

Implementation of Interinstitutional and Transnational Remote Terminology Services

Daniel Luna M.D.^a, Gastón Lopez^a, Carlos Otero M.D.^a, Alejandro Mauro M.D.^b, Claudio Torres Casanelli^b, Fernán González Bernaldo de Quirós M.D.^a

^aDepartment of Health Informatics, Hospital Italiano de Buenos Aires, Argentina

^bMegasalud, Chile

ABSTRACT

In view of the benefits yielded by the Terminology Server running in the Healthcare Information System of Hospital Italiano de Buenos Aires, the Institution implemented all the necessary changes in order to offer such services to other health-care institutions, using the Internet as communication vehicle. Megasalud is the most integrated healthcare network of Chile and decided to change their legacy system and develop their own Healthcare Information System and started to use Remote Terminology Services. After the implementation of these Terminology Services we tested the performance for identifying free text added in their electronic health record. Between 78% to 89% of text entered was recognized. The task of creating an institutional interface terminology provides an excellent service to the users, as they have the freedom to use free-text entry.

INTRODUCTION

Data entry is an obstacle for the usability of Electronic Health Records (EHR) applications and the acceptance of physicians, who prefer to document health care findings, processes, and outcomes using "free text" [1-2]. However, Healthcare Information Systems (HIS) should capture the clinical data in structured and preferably coded format to support research and enable decision support programs used at the point of care [3]. In order to address this point it has been developed numerous terminological systems for the systematic recording of clinical data. These systems inter-relate concepts of a particular domain and provide concepts with related terms and possible definitions and codes [4]. Terminological systems can be distinguished as aggregate terminologies, reference terminologies, and interface terminologies, each used for different purposes and each serving different requirements regarding their intended use and domain. The aggregate terminology (e.g. the International Classification of Diseases in their different editions) suffers of lack of structure and formal semantic definitions that results in shortcomings when aiming for data reuse [5]. Reference terminologies are those designed to

provide exact and complete representations of a given domain's knowledge, and their interrelationships, which are typically optimized to support the storage, retrieval, and classification of clinical data [6], the best representative of this group is the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [7]. The interface terminologies are systematic collections of clinically oriented phrases aggregated to support clinician's entry of patient information directly into computer programs, such as clinical documentation systems or decision support tools [8]. These terminologies facilitate the display and collection of clinical data in a simple way while simultaneously linking user's own descriptions to structured data elements in a reference terminology or aggregate terminology. All of these types of terminological systems can be grouped in a terminology server and moving from the basic model composed of a list of codes and descriptions, to a complex system of conceptual representation of the medical vocabulary [9].

The Hospital Italiano de Buenos Aires (HIBA) has developed a local interface terminology [10] in a context of a terminology server [11] with the aim of providing support to clinical documentation and autocoding of clinical data in the context of their HIS [12].

The most integrated health network of Chile, Megasalud, was using for a decade an EHR named SiapWin. In 2007 decided to develop their own HIS allowing longitudinal care of patients treated in the network with the mentoring of the medical informatics expertise of HIBA. In behalf of this project HIBA decided to modify the functionality of their terminology server to provide terminology services to other institutions.

The aim of this study is to describe and quantify the use of Remote Terminology Services (RTS) provided by HIBA through a transnational and interinstitutional implementation.

MATERIALS AND METHODS

The Terminology server of HIBA

The terminology server of HIBA is composed of a local interface terminology (thesaurus) [10] mapped to a reference terminology, SNOMED CT [11]. The

thesaurus consists of a list of terms created from almost 2 million free text inputs extracted from the clinical data repository. The terms included in the thesaurus are divided into concepts (real clinical entities) and descriptions (different ways of naming these clinical entities). Physicians entered these terms into specific domains (problem list, procedures, discharge summary, etc.) in the problem-oriented EHR.

The interface terminology is updated daily by a professional team who audit, code and link each new term to the SNOMED CT as a reference terminology, and use the official mapping into SNOMED to another classifications (like ICD 10). When SNOMED doesn't offer an official mapping, the team generates a manual cross-link through functionality on the terminology server.

