URT "Indexing and Classification Systems" Projects and Biomedical Knowledge Standards

Maria Teresa Chiaravalloti*, Erika Pasceri**, Maria Taverniti***

* Università degli studi della Calabria, via Pietro Bucci cubo 17/b, Rende (CS), Italy,
<mariateresa.chiaravalloti@unical.it>
** Università degli studi di Udine, via Palladio 8, 33100 Udine, Italy,
<erika.pasceri@unical.it>
*** Consiglio Nazionale delle Ricerche, piazzale Aldo Moro 7, 00185 Roma, Italy,
<maria.taverniti@unical.it>

Maria Teresa Chiaravalloti graduated with excellence in Computer Sciences for the Humanities at University of Calabria, Italy. In August 2009, she won a research grant contract with the Consiglio Nazionale delle Ricerche. In December 2009, she passed the selection of the PhD School "Andre Gunder Frank" at University of Calabria, focusing her interest on clinical classifications and biomedical terminologies. Since December 2010, she has been a short-term scholar at Regenstrief Institute, Indianapolis, working on the Italian translation of LOINC (Logical Observation Identifiers Names and Codes). Her main interests are in knowledge management, terminology, automatic processing of texts, information retrieval, and document management.

Erika Pasceri has been a PhD student in "Bibliographic, Archival, Documentary Sciences" at the University of Udine, Italy since January 2010. She earned her degree in July 2009 in Informatics for the liberal arts and then started to work as a research fellow for the Italian National Council of Research (CNR) in a project on biomedical terminologies. She works at the URT-CNR, University of Calabria, Rende, Italy. Her main interests are in knowledge management, terminology, automatic processing of texts, information retrieval, and document management.

Maria Taverniti works for the CNR of Italy and collaborates with the Labdoc (University of Calabria) as researcher assistant. She is interested both in knowledge management activities, like analysis and representation of tacit knowledge, and in the construction of terminological controlled tools, such as faceted thesauri. She teaches cataloguing and indexing methods at the University of Salerno. She has collaborated on projects aimed at the formalization of tacit knowledge of the craftsmen of Calabria by performing activities of knowledge acquisition, elicitation, and modeling. Since 2005, she has been the educational trainer for Cataloguing and Classification teaching at the University of Calabria.

Chiaravalloti, Maria Teresa, Pasceri, Erika, Taverniti, Maria. **URT "Indexing and Classification Systems" Projects and Biomedical Knowledge Standards.** *Knowledge Organization, 39(1)*, 3-12. 6 references.

ABSTRACT: This paper presents an overview of the activities of the "Unità di Ricerca presso Terzi" (URT) – "Indexing and classification systems" of the National Research Council (CNR), which deals with research in organization and document management systems, indexing, classification, knowledge, and content management. Particular attention will be focused on the project of the Electronic Health Record Technological Infrastructure (InFSE), in which the URT works on the definition of ontological models, and vocabularies supporting EHR with the aim to build an updated Italian version of a unified medical reporting language. Although current Italian legislation makes mandatory the use of internationally recognized classification systems to help interoperability and exchange of information, the current structure of those systems does not allow their full use in the Italian context. For all these reasons, it is necessary to make some adapta-

tions in the International Classification of Disease 9th Revision terminology, used for diagnosis encoding, and in the Logical Observation Names and Codes, used for encoding laboratory observations.

Received 22 July 2011; Accepted 22 July 2011

1.0 Introduction

The research unit "Unità di Ricerca presso Terzi" (URT, "Indexing and classification systems") was established by a December 6, 2006 resolution of the National Research Council (CNR) on the basis of an agreement among University of Calabria, Calabria Region, and CNR. Among the main objectives of URT, there are: the study of fundamental indexing and classification systems; the definition of models and documental languages; and the analysis, recovery and formalization of knowledge, with a special focus on entrepreneurial and handcraft activities. Among the most important activities are two projects, respectively, of the National Centre for IT in Public Administration (CNIPA) and of the Department for Innovation of the Presidenza del Consiglio dei Ministri, the first aimed at the automatic classification of digital documents, CLAUDIO (Cavarretta et al. 2009) and the second, discussed in detail in the following paragraphs, about biomedical data encoding within the Electronic Health Record (Aracri et al. 2011).

