A Tagging System for Section Headings in a CEN Standard on Patient Record

Angelo Rossi Mori, Fabrizio Consorti, Elena Galeazzi
Reparto Informatica Medica, Istituto Tecnologie Biomediche, CNR, Roma, Italy
Istituto IV Clinica Chirurgica, Università La Sapienza, Roma, Italy
TESEO Telematic Services, Roma, Italy

ABSTRACT. CEN is developing a series of standards for transmission of patient records. We present here preliminary results to produce a standard on section headings in record systems. Each record system uses its own scheme of headings, depending on specialty, National regulations and tasks; users cannot accept a unique, standard scheme. Enumerating all possible headings in all National languages is not necessary for a safe and faithful transmission of records. The focus of CEN standard is shifting from the headings to a system of tags that can be tied to them. We want to convey just the crucial properties to assure correct data handling by receivers. Our tagging system is organized into 5 clusters:
C0 Nature, i.e. tags to identify the nature of data;
C1 Safety context, i.e. essential tags that convey the main context of data;
C2 Interpretation, i.e. tags about interpretation of data in the original context by the original user;
C3 Intention, i.e. additional tags to make explicit the links that reveal sender's intentions and goals;
C4 Organization, i.e. further tags to show the organization of the original record.
Although tags were conceived for messages, they should be present in local systems since data generation. In fact, tags can be used to organize data within a record and allow for multiple views, without imposing a unique, fixed structure to the record.

1. INTRODUCTION
CEN/TC 251 — the Technical Committee for Healthcare Informatics of the European Organization for Standards — is developing a series of standards on transmission of Electronic HealthCare Record (ECHR), upon mandate from the European Union [2]. One of the standards should assure the correct interpretation of section headings [13]. It assumes that a heading scheme used in a particular record system depends on:
- different views on data according to tasks,
- National regulations and local customs,
- needs of healthcare professionals in various environments (considering specialties, sub-specialties and allied professions, including general practice, hospitals, home care).
Therefore it will be inappropriate to develop a unique scheme of general-purpose headings (to be adopted in all European Countries). In US, LOINC is distributing a collection of headings from various sources, from which users could select their particular subset [18].

But safe and faithful transmission of patient records does not necessarily require the enumeration of all possible headings in all National languages. In fact, just a precise set of crucial properties of record elements is needed to guide the receiver in the correct interpretation of data. Therefore CEN is adopting a complementary approach, and focus of the standard is shifting from the headings themselves to a system of tags that can be attached to them.
In this paper we present the work to prepare that standard. We put the basis for comparison of headings from different record systems. As a by product, we also obtain a framework to support structured interfaces and systematic representation of information in local record systems, and a mechanism to facilitate multiple views.

2. METHODOLOGY
The GALEN-IN-USE Project — in synergy with CEN/TC251/WG II on terminology, according to CEN ENV 12264 [3] — developed tools, methodologies and skills to manage semantics in healthcare applications, based on the production of systematic dissections in a cooperative environment [5, 6, 11, 12]. We applied this know-how to the structure of actual record systems, to reconstruct the (structural) context for correct data interpretation. A scenario about development and potential content of the standard was presented in [16]. The process is evolutionary and is made of cycles of iterative refinement of 4 components (see [14, 11]: categorial structure [3], cross-thesaurus, list of relevant headings, database of systematic dissections). Results of a preliminary cycle were presented in [9].
We then systematically analyzed about 600 headings from various sources [13] to work out a tagging system for the characterization of headings. After refinement, this tagging system should be adopted as a CEN standard. In addition, the standard will provide structured definitions and systematic names for a significant list of headings from various sources. Definitions and names will be produced according to the same categorial structure used to develop the tagging system, i.e. within the same system of concepts [3]. The standard will not impose a fixed list of headings, but just provide tags. Developers and users of record systems will be allowed for systematic extensions according to defined rules and constraints. Other standards will describe how to use the tagging system in transmission and design of ECHR (e.g. in an XML environment).
3. RESULTS

3.1. Clusters of tags
The standard will consist of about 50 tags, i.e. the refinement of the values listed in lowercase below. Tags are presented here by semantically homogeneous sets, in UPPERCASE. Tags reflect the present cycle of development. Decisions should be refined by CEN/TC251/ WGI and WGII, also with limited tests, to obtain the optimal tagging system that combines easiness of usage with coverage of the vast majority of cases. We used a pragmatic criterion to organize our tag sets into 5 clusters. In fact, decisions were based on safety principles that envisage uses of the tags into messages (see 4.2).