When the performance of autocoding was measured, we found that about 80% of the texts found in a problem list [13-14] or discharge summary [15] were automatically coded; the remaining 20% required manual coding.

Design, develop and implementation of Remote Terminology Services

Modifications were undertaken in our terminology server with the goal of providing remote terminology services to other institutions, both in Argentina and other Spanish speaking countries.

Technological issues

In the layer of access to information Web Services developed with JAVA, JDK 1.6 was used. The Web Services (WS) were deployed in a SUN's Glassfish application server, and the data was stored in an Oracle 11g database.

The WS were published in the Internet for the remote access of the applications of other institutions.

Terminological Web Services

Published WS allow the most of the progress achieved by HIBA in the management of terminological domain. There are several services that can be used to process the text entered by a physician in their distance applications. Table 1 shows the complete list of services and their significance.

Implementation

The first step in the new data model of the Terminology Server was the creation of an institutional interface terminology (thesaurus) for each organization interested in use of the RTS. This thesaurus is created from the normalization of terms included as free text within specific fields, like problem list or discharge summary. Each of these thesauruses was stored in the Terminology Server of

HIBA. This replication process allowed them to obtain the experience of one decade of auto-coding at their start-up. After deployment was completed, auditing professionals of HIBA coded each new term entered into the interface, tailoring the terminology to each institution's own vocabulary.

| Service | Description |
|------------------------|--|
| Intelligent prompting | Perform a preliminary search entering the first three characters. |
| Term Recognition | Search for the text entered in the interface vocabulary and offer the alternative to improve the medical record. |
| Creation of a new Term | Enter new term in the interface vocabulary and it is entered into the audit circuit. |
| List Classification | Return back available Classification. |
| Assign classifier | Valid term plus classification return back the corresponding code. |
| Assign DRG | From a discharge summary encoded with ICD9-CM and other metadata, returns back DRG code. |
| List Domains | Return the domain available (Problems, Procedures, Medications, etc.). |
| List domain elements | Returns back terms contained in a domain. |

Table 1- Terminological Web Services provided by HIBA

When the EHR is in use, a physician enters the first three characters from a text, the EHR requests the *Intelligent Prompting Service*. This service searches in the interface terminology of the specific institution and returns the 5 most common alternatives, with an appropriate level of detail. More specific searching is given when increasing the quantity of characters.

If the clinician selects a term from the list, the application records the corresponding code (Thesaurus code)¹, ending the registration process.

If the professional prefers not to select a term from the list and continues his input, the EHR application requests the *Term Recognition Service*. This service searches for the text in the interface terminology and one of the alternatives described below can occur:

- **Text is valid and it's entered into an adequate level of detail.** The application records the corresponding code of Thesaurus concluding the registration process. e.g. Measles².

¹ Thesaurus code: each term has a unique number in the interface terminology.

² The original language of all these examples is Spanish

- **Text is valid and there are more specific options.** The application offers a list of options; the physician can select one of them or keep the originally entered text. e.g. AH (Arterial Hypertension).
- **Text is valid but requires more specific level.** The application offers the list of options and the physician must select one of them, with a compulsory refinement. e.g. Diabetes Mellitus.
- **The text contains more than one medical concept.** The application displays the list of terms, where the physician can select one or all of them. e.g. cardiac insufficiency and pneumonia.
- **Text is ambiguous.** The application offers the list of options so the physician must select one of them. e.g. Angina (Angina in Spanish meaning chest pain or faryngitis).
- **Not valid text.** The application displays the corresponding explanation. e.g. Vaccines cardboard.
- **Text is not recognized but there are alternatives with similar syntax.** The application offers the list of options where the physician could select one of them or keeps the original text entered. e.g. Migraine instead of Migraine.
- **Text is not recognized as a health problem.** The application offers the physician the opportunity to keep the text entered by calling *Creation Term Service* and creates the term in the interface terminology and stores for the manual audit process. e.g. Stener Lesion.

To code the terms in the EHR by a specific classification, the coding application requests the *List Classification Service*, to select the appropriate classification. The system displays a list of classifications available and the operator must select one of them.

The system then uses the *Assign Classifier Service* to assign the code for each term.