CLAUDIO is an open source software able to process and classify, in a semi-automatic way, documents recorded in a public records system. At the moment it is used in Public Administrations. CLAUDIO aims to realize an open-source software application that can operate an automatic or semi-automatic classification of documents, stored in a computer with a protocol system and which could be integrated into the main part of the systems currently used by public administrations. This prototype software was developed in consideration of the fact that in each public administration exists a set of properly classified documents that can be used to train the automated classification system after being validated by domain experts. The already classified corpus of documents provides an opportunity to build a Machine Learned Model (MLM) proposing a classification code for a document basing on a computational analysis of the words entered in the object field by an operator. The critical point of automated classification systems lies in the methodology of similarity calculation among documents. The developed prototype uses the TF-IDF weight (term frequency-inverse document frequency), a function used in information retrieval to measure the importance of a term relating to a document or a collection of documents. This function increases with the number of times the term is contained in the document, but it grows in inverse proportion to the frequency of the term in the collection. The base is to give more importance to terms that appear in the document, but they are generally infrequent. To each query, based on the results obtained, is assigned a score of similarity. The score measures the similarity of *q*, the given query, with *d*, an object indexed in search engines. The result obtained is the record associated to the query and each record containing the relative rankings. All the records are arranged according to a pivot method based on class value that serves as primary key. A classification score is generated by averaging the highest score of similarity and the geometric mean of all similarity scores. The class that gets the highest point is proposed as first possible candidate for classification.

In collaboration with ITC-CNR, the URT worked on the project InnovANCE-Industry 2015, aimed at defining a specialized thesaurus (Cardillo et al. 2009) in the construction and energy efficiency domain, which can manage language and technical evolution, constituting the national reference for all phases of the construction process (design, implementation, and management), as well as the operational phases associated with them. The activity aims to create a free accessible database containing all relevant information, whether technical, scientific, economic, legal, etc., that is useful for creating the first national database in the construction field. The system will allow the networking of all involved users in order to facilitate the circulation of know-how among the different parties involved. The system, thanks to collecting, cataloguing, and distributing data, will promote the integration of subjects and process phases, exploiting the existing potentialities for interoperability among all the different existing software (CAD, management, energy, etc., according to already available ISO standards) ensuring a continuous updating of the information collected and distributed. The program will also allow the use of a common language, through a shared semantic network among different users: clients, component manufacturers, construction companies.

Between July 2007 and July 2008, the URT coordinated and implemented as lead agency, a project funded by Region Calabria on March 15, 2002 (under the regional law, *Rules on the protection, restoration and promotion of artistic and typical handcraft in Calabria*, which was realized with the Co.Se.r. (Consortium of Services for Handcraft), Calabria). This collaboration, whose goal is the definition of methodologies and tools for tacit knowledge recovery and accumulation in the Calabrian gold handcrafts domain, produced Goldthes, the faceted thesaurus in the gold handcraft field (Cardillo et al. 2010). Methodologically, the project included different steps in order to adopt the CommonKADS standard: by acquisition and elicitation of knowledge through techniques such as interviews and self-reports with domain experts, and by the knowledge analysis acquired both in terms of key concepts and relations between them, and by creating some models provided by the reference methodology as well as models built using the tools of knowledge representation PCPACK5. The results obtained consist of two interdependent products: a knowledge base (KB) and a thesaurus.

2.0 InFSE Project

The project "Infrastruttura tecnologica del Fascicolo Sanitario Elettronico" (InFSE; FSE is the Italian acronym for EHR) is intended to implement solutions to ensure semantic interoperability among different e-health infrastructural architectures for the construction of ontological models supporting EHR. EHR National guidelines, approved in February 2011 from Ministero della Salute, established (translated by the authors):

> The Electronic Health Record is a set of digital data and documents relating to health and social health clinical events generated by present and past, about the patient. The Electronic Health Record, which has a time frame that covers the entire life of the patient, is fed continuously by those who take care of the insured under the National Health Service and regional health and social services.

The project was motivated by the need to define a standardized terminology to be incorporated into the EHR for the unique diagnoses and laboratory data encoding both in prescription and reporting processes. The basic idea of the project is to create a national reference model of EHR that imposes the need to adopt open technologies and shared systems of encoded information, based on internationally recognized standards in order to ensure semantic interoperability in multple applications. The Technical Rules published on December 1, 2009 by the Presidenza del Consiglio dei Ministri, Dipartimento per la digitalizzazione della Pubblica Amministrazione e l'innovazione tecnologica in cooperation with the permanent work table for the Electronic Health Regions and Autonomous Provinces titled *Standard tecnici per la creazione del 'Documento di Prescrizione' secondo lo standard HL7-CDA Rel. 2* prescribe the use of ICD9-CM (International Classification of Disease 9th Revision) codes for diagnosis in prescription and reporting and LOINC (Logical Observation Names and Codes) codes to codify electronic prescription sections. In compliance with current legislation, we proceeded to analyze the two coding systems by both structural and conceptual points of view in spite of a purely practical perspective, in order to evaluate their applicability and usability into the context of use.