C0 Nature
Nature of clinical data is disparate; tags should identify the kind of data to guide further processing.
FOCUS: condition, care act, management act, event
ASSOCIATE TOPIC: device, medicinal product, method

C1 Safety context
Knowledge on instantiation of concepts within the original context is crucial for correct management of information. These tags are essential in a message, as they determine its safe interpretation by the receiver.
ACTUAL SUBJECT: patient, donor, fetus, relative of patient
ACTING MODE: history taking, physical examination
KNOWING MODE: reported by patient, observed by recorder, observed by external service
URGENCY: emergency, routine
CERTAINTY: confirmed, ruled out, hypothesized
PROCESS STATUS: past, present, future
PATIENT STATE: condition xxx
META-INSTRUCTION: tags are not reliable for advanced usage (i.e. tags may be used perhaps to guide display or to retrieve information, but not for decision support systems)

C2 Interpretation
These tags are added to the previous ones, to convey knowledge about the interpretation by the HC professional within the original context.
INTERPRETATION ROLE: diagnosis, problem, side effect, complication
INTERPRETATION LINK: assigned to problem #, RELEVANCE (within a contact): primary, secondary

C3 Intention
These tags are added to give a cue on motivations and goals, i.e. to reveal the intention of the original user or patient. They allow to group data for convenience (e.g. according to actions of the professional or to triggering problems). They convey little additional factual knowledge about the patient; nevertheless they describe relevant features of the process of care and are useful for audit and quality assurance.
INTENTION ROLE: goal for professional, alert, reason for contact
INTENTION LINK: has goal xxx

C4 Organization
Tags on the organization used in the original record system allow a better understanding of the user's behavior. They do not provide additional knowledge to a professional: terminological knowledge carried by these tags could be also derived from data themselves by appropriate processing or by local look-up tables.
ABSTRACTION: cardiovascular disease, imaging procedure
ANATOMY: digestive system, respiratory system, ...

3.2. Applying tags to record entries
We are processing headings from various sources, to produce a table that associates - in principle - the most suitable tags to each heading [13].

We developed the following set of rules to apply the tagging system in a real patient record system:
- the tagging system applies not only to section headings, but also to all record entries (i.e. also to data elements and values, see below);
- tags will not replace the original text, but they assist in displaying and processing of data;
- an entry may be associated to multiple tags;
- the association of an entry to a tag can be proposed in abstract, but has to be checked when the entry is used in a particular context;
- in a coherent record system, a lower-rank entry is associated by default to all the tags defined for higher-rank entries; a more specific tag value replaces a more generic value.

The following two examples simulate the application of our tagging system to record entries.
The record system in Example #1 has a section on "history" (a record item complex in CEN ENV 12265 [4]), and a sub-section on "cardiovascular diseases" (a record item). Let us assume that the patient-related content for this record item is "myocardial infarction" (the value, in courier).
In column 1 below we represent this nesting, and we also simulate some more record entries (we skip other nested elements under "...").
In column 2, we call "function" the nature of the entry according to CEN ENV 12265, i.e. whether it is a container ("record item complex", or "record item"), or a content ("value").
In column 3 we propose the standard tags. Note that their scope propagates down in the chain.

Other record systems could be organized in different ways. In Example #2 we suppose that a record is organized by system (for example in a cardiovascular department).

In both examples, a message could transmit either a whole record, or the whole "history", or elementary facts, e.g. the value = "myocardial infarction".
example #1

<table>
<thead>
<tr>
<th>function</th>
<th>tags</th>
<th>cluster</th>
</tr>
</thead>
</table>
| history  | complex | ACTING_MODE = history taking  
ACTUAL_SUBJECT = patient  
KNOWING_MODE = reported by patient | C1 |
| cardiovascular diseases | item | **item**  
ABSTRATION = cardiovascular disease  
ANAT = cardiovascular system  
FOCUS = condition | C4 |
| myocardial infarction | **item** | **item**  
ANAT = heart  
PROCESS_STATUS = past | C4 |
| hypertension | **item** | (note: active or not? not reported here) | C1 |
| gastrointestinal diseases | item | **item**  
ANAT = gastrointestinal system | C4 |
| duodenal ulcer | **item** | **item**  
ANAT = duodenum | C4 |
| infectious diseases ... | item | **item**  
ASSOC_TOPIC = infectious disease | C4 |
| operations | item | **item**  
FOCUS = surgical procedure  
PROCESS_STATUS = past | C0 |
| aortocoronary bypass | **item** | **item**  
ANAT = cardiovascular system | C4 |
| physical examination ... | complex | ACTING_MODE = physical examination  
KNOWING_MODE = observed by physician | C1 |
| investigations ... | complex | ACTING_MODE = investigation  
FOCUS = property | C1 |