Using this mechanism it is possible to select the classifier ICPC-2 for the epidemiological analysis from a problem list of the outpatient EHR, ICD-9 CM and ICD-10 for a discharge summary in the inpatient EHR. This mapping is possible because we used the official cross-match offer by our reference terminology (SNOMED) or creates our own mapping by the specific terminology team.

From a discharge summary coded in ICD-9 CM it may apply the *Assign DRG Service* to obtain the corresponding code.

An alternative implementation for free text entry is the creation of a structured entry interface based on

lists. The EHR application must request the *List Domains Service*. Once the physician selects the domain, the application must request the *List Domain Elements*, which will provide a complete list of terms, allowing a selection.

Megasalud HIS

In 2007, Megasalud decided to undertake the process of migrating their existing legacy system (and its stored data) into a new HIS. The new system would enable the development of a problem-oriented EHR where information entered in free text by the physician's would be coded by using RTS.

Remote Terminology Services utilization

The implementation was carried out in two stages, the start-up and the real time utilization of RTS.

Start-up

The aim at this point was to extract the greatest amount of clinical information possible from the existing system (mostly in free text), and add this information into the new clinical data repository by coding it. To this purpose, extracted data were processed by the RTS and coded it when it was possible. This data included allergies, reason(s) for the consultation, habits, risk factors, symptoms and diagnosis entered by physicians in a free text form, and only coded diagnoses when they felt it particularly necessary. This processing was undertaken in September of 2008 [16].

Real time

Once the problem-oriented EHR of Megasalud was implemented, it began using some of the Terminological Web Services provided by the RTS in real time. The Terminological Web Services were used for the free text data entry in the problems list and procedures of the EHR.

Evaluation of the RTS

With the objective of quantifying the use of RTS provided by HIBA through a transnational and interinstitutional implementation we undertook three different performance indicators:

- Amount of data recognized and autocoded in the start-up of utilization of RTS, September 2008.
- General performance of the real time utilization of the RTS from March 1st to October 1st 2009.
- Detailed analysis of the real time utilization of the RTS from August 28th to October 1st 2009.

RESULTS

The Start-up performance

The clinical data stored in the legacy system of Megasalud were 14,120,751 single text phrases enabled to process by the RTS.

With the batch processing of these data, the RTS recognized and auto coded 11,118,760 (78.74%) texts (included valid and not valid text), and did not recognize 3,001,991 (21.26%) of the original data. The different categories from where clinical data was extracted, with the performance of the RTS in detail are shown in Table 2.

General real time performance

In the period between March 1st to October 1st 2009 the physicians at Megasalud entered 592,249 pieces of text in the problem-oriented EHR, 530,897 (89.64%) of them were successfully recognized in the interface terminology of Megasalud by the utilization of RTS in real time. The remainder 61,352 (10.36%) went under the audit process and manual modeling.

| Categories | Recognized n (%) | | Not Recognized n (%) |
|-------------------------------|-----------------------|-------------------|----------------------------|
| | Valid | Not Valid | |
| Allergies | 36,539 (0.26) | - | 11,359 (0.08) |
| Not coded | 6,660,487 (47.17) | 557,025 (3.94) | 1,037,237 (7.35) |
| Diagnoses | 1,543,798 (10.93) | 20,482 (0.15) | 809,867 (5.74) |
| Coded | 119,452 (0.85) | 4,292 (0.03) | 8,630 (0.06) |
| Risk Factors | 577 (0.00) | 27,006 (0.19) | 15,943 (0.11) |
| Habits | 2,012,423 (14.25) | 136,679 (0.97) | 1,118,955 (7.92) |
| Reason(s) for consultation | 10,373,276 (73.46) | 745,484 (5.28) | 3,001,991 (21.26) |
| Total | | | |