A preliminary analysis revealed that ICD9-CM is extremely complex in its structure and unrepresentative for the reference domain, rarely used in prescription and reporting; on the other hand, LOINC is suitable not only to codify electronic prescription sections, but also to standardize laboratory observations in Italy. Because of this, the URT oriented its activities to the improvement of the two coding systems trying to adapt them to the context of use and in relation to real needs of the final users. Concerning ICD9-CM, the main issue was its inadequacy to represent the terminology currently used by physicians, so it was decided to develop ICD9-CM Plus: an enriched version of ICD9-CM created through VILMA (Vocabolari Italiani Lessici di Medicina e sAnità) to facilitate the access to specialized information. A software tool created for this purpose, VILMA is based on a multi-user, centralized architecture that allows enriching the master terminology (ICD9-CM) with approximately 45,000 thesaural relationships among terms included as an extension of ICD9-CM. We proceeded also to the creation of the Italian official translation of the LOINC database (in collaboration with Regenstrief Institute, Inc., who is creator and curator) and the construction of software to support Italian laboratories in mapping local tests to the LOINC database, in order to enable semantic interoperability in the exchange of information among different health facilities.

An internationally respected informatics and healthcare research organization, the Regenstrief Institute is recognized for its role in improving quality

of care, increasing efficiency of healthcare delivery, preventing medical errors and enhancing patient safety. Established in 1969 by philanthropist Sam Regenstrief on the Indiana University-Purdue University Indianapolis campus, the Institute is supported by the Regenstrief Foundation and closely affiliated with the Indiana University School of Medicine and the Health and Hospital Corporation of Marion County, Indiana (http://www.regenstrief.org/).

3.0 Biomedical reporting thesaurus

The first step in creating the thesaurus was collecting national EHR projects. This helped to identify objects constituting EHR models. Particular attention was focused on the language sources for disease description. International biomedical lexicons and taxonomies were also analyzed, and those with a native Italian version and an official translation were isolated. This activity allowed the identification of lexicons/dictionaries (some of them structured and others just simple term lists) to be used for the growth and enrichment of the classification system required by technical rules.

Lexicons, merged into a single set of terms conventionally named UNICO, were integrated into ICD9-CM through VILMA software. In addition, deleting terms already present in ICD9-CM allowed the identification of redundancy due to repetition often difficult to detect manually; the tool also allowed the defining type of relations (synonymic, hierarchical, or associative) to link different terms in UNICO with the terms contained in the master terminology. Despite additions to the original classification system, ICD9-CM Plus results are still inadequate for physicians' requirements, because, in some cases, strings are too complex in describing diagnoses, such as '*frattura esposta della volta cranica con emorragia subaracnoidea, subdurale ed extradurale con perdita di coscienza di durata non specificata*' (code 800.76). The system doesn't manage synonymy relations well; for example, physicians often use 'rinorragia,' but, in ICD9-CM, there is only the term 'epistaxis' (code 784.7), which is a synonym. Another case is the absence of simple names of diseases; they are accompanied by modifiers and specifications, such as in the case of 'obesità' that is present only as 'sovrappeso e obesità' (code 278.0) and 'obesità grave' (cod . 278.01).

To better understand methodologies and techniques, main features of the software created are described below (as shown in Figure 1). After term recognition and interpretation rules were established, a term list composed of distinct terms was semi-automatically created; known conventionally as UNICO, it maintained the original link between each term and its original lexicon. Some relationships were proposed automatically by the software, based on semantic rules established during the processing phase. They were validated and maintained if correctly defined; if erroneously constructed by the software, they were deleted or reformulated.

All the terms in UNICO were related to ICD9-CM through different types of relationships established by operators. As shown in Figure 2, the system allows each operator to work on a term in mutual exclusion, and this prevents different operators from working simultaneously on the same term.

Researching terms with which to establish relationships can be done through different types of filters (Figure 3):

- "**contain**": results are all ICD9-CM Plus terms that contain the term even as part of other terms;
- "**starting with**": results are all ICD9-CM Plus terms that begin with the string specified;
- "**end up**": results are all ICD9-CM Plus terms ending with the string specified;
- "**exact match**": result is the exact correspondence of the string specified;

Figure 1. Integrating terms through V.I.L.M.A.