example #2

<table>
<thead>
<tr>
<th>function</th>
<th>tags</th>
<th>cluster</th>
</tr>
</thead>
</table>
| clinical assessment | complex | ACTING_MODE = clinical assessment  
ACTUAL_SUBJECT = patient | C1 |
| cardiovascular system | item | **item**  
ANAT = cardiovascular system | C4 |
| patient history | **item** | **item**  
ACTING_MODE = history taking  
KNOWING_MODE = reported by patient | C1 |
| myocardial infarction | **item** | **item**  
FOCUS = disease  
ANAT = heart  
PROCESS_STATUS = past | C0 |
| hypertension | **item** | **item**  
FOCUS = disease  
ANAT = cardiovascular system | C4 |
| aortocoronary bypass | **item** | **item**  
FOCUS = surgical procedure  
PROCESS_STATUS = past  
ANAT = cardiovascular system | C4 |
| family history | **item** | **item**  
ACTING_MODE = history taking  
KNOWING_MODE = reported by patient  
ACTUAL_SUBJECT = relative of patient | C1 |
| physical examination ... | **item** | **item**  
ACTING_MODE = physical examination  
KNOWING_MODE = observed by physician | C1 |
| investigations ... | complex | ACTING_MODE = investigation  
FOCUS = property | C1 |

In the last case, to assure that information is correctly interpreted, a message should also contain:

\[
\begin{align*}
\text{FOCUS} &= \text{condition}; \\
\text{ACTING\_MODE} &= \text{history taking}; \\
\text{PROCESS\_STATUS} &= \text{past}; \\
\text{ACTUAL\_SUBJECT} &= \text{patient}; \\
\text{ANAT} &= \text{heart}
\end{align*}
\]

Receiving software can use this tagging information, even without interpreting the actual content, to suggest the proper place for the entry in the structure of the receiving record.

4. DISCUSSION

First we present the requirements from healthcare and telematics where our results should be applied.

Then we discuss some issues about our tagging system and its usage.

4.1. Requirements for a standard on headings

Patient record is a collection of documents; messages consist of either entire documents or information rearranged from these documents.

Telematics and the web support sharing (i.e. either exchanges or browsing) of patient-related documents. Appropriate standards should set up the mechanism to convey the structure around data [10]. Moreover, whatever is the input and sharing solution, internal representation of relevant information should be suitable for advanced processing of its semantic content.

In fact, if data are adequately structured: i) quality and completeness of data can be checked against well-defined requirements; ii) the receiver can correctly interpret the information and can integrate the received messages with the existing collection of documents.

757
Continuum, containers and contents

In a previous paper [15] we argued that there is a continuum in the patient record systems across the headings of sections and sub-sections (containers), until the values (content) in lowest-rank record elements. We assume an operational definition of an information unit, as the minimal amount of details that can influence the behavior of someone (in a given context) in a significantly specific way.

An information unit can be arranged in different ways in different record systems, allocating details either into one record element or into multiple containers, each containing as value a fragment of the original information. To transmit this information unit in order to assure its faithful re-use with minimal human intervention, we have to reconstruct each container (or better each chain of containers of increasing rank, as in a set of Chinese boxes) and thus we must transmit also an adequate amount of the structural context from the original document. A standard cannot be limited to section headings, but it must deal with all the record entries that make a container. In this paper we are not dealing with representation of content [7, 14]. In particular, we ignore the non-structuring details, i.e. minor qualifiers such as "laterality" for anatomy, "extent" for surgical procedures, or "measurement circumstances" in laboratory tests. In fact, they should be considered in domain-specific standards about semantic templates (categorial structures in [3]).

Analogously, we don't process attributes already properly considered in the information model, such as dates, names, identification numbers, etc.

4.2. A pragmatic criterion to cluster tag sets

Our present scale of priority (see 3.1) is mainly guided by safety reasons. Cluster C0 (by an explicit or implicit mechanism) should be mandatory in each application because it gives the first identification of the nature of an entry (values could correspond for example to the names of the DICOM templates [19]).

C1 is also needed because it carries crucial information about context that cannot be derived from data. Then we have C2 for effectiveness, C3 mainly for friendliness. Finally C4 is for local needs of browsing and clustering (tags are redundant because they rely on medical knowledge, not on patient information or clinical context of the sender).

The tagging system will result from a compromise and will not exactly replicate the underlying categorial structure: eventually about 50 details should be selected, to become "tags" with a particular level of granularity, and pragmatically organized. For example, the categories "disease", "medicinal product" and "anatomy" have similar semantic features, but we decided to manage them in different ways. In fact, in §3, we used "disease" as a value for FOCUS, "medicinal product" as a value for ASSOC-TOPI and ANATOMY as a new tag name (its values represent different body systems and body parts). We actually considered that: i) given an ASSOC-TOPI = "medicinal product", the FOCUS should be either the "medication act" (prescription, dispensing or administration) or an "event" (accidental ingestion) or an "allergy state" with role "alert"; ii) users are more likely to organize views on the record by body systems than by generic kind of medicinal product.

