DIGITAL ONCOLOGY PATIENT RECORD - HETEROGENEOUS FILE BASED APPROACH

Nikolay Sapundzhiev
Department of Otorhinolaryngology,
Medical University "Prof. P Stoyanov", Varna, Bulgaria

SUMMARY:

Introduction: Oncology patients need extensive follow-up and meticulous documentation. The aim of this study was to introduce a simple, platform independent file based system for documentation of diagnostic and therapeutic procedures in oncology patients and test its function.

Material and methods: A file-name based system of the type M1<separator>M2<separator>M3.F2 was introduced, where M1 is a unique identifier for the patient, M2 is the date of the clinical intervention/event, M3 is an identifier for the author of the medical record and F2 is the specific software generated file-name extension.

Results: This system is in use at 5 institutions, where a total of 11 persons on 14 different workstations inputted 16591 entries (files) for 2370. The merge process was tested on 2 operating systems - when copied together all files sort up as expected by patient, and for each patient in a chronological order, providing a digital cumulative patient record, which contains heterogeneous file formats.

Conclusion: The file based approach for storing heterogeneous digital patient related information is an reliable system, which can handle open-source, proprietary, general and custom file formats and seems to be easily scalable. Further development of software for automatic checks of the integrity and searching and indexing of the files is expected to produce a more user-friendly environment.

Keywords: Head and neck cancer, cancer informatics infrastructure, cancer registries, information storage and retrieval, data warehouse
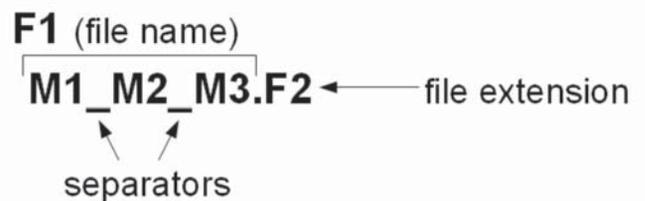
INTRODUCTION

Oncology patients need extensive follow-up. From diagnosis till the end of follow-up/life they will have a very long period of medical observations and will meet many health-care professionals at different health care facilities. Respectively the medical documentation piles up, but is often scattered throughout different healthcare centers and only selected parts of it get to the next treatment level [4]. The importance of the medical record is indisputable for the individual patient - treatment/follow-up are based on it [3]. The aim of this study was to introduce a simple, platform independent file based system for documentation of diagnostic and therapeutic procedures for the long-term follow up of oncology patients and test its function at several medical centers.

MATERIAL AND METHODS

The digital patient record system was initially introduced in 2004 for the purpose of follow-up of patients with upper airways stenosis as a part of a prospective clinical observation, carried by the author. Further improvements were introduced in cooperation with Bapha Consult Ltd. (Varna, Bulgaria) and tests were carried out on different hardware and software platforms (MS-Windows and Linux). In this way a file-name based system of the type M1<separator>M2<separator>M3.F2 was introduced. M1 is a unique identifier for the patient. The Unified Citizen Number (so called EGN) of all Bulgarian citizens serves perfectly the role of M1. M2 is the date of the clinical finding/ intervention/ event. The M2 is presented in the YYYYMMDD HHMMSS** format. The asterisk (*) is for optional numerical positions. M3 is an identifier for the author of the medical record. F2 is the specific software generated file-name extension. M4 is the content of the file - the medical information itself and is not a part of the file-name system. A schema of the file name is presented on Figure 1.


RESULTS

Introduced in 2004 by the author the system was upgraded and improved in cooperation with Bapha Consult Ltd. (Varna, Bulgaria). It was tested on several different hardware platforms with two operating systems (MicrosoftWindows and Linux). This test version proved its usefulness and since 2006 is used for prospective follow-up and documentation of oncology patients at the Department of otorhinolaryngology, head and neck surgery at the "Sveta Marina" university hospital in Varna. Till 2009 it is in use at 5