Table 2- Start-up performance of RTS

Detailed analysis

For the purpose of a detailed analysis of the RTS from Megasalud we analyzed log files from WS consumption from August 28th to October 1st 2009. The physicians performed 82,072 queries to the WS during the period under evaluation. The WS recognized the text entered 62,186 times (75.77%), and was not able to recognize the entry text 19,886 times (24.23%). Whenever the entry text was not recognized, the user was allowed to enter the text *not-recognized as a health problem* and it was sent to the coding team to code it, or to select an option that the RTS could bring up based on the matching root algorithm (e.g. if the user enters "ulcerative colitis" and the system does not have that text entry modeled as a "valid clinical entity but wrongly typed", it will look at the root text entered "Ulc + Col" and will try to give up some options, like "ulcerative colitis", "acute ulcerative colitis" or "exacerbation of ulcerative colitis")

Nearly 70% of the time (55,253 texts representing 67.32%) the entry text was recognized as a valid

clinical entity and the user was given options to extend that clinical entity providing a more detailed description. 43,874 times (79.41%) the text string enter was an correctly written and clinically relevant (Valid Description), and 33,823 times (61.21%) the entry text entered was the preferred description for that health problem (e.g. for CVD, the preferred description is "Cerebrovascular Disease")

In 2.47% of the times the entry text entered had no refining options, being a valid description only 40 times and a preferred description 13 times.

In 4,785 times (5.83%) the user was given the response "Not Valid Text". The most frequent response was "not valid because it is not a health problem". In 29.30% of the times (1,402) the response was "Add a finding site to this problem" and less frequent were "Invalid Acronym", "Text is ambiguous" or "The text contains more than a medical concept ". Error search strings were rare (0.01%).

DISCUSSION

It has been previously described that interface terminologies are unique vehicles for supporting efficient and accurate interaction between healthcare providers and computer-based clinical applications [8]. The use of a standard terminology provides interoperability among the HIS used by different stakeholders of the healthcare system, allowing the exchange of information between health institutions, suppliers, government offices, researchers, etc. The task of creating an institutional interface terminology demands a lot of work, but provides an excellent service to the users, as they have the liberty to enter information in free-text style.

We consider of great value to provide services to other institutions by our RTS. Creating and maintaining a sharable Spanish interface vocabulary database between different countries is a big task as medical Spanish is a rich vocabulary and there are different ways of naming the same clinical entities (polysemy), and different acronyms and synonyms between countries.

The overall recognition rate in this experience was higher than 75% providing with the possibility to Megasalud to have clinical data coded both from their historical stored data and the data of their new EHR.

The highest percentage of terms not recognized in the period of detailed analysis (1/3 of the total since starting using the RTS) can be explained by the progressive increase of physician's using the EHR; which was implemented gradually in each of its 30 facilities across Chile.

Our work has some limitations to be considered. In the start-up stage, many of the previously stored text were recognized as acronyms. These texts may have

different meanings in Argentina and in Chile. A lot of work is needed to perform a manual revision to assure a good quality in the Megasalud interface terminology.

After the implementation of the problem-oriented EHR was necessary undertake changes in the some preferred clinical terms because the physicians in Chile named in different ways some conditions.

In the near future we will start the implementation of these RTS in a large HIS of Uruguay.

CONCLUSION

The use of Terminology Services via the Internet will enable healthcare institutions to quickly implement a complex and comprehensive solution for their coding issues. Institutions will be able to enhance their own HIS by providing functionality that will be simple and intuitive to health care professionals.

Since the implementation, participating institutions have had an extensive vocabulary for the start-up and were provided with close to a decade of continuous maintenance and updating. They are now able to have an independent vocabulary supported by a highly trained staff working at the Terminology Area of Health Informatics Department of the HIBA.

In the future these institutions will be able to run their management processes automatically, using the classification that best adapts to their local requirements, without the need to introduce changes in the stored information.

This system will pave the way to the integration of EHR and clinical decision supporting systems (CDSS).