The "Abstraction" tags deal with subsections actually used in a given record system. Information is provided to the receiver for awareness. If the receiving system uses the same label, it can use the value to properly classify the entry. Otherwise, the system can use its own abstractions (if it knows how to calculate them from attributes and values) or none. If the content is a "coded entry" (i.e. not just a free-text string), similar information could perhaps be generated by the receiving application — i.e. retrieved in some local look-up table or calculated from the FOCUS plus the associate topic; e.g., one could produce "ABSTRACTION = CV disease" from "FOCUS = disease" + "ANAT = CV system" (note that processing of sensible combinations is not a trivial task, i.e. it requires a terminology server [7]). However, this cannot replace the information about the original organization of data. Finally, it is more safe and efficient if tagging is prepared from the original context — and validated according to it — by the sending system.

4.3 XML and semantic tagging

Integration of clinical data is the next challenge for Healthcare Informatics [1,17]. It can be realized at various levels, as in the Kona proposal [8], developed by the SGML/XML SIG of HL7.

Our approach is compatible with Kona, and could bring to a richer implementation of its level 2. Each heading and each record item implies various details (descriptors) that have specific functions. In fact, many entries can be classified in multiple ways, according to different perspectives (i.e. depending on which detail we are considering).

A standard on headings in patient records should shift to a system of tags to describe all record entries. Each standard tag will be the representation of a relevant detail from either the content or its "structural context". Our goal is to allow for safe interpretation first, and then to tie knowledge to data for their re-use.

For example, using the information provided by the tags, an application could organize data in many ways:

- select all entries on "history of diseases" (FOCUS = disease, ACTING MODE = history taking). The system will include entries on various body parts;
- select all entries on "heart diseases" (FOCUS = disease, ANAT = heart). The system will include entries on both history and current diseases;
select all entries on "history about heart" (ACTING MODE = history taking, ANAT = heart). The system will include entries from the history, about diseases and surgical operations on heart;

* select " all data about heart" (ANAT = heart). The system will include also entries on lab data (EKG) and images.

Appropriate tagging will make documents more independent from applications (i.e. data will be re-usable in various contexts). Semantic tags — as opposed to formatting tags in HTML — can be effectively exploited by XML. Other standards on architecture and messaging (for example in CEN and HL7) could put the precise requirements on the actual usage of the tagging system in XML.

In addition to the standard on tag names and potential values for descriptors, we also need appropriate separate standards on codes for the values and perhaps on abstractions.

5. CONCLUSIONS

Clinical information is facing the Web revolution. Patient record is a collection of documents, made of information units that can be clustered in many ways, according to user’s purposes. Standards will support faithful and sure exchange of either whole records or selected information units, and remote browsing of documents. Our tags convey crucial details (see clusters C0 and C1) and some additional knowledge mainly on the “container part” of information units. Our goal is to avoid misleading processing of clinical information in situations that imply (Web-based) re-uses of data in various contexts.

*Unless we put severe constraints on communication of clinical information (i.e. on topics, granularity or formats), or we enforce a complete human control, crucial tags must be always made explicit.*

The cost of tagging is balanced by improvements not only in safety and quality, but also in effectiveness of processing. In fact, they can be used not only by further standards in CEN or HL7 on record architecture and messaging, but also for structured input and for XML-based applications: e.g. internal representation of documents, semantically driven browsing, rearrangement of documents into messages and in the near future for indexing in document databases.

Acknowledgements. Work partially financed by the EU Project HC2018 GALLEN-IN-USE. We are grateful to David Markwell and to members of CEN/TC251/PT27. This paper expresses views of the Authors and is not on behalf of CEN/TC251 WGII or PT27.

REFERENCES


2. CEN TC251 “Health Informatics” at http://www.centc251.org


7. GALLEN and GALLEN-IN-USE (1992-98) documentation available from AL Rector, Medical Informatics Group, Dept. Computer Science, Univ. Manchester, Manchester M13 9 PL, UK e-mail galen@cs.ac.man.uk; see http://www.cs.man.ac.uk/mig/galen and http://www.cs.man.ac.uk/mig/giu


9. Tomelo (an European Project) see documentation at http://www.chem.kun.nl/tomelo/#Workshops


13. Rossi Mori A. (ed) Preliminary material of CEN/TC251/PT27 on WI 1.8-2 ECHR-part 2 Domain Templet, (available on request at rossi@color.imrkant.rm.cn.it )


15. Rossi Mori A, Galeazzi E, Consorti F, Bidgood WD Conceptual schemata for terminology: a continuum from headings to values in patient records and messages JAMIA symp. suppl., 1997