institutions. Records were inputted by 11 persons on a total of 14 different workstations. A total number of 16591 entries (files) for 2370 patients were entered. The merge process was tested on 2 operating systems - MS-Windows and Linux - and worked in a similar way. When copied together on a single storage place all files lined up as expected by patient, and for each patient in a chronological order. Files form different workstations, but for the same patient automatically were listed together providing a cumulative patient chart, which contains heterogeneous file formats, but listed consecutively. Example listing of heterogeneous patient records is presented on Figure 2 (all identifiers have been changed for privacy reasons).
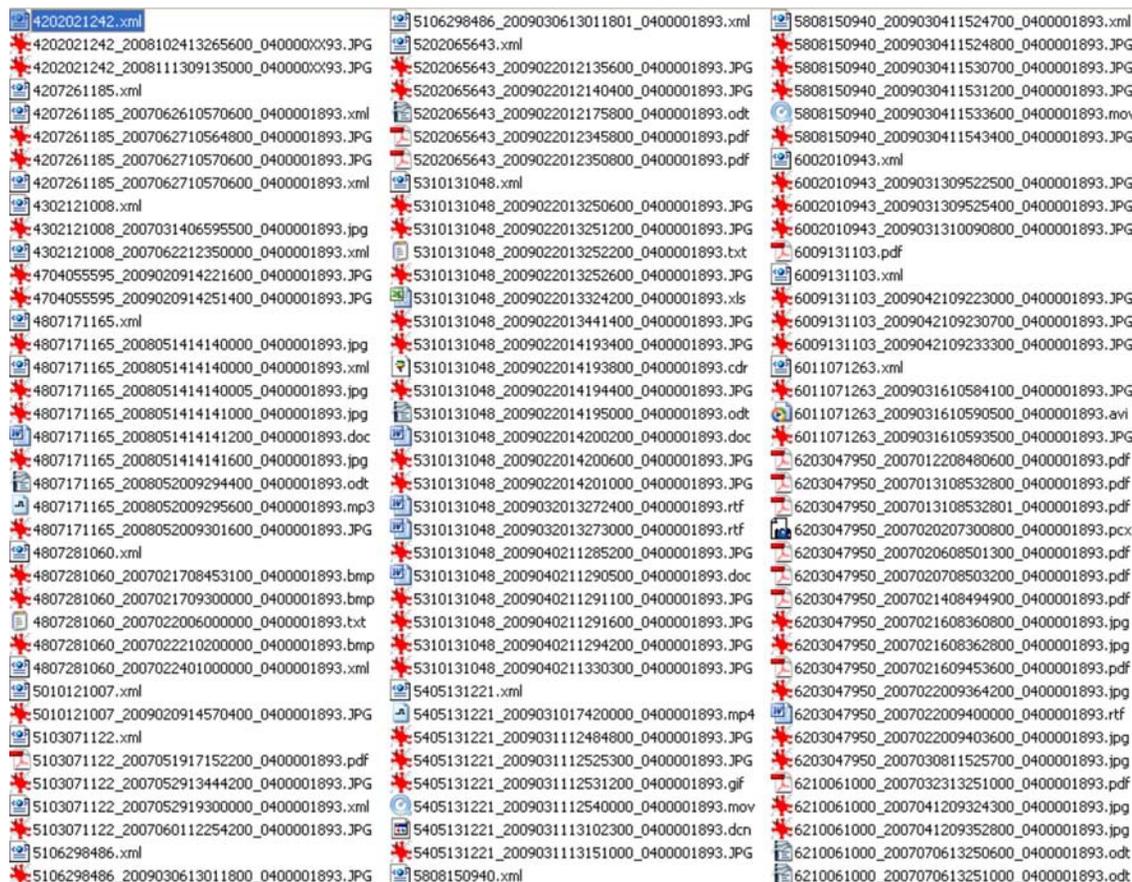
DISCUSSION

The medical documentation of the cancer patients is important for decision-making and follow-up by the health care professionals [3]. On the other hand it is a source of statistical and scientific information [4]. In the system presented the emphasis is on the patient himself. The aim is to have as detailed information as possible, stored in an uniformed manner, which can be easily retrieved and viewed and could be updated by newer records independently at various sites.

Some databases are based on a unique identifier of the tumor - the rationale is, that every different malignancy in the same patient is a different disease [7]. Our approach still has the ability to document multiple tumors and events for the same patient. The use of EGN as a main identifier will allow merging heterogeneous data from different medical facilities. Studies in a particular institution have a natural tendency to present only limited data on particular time period [3,4]. False cures or failures could be calculated if the patients will be followed in another region of the country in other medical facilities [3,8]. Even countries like Germany do not have a structured national oncology register, but rely on the local registers of the provinces [1]. A change in the settlement of a patient will result in a "new oncology case" and "new record". This will perturb again the overall statistics. Tumor classifications change, even ICD changes [4,5].

Large nation-wide oncology databases with specific architecture are hard to be linked to provide current information for the patient and be used as a source of medical knowledge for analysis [6]. Such large databases need time to gather and summarize the result, the reports often follow on an yearly basis and give a good demographic overview, but are of little significance for the

individual patient [2,4]. The linking of a particular small database of a prospective study for a given parameter to a large pre-structured general oncology data base will end up in adding tens of columns, which will but contain info only for the study patients [7]. The other approach is the use of relative databases, where the small sub-database will be external to the general database, but only linked to it via the identifiers of the patients. So again we need an unique identifier for the patient, which will be reliable. Some large centres have developed internal patient record systems with unique identifiers, which but can not be linked on a nation-wide level or be accessible for the physicians in the follow-up phase [3,4]. With the file system presented each medical result/intervention for one patient is a separate piece of data. There is no predefined database structure. For particular tests/ study protocols a separate report form could be created, which will easily and naturally integrate into the patient record.

When asking some particular medical/scientific questions the medical society initiates studies. The prospective studies are highly valued because of their precise design, aimed at the specific topic. The databases, which are created for such studies are quite limiting, focused only on the particular parameters and end up on the hard-drive of some computer once the study is carried out, published and the working group separated. Such specific databases are hard to be integrated in an general oncology record system [3,7]. Being to particular they are of little value for the routine follow-up of patients. In this way they represent a precisely answered medical question, but a lost piece of information for the oncologic community and inaccessible protocol for clinicians/researchers, who my be interested in the procedure itself.