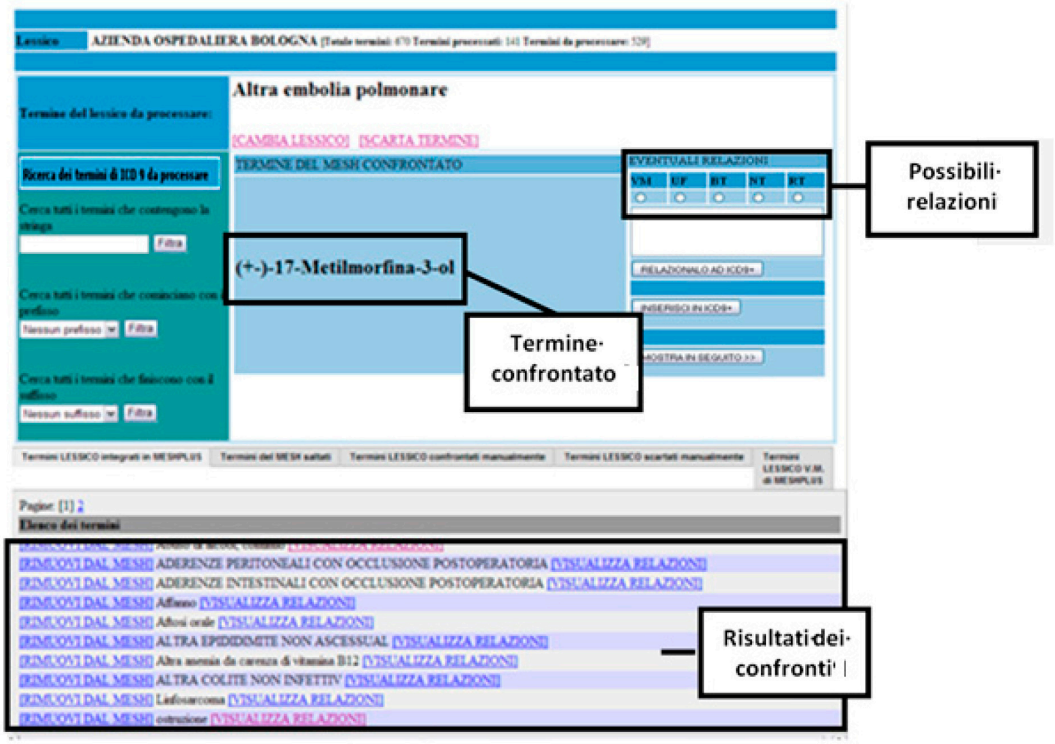

Figure 2. Screen shot of V.I.L.M.A. software.

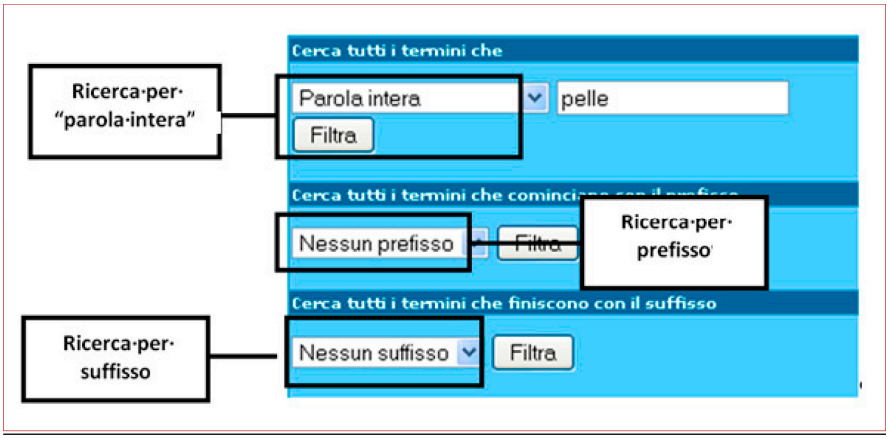

Figure 3. Searching terms within ICD-9-CM.

– "**whole word**": results are all ICD9-CM Plus terms that contain the full term;
– **by prefix:** result is a list of all the terms beginning with the selected prefix (e.g., iper-, ipo- etc..);
– **by suffix:** result is a list of all the terms ending with the selected suffix (e.g., -ite, -osi, etc.);

The type of relationships that can be established between a UNICO term and a term already in ICD9-CM Plus are:

– **VM** (Morphological Variant): UNICO term vary in form compared to a term already in ICD9-CM Plus;
– **UF** (Used For) - UNICO term is a synonym of a term already in ICD9-CM Plus;
– **BT** (Broader Term) - UNICO term is included as broader term of a term already in ICD9-CM Plus;
– **NT** (Narrower Term) - UNICO term is a subordinate of a term already in ICD9-CM Plus;
– **RT** (Related Term) - UNICO term is an associated term of a term already in ICD9-CM Plus.

If the term has no relationships with ICD9-Plus terms, it can be included as a principal term in ICD9-CM Plus without defining any type of relationship. The final product is exportable in any format for further use, both as a full list of terms and as a thesaurus.

3.1 Final product

To facilitate using the thesaurus, a searching system and a user-friendly display were created, in order to allow navigation among different hierarchical levels and associations previously defined for each term (Figure 4).

In order to evaluate the representativeness of ICD9-CM Plus compared with ICD9-CM, a comparison was made between terms in ICD9-CM Plus and terms extracted from a sample of prescriptions of two hospitals. The results obtained show an increase in the percentage of overlap compared to the values previously obtained only by comparing the terms of ICD9-CM (see Table 1). It confirms validity and effectiveness of the work done.

Sovrapposizioni Campioni Az. Ospedaliere / ICD9-CM Plus		
	Numero Sovrapposizioni	% Lessico contenuta in ICD9 Plus
AOBO (codified prescriptions)	5918	93,46% (+ 2.92%)
AOBO (free prescriptions)	863	2,84% (+ 1,92%)
AOCS (free prescriptions)	342	31,61% (+ 28.46%)
MMG	518	15,14% (+ 9.23%)

Table 1. Percentages obtained after overlap among samples of prescriptions of two hospitals and ICD9-CM Plus.

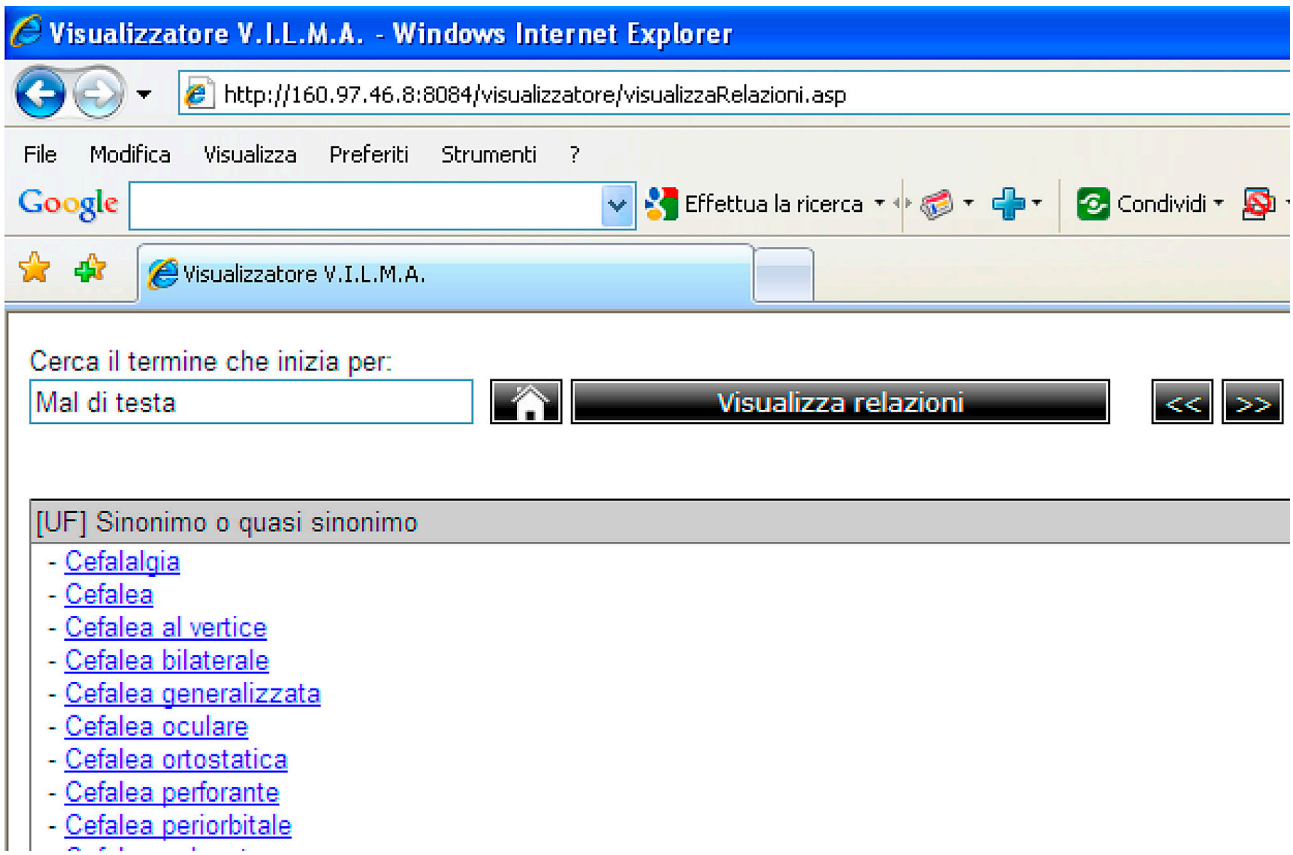

Figure 4. V. I. L. M. A. display

4.0 LOINC: translation and adaptation of the laboratory coding system to the Italian context

A standardized coding system for laboratory observations does not exist in Italy. Codes currently used to identify laboratory observations were created for economical purposes only; therefore, they are not based on a structured classification of the reference domain. On the other hand, each laboratory has developed its own coding system according to its needs, so they all use different types of codes (alphabetic, numeric, or alphanumeric). This highly fragmented situation becomes extremely critical in the frame of the automated management of laboratory data within the national EHR project.

LOINC is an international standard for coding clinical and laboratory observations, and it is the most suited to solve these kinds of issues, because it is highly detailed in test descriptions and recommended for encoding identifiers in electronic observations by HL7 standards. The aim of the system is to encode the observation and not the value that it can assume, so that, if the observation is considered as a question (e.g., what is the patient's hemoglobin?) and the resulting values as a response, it is possible to say that LOINC encodes the question (718-7: Hemoglobin: MCnc: Pt: Bld: Qn), while codes for the answers are provided by other classification systems such as ICD or SNOMED (Systematized Nomenclature of Medicine−Clinical Terms). SNOMED is a systematically organized computer processable collection of medical terminology covering most areas of clinical information such as diseases, findings, procedures, microorganisms, pharmaceuticals etc. It allows a consistent way to index, store, retrieve, and aggregate clinical data across specialties and sites of care. It also helps organizing the content of medical records, reducing the variability in the way data is captured, encoded and used for clinical care of patients and research. The growing importance this coding system is assuming at the international level makes it the ideal candidate to resolve the confused situation in Italy. To this end, different institutions and organizations interested in the health field and, especially, in the definition of a standard for encoding laboratory observations, decided to set up a working group named "LOINC Italia" (http://groups.google.com/group/loinc-italia?hl=it). The basic principle of the group is to share experiences and knowledge to realize the Italian LOINC version, as close as possible to the needs of different stakeholders. Members involved are: CNR-URT Indexing and Classification Systems (Cosenza), CNR-ITB Electronic Health Unit (Roma), University Hospital Molinette (Turin), Arsenàl Consortium (Treviso), Castelfranco Veneto Hospital, ULSS Asolo (TV), Provincial Health Services (Trento), Bruno Kessler Foundation(Trento).

The first step of the work was to translate the LOINC database. It was not a simple translation of the single codes of the Regenstrief database, but a deeper analysis of the Italian context, which allowed the identification of Italian laboratory needs. Thanks to the association with the University Hospital "Molinette" in Turin, it was also possible to start the process of mapping the local codes of eight Region Piemonte laboratories to the LOINC codes. A web application, developed by CNR URT, assisted and supported laboratory technicians in this process. It is important to highlight that mapping operations are not intended to replace local codes with LOINC codes, but only to integrate them into work practices and the information exchange of Italian laboratories, so that they can become unique identifiers as a key to access international databases for laboratory tests results.

4.1 LOINC: Names translation

To facilitate the mapping process of local codes to LOINC and the exchange and sharing of information related to laboratory tests of each patient/citizen, it was necessary to translate the American master of the LOINC database. The translation (Regenstrief Institute 2009) incorporated about 12,000 atomic parts of the so-called compositional database, then was used to generate the entire database in Italian. On one hand, the recommended approach reduced the translation time, but, on the other hand, it raised a number of critical issues related to linguistic differences. To remain as faithful as possible to the original, we tried to adapt and contextualize the individual units to the Italian by using synonyms. The review and validation were done by a group of experts in the biomedical field (experts involved were: Bonofiglio Daniela, Associate Clinical Pathology, University of Calabria; Catalano Stefania, Associate professor in General Pathology, University of Calabria; Domenico Sturino, Researcher in English Language and Translation, Linguistics Department, University of Calabria).

Before starting the translation process, some rules were established:

– Always translate strings of words or terms that could answer to conceptual frameworks and conventions established in LOINC system and that

do not find any direct correspondence in Italian; example:
Formal_Name: A little u super little a
Display_Name: Au ^ a
Your_Translation: A u minuscola apice a
Your_Synonyms: Au ^ a

– Keep Latin in species names, only translate English common names, if present; example:
Formal_Name: Saccharomyces cerevisiae
Display_Name: Baker's yeast
Your_Translation: Saccharomyces cerevisiae
Your_Synonyms: Lievito di birra

– Manage acronyms and abbreviations after creating tables of correspondences, using three different strategies:
1. Explain and translate acronyms if there is an Italian correspondence; example:
Formal_Name: CSF
Display_Name: Cerebral spinal fluid
Your_Translation: LCS
Your_Synonyms: Liquido cerebrospinale
2. Mantain English acronym only if it is widely used in Italy; example:
Formal_Name: HIV
Your_Translation: HIV
3. Keep English form and proceed to clarify it only if it is not widely used in Italy; example:
Formal_Name: Bil fld
Your_Translation: Bile

– Normalize abbreviations referring to administration time of substances that act as a stimulus in the so-called *challenge test*, in order to limit, as much as possible, ambiguous cases; example:
Formal_Name: 1.3H post dose insulin IV
Display_Name: 1.3 hours post dose insulin IV
Your_Translation: 1.3h post dose insulina EV
Your_Synonyms: 1.3 ore dopo dose insulina EV

The translation fields are shown in Figure 5:

– *Formal_Name*: Specify LOINC name of each element;
– *Display_ Name*: Specify the correspondent English synonym;
– *Your_Traslation*: Contains the Italian translation of *Formal_Name*;
– *Your_Synonyms*: Contains the Italian correspondence, if exists, of *Display_ Name* and other possible synonyms in Italian.

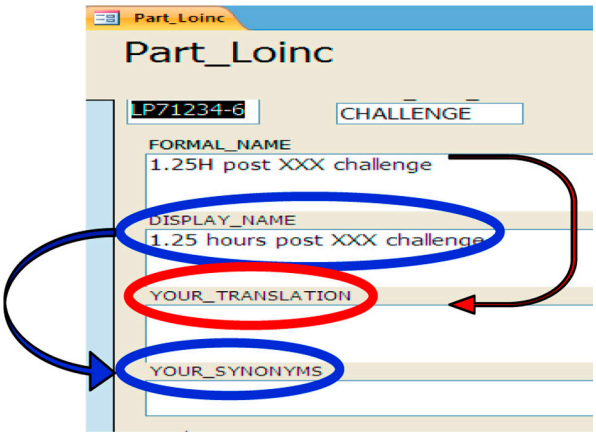

Figure 5. Translation mask.

A certain number of issues related to the linguistic distance between English and Italian were found during the translation work, as mentioned earlier. Here are some examples:

1. Presence of particular types of bacteria and/or viruses of North America. They were translated anyway and maintained in the Italian database, even if not currently tested; example:
Formal_Name: Dermatophagoides Farinae
This is a species that belongs to the family of "House dust mild" typical of areas of North America.
Display_Name: American House Dust Mite
Your_Translation: Dermatophagoides Farinae
Your_Synonyms: Acari della polvere delle case americane

2. The lack of a direct correspondance among two languages; example:
Formal_Name: Cat dander
The English term "*dander*" refers to three different conceptual entities: hair, dandruff and dry skin. In Italian there is no single term to describe these three concepts, the translation of this term "dander" is not entirely correct.
Your_Translation: Forfora di gatto

The translated database was sent to Regenstrief Institute. The first Italian version of LOINC was included in the December 2010 release (LOINC 2.34 and RELMA 5.2); thanks to a further review of the translated items, a more complete and organic Italian translation of the LOINC names was included in the LOINC 2.36 and RELMA 5.3 release in June 2011.

4.2 Italian Loinc Tool

In order to provide technical support to the project, a web application was implemented. It allows mapping operations of local tests to the international standard LOINC and to the specific regional reimbursement list. The Nomenclatore Tariffario Nazionale, a no longer valid national list of codes and standardized names, was embedded in the system and linked through many-to-many (N:N) relations to the regional reimbursement lists, which also were embedded in the system. Furthermore, each laboratory mapped its local test list to the codes of its regional reimbursement list through zero to many (0:N) relations. That is to say, each regional code could be mapped to multiple local tests. The lack of mapping between some of the local tests and the regional codes was solved, embedding into the tool the ability to map towards the regional codes, thus creating an incremental knowledge base exploitable by further users for each new iteration. In this scenario, we established the following mapping rules:

– a regional code could be mapped to more than one local tests (1:N);
– a local tests could be mapped to multiple codes of the regional reimbursement list (N:1) (Figure 6).

In mapping to LOINC codes, a laboratory employee can search in different sources—native English LOINC, the new Italian LOINC version, CUMUL (the Swiss- Italian translation of the standard, which does not always conform in grammar and syntax with the Italian language)—to find out the most suitable LOINC code to map to, viewing also the mapping choices already done by the others. At the end of the mapping process, a scientific committee, including domain experts and translators, will validate the choices made by each laboratory.

5.0 Conclusions and perspectives

The lack of a previous study of ICD adaptability to the needs of Italian physicians has resulted in a lot of difficulties in using it for encoding diagnoses in electronic prescriptions, although what mentioned Technical Rules states. Integrating ICD9-CM with the diagnosis of both a real sample prescription and other biomedical terminological sources improved its representativeness. Nevertheless, results obtained so far are not fully satisfying, because they show some gaps of the classification system, especially about comorbidity reporting and suspected diagnosis. It was thus decided to work more on adaptating ICD9-CM to the Italian scenario, by extracting a new sample of

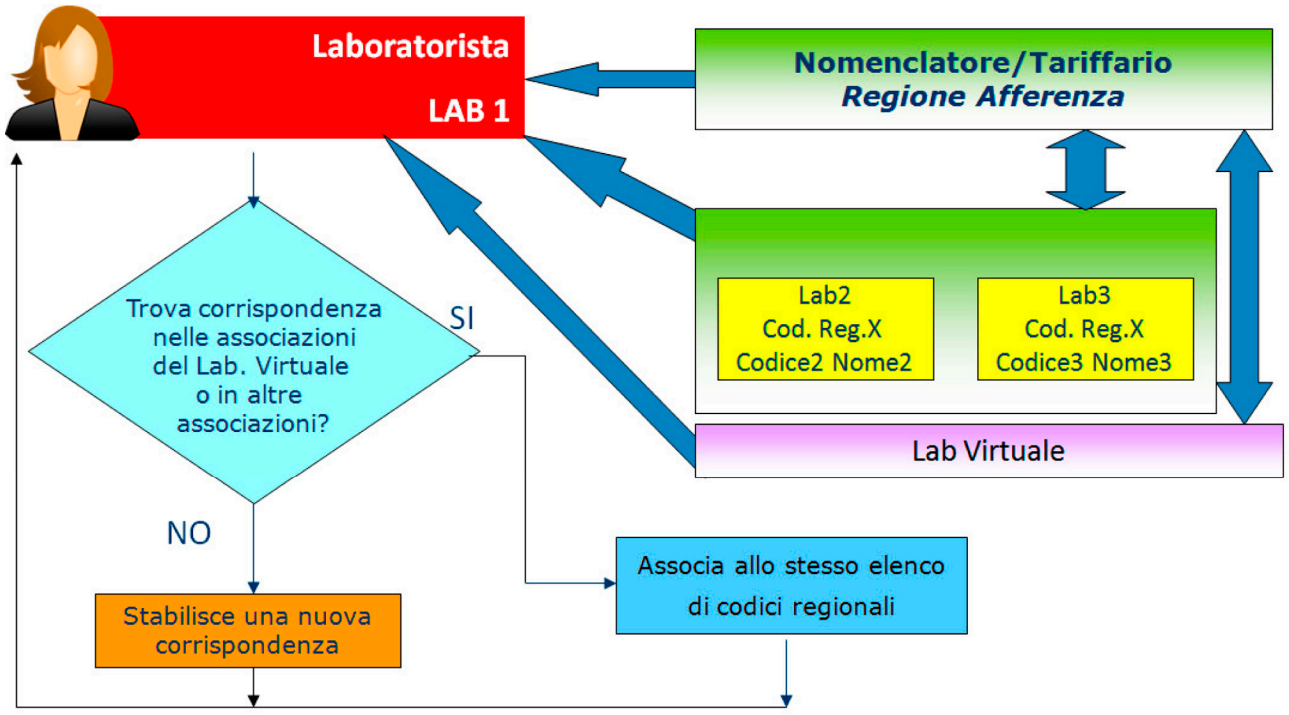

Figure 6. The work flow towards the regional reimbursement lists.

diagnoses according to more precise statistical criteria. It should allow a deeper analysis of the data and a better evaluation of healthcare professionals' needs. The new thesaurus will be tested by Federazione Italiana Medici di Medicina Generale in order to evaluate whether it is suitable to be used in the emerging Italian EHR technological infrastructure.

In relation to the work of creating an Italian standardized coding system for laboratory observations, the Italian LOINC version was first released in December 2010. Subsequently, a revision of all the translated LOINC names resulted in a new Italian LOINC database, syntactically and semantically improved. It was embedded in the official June 2011 Regenstrief Institute releases. Mapping operations, which Region Piemonte is working on, will give us feedback in order to evaluate the extension of the system to the other Italian Regions.

References

Aracri, Giovanna, Branca, Biancaluisa, Chiaravalloti, Maria Teresa, Guaglianone, Maria Teresa, et al. 2011. Verso