References

- [1] Middleton, B., K. Renner, and M. Leavitt, *Ambulatory practice clinical information management: problems and prospects*. Healthc Inf Manage, 1997. **11**(4): p. 97-112.
- [2] Rector, A.L., *Clinical terminology: why is it so hard?* Methods Inf Med, 1999. **38**(4-5): p. 239-52.
- [3] McDonald, C.J., *The barriers to electronic medical record systems and how to overcome them*. J Am Med Inform Assoc, 1997. **4**(3): p. 213-21.
- [4] de Keizer, N.F., A. Abu-Hanna, and J.H. Zwetsloot-Schonk, *Understanding terminological systems. I: Terminology and typology*. Methods Inf Med, 2000. **39**(1): p. 16-21.
- [5] Rose, J.S., B.J. Fisch, W.R. Hogan, B. Levy, P. Marshall, D.R. Thomas, and D. Kirkley, *Common medical terminology comes of age, Part Two: Current code and terminology sets--strengths and weaknesses*. J Healthc Inf Manag, 2001. **15**(3): p. 319-30.
- [6] Rose, J.S., B.J. Fisch, W.R. Hogan, B. Levy, P. Marshall, D.R. Thomas, and D. Kirkley, *Common medical terminology comes of age, Part One: Standard language improves healthcare quality*. J Healthc Inf Manag, 2001. **15**(3): p. 307-18.
- [7] *The International Health Terminology Standards Development Organisation (IHTSDO)*. [cited 2008 10-Oct-2009]; Available from: <http://www.ihtsdo.org/>.
- [8] Rosenbloom, S.T., R.A. Miller, K.B. Johnson, P.L. Elkin, and S.H. Brown, *Interface terminologies: facilitating direct entry of clinical data into electronic health record systems*. J Am Med Inform Assoc, 2006. **13**(3): p. 277-88.
- [9] Chute, C.G., P.L. Elkin, D.D. Sherertz, and M.S. Tuttle, *Desiderata for a clinical terminology server*. Proc AMIA Symp, 1999: p. 42-6.
- [10] Lopez Osornio, A., D. Luna, M.L. Gambarte, A. Gomez, G. Reynoso, and F.G. de Quiros, *Creation of a Local Interface Terminology to SNOMED CT*. Stud Health Technol Inform, 2007. **129**: p. 765-9.
- [11] Gambarte, M.L., A. Lopez Osornio, M. Martinez, G. Reynoso, D. Luna, and F.G. de Quiros, *A practical approach to advanced terminology services in health information systems*. Stud Health Technol Inform, 2007. **129**: p. 621-5.
- [12] Luna, D., P. Otero, A. Gomez, M. Martinez, S. García Martí, M. Schpilberg, A. Lopez Osornio, and F.G. Bernaldo de Quiros, *Implementación de una Historia Clínica Electrónica Ambulatoria: "Proyecto ITALICA"*. in *6to Simposio de Informática en Salud - 32 JAIIO*. 2003. Bs As, Argentina: Soc. Argentina de Informática e Investigación Operativa (SADIO).
- [13] Lopez Osornio, A., S. García Martí, M.L. Gambarte, G. Staccia, M. Schpilberg, E. Soriano, P. Otero, D. Luna, and F. Gonzalez Bernaldo de Quiros, *Creation of a Web-based ICPC encoder and a secondary codification variability Developing tools from a free text variability analysis of diagnostic descriptions records*. in *Medinfo*. 2004. San Francisco, California, USA.
- [14] Lopez Osornio, A., S. Montenegro, S. García Martí, L. Toselli, C. Otero, I. Tavasci, D. Luna, A. Gomez, and F. Gonzalez Bernaldo de Quiros, *Codificación múltiple de una lista de problemas utilizando la CIAP-2, CIE-10 y SNOMED CT*. in *3er Virtual Congress of Medical Informatics - Informedica*. 2004.
- [15] Navas, H., A. Lopez Osornio, A. Baum, A. Gomez, D. Luna, and F.G. de Quiros, *Creation and evaluation of a terminology server for the interactive coding of discharge summaries*. Stud Health Technol Inform, 2007. **129**: p. 650-4.
- [16] Torres Casanelli, C., H. Navas, S. Benitez, L. Biaggini, G. Morales, P. Navarro, D. Luna, F. Gonzalez Bernaldo de Quiros, and M. Maira, *Implementación de servicios terminológicos en una red de atención ambulatoria*. in *INFOLAC 2008 - 3er Congreso Latinoamericano de Informática Médica*. 2008. Buenos Aires, Argentina.