The heterogeneous file-based system proposed is based on 2 practical principles. The first one is the medical one (M): for every patient record the most important parameters are: M1 - who is the patient; M2 - chronology of the event; M3 - ho performed the event and M4 - nature of the event. These are the classical attributes to every patient record, no matter the setting, or the form of storage of the information. Medical information makes no sense if it is not patient related (M1). The second important issue is the patient history - when happened the event (M2) and what exactly happened (M4). Examinations, procedures and documentation are responsibilities of a certain person - the health-care professional or (M3). When talking of documents it appears more important to identify the author of the document. This could be any healthcare professional of documentation officer, while the physician, who had performed the intervention/ diagnostic could be another person or a group of persons. Their names ought to be included in the medical report itself.

The second principle is that nowadays everything could be digitized. No matter what kind of information is digitally stored it is in the form of files (F). And every file is identified by its name (F1) and extension or file type (F2), which determine its uniqueness. M1 together with M2 and optionally M3 build the file name F1. F2 is technically determined and is specific of the file type (text, image, video etc.). This gives us an unique file name, which will represent an unique action on an unique patient. The content of the file is actually the particular medical information M3. Optionally the file could contain control entries about the M1, M2 and M3 if supported by file format.

In this way the file-name system of the type M1<separator>M2<separator>M3.F2. was introduced. The identification of the patient is done by his Unified Citizen Number (EGN). It serves perfectly the role of M1. The test carried so far revealed several problems with this approach. Foreign citizens need another ID number to serve as M1. Another limitation is the fact, that according to the Low for personal information the EGN is a piece of personal information and this raises legislative issues, which are practically irrelevant to the hospitals.

The M2 is presented in the YYYYMMDD HHMMSS** format. The asterisk (*) is for optional numerical positions. This date format appears not to be user friendly. The typical date format for Bulgaria is DD.MM.YYYY HH:MM:SS. It will but complicate the computer algorithm. The files will not sort in the most "natural" to the computers way in a chronological way if the data format starting with the day is used. The YYYYMMDDHHMMSS** format is closer to the date format coded in the EGN and will complementary allow for easier calculation of the patient age for example.

CONCLUSION

The file based approach for storing heterogeneous digital patient related information is an advantageous system, which can handle open-source, proprietary, general and custom file formats. All files organize spontaneously based on the unique identifier of the patient in the file-name. The natural sorting order of any OS will order the files chronologically, thus creating an clear medical record, close to the standard paper form. No complicated central database is needed to handle and index the information in predefined structures. Stored on a central server with read only access, the files should be further protected as any other crucial digital information. Automatic robots and queries could perform checks of the integrity of the identifiers in the file name and the meta-data, contained in the file itself. Queries could perform automatic large scale retrospective data retrieval. For prospective studies particular forms in widely accepted file formats could be created, which will again align with all other patient records chronologically.

REFERENCES:

1. Ambrosch P. Screening untersuchungen zur Frьherkennung von Karzinomen der oberen Luft- und Speisewege. HNO 1996;44:609-611

2. Bhattacharyya N, Fried MP. Benchmarks for mortality, morbidity, and length of stay for head and neck surgical procedures. Arch Otolaryngol Head Neck Surg 2001;127:127-132

3. Coatesworth AP, Tsikoudas A, MacLennan K. The cause of death in patients with head and neck squamous cell carcinoma. J Laryngol Otol 2002;116:269-271

4. Jones AS, Beasley N, Houghton D, Husband DJ. The effects of age on survival and other parameters in squamous cell carcinoma of the oral cavity, pharynx and larynx. Clin Otolaryngol Allied Sci 1998;23:51-56

5. La Vecchia C, Lucchini F, Negri E, Levi F. Trends in oral cancer mortality in Europe. Oral Oncol 2004;40:433-9

6. Reid BC, Alberg AJ, Klassen AC, Rozier RG, Garcia I, Winn DM, Samet JM. A comparison of three comorbidity indexes in a head and neck cancer population. Oral Oncol 2002;38:187-194

7. Rowley H, Roland NJ, Helliwell TR, Caslin A, Kinsella AR, Jones AS. p53 protein expression in tumours from head and neck subsites, larynx and hypopharynx, and differences in relationship to survival. Clin Otolaryngol Allied Sci 1998;23:57 62

8. Weymuller EA, Yueh B, Deleyiannis FW, Kuntz AL, Alsarraf R, Coltrera MD. Quality of life in patients with head and neck cancer. Lessons learned form 549 prospectively evaluated patients. Arch Otolaryngol Head Neck Surg. 2000;126:329-335

Address for correspondence:
Dr. Nikolay Sapundzhiev, MD
Department of Otorhinolaryngology,
Medical University "Prof. P Stoyanov",
55, Marin Drinov Str., Varna 9002, Bulgaria
e-mail: n.sapundzhiev@gmail.com