Fawcett, P., Jakobsen, L., Liguori, R., Nix, W., Otte, G., Proenca, J., Schofield, I., Sieben, G., Veloso, M. and Vila, A.: A common structure for the representation of data and diagnostic processes within clinical neurophysiology, In: Proc. of the 12th Int. Congress of the European Federation for Med. Inf., EFMI, Lisboa, 1994, 150-155.
8. Makula, M., Milička, T., Noris, I., Vasilovčík, B. and Vlasko, K. Support of EMG diagnostics. Technical report, Department of Computer Science and Engineering, Slovak University of Technology in Bratislava, 2000.
   $http://www2.dcs.elf.stuba.sk/TeamProject/1999/team05/$
9. Ziebelin, D., Vila, A., Rialle, V.: Neuromyosys - a diagnosis knowledge based system for EMG. In: Proc. of 12th Int. Congress of MIE, Lisboa, Portugal, 1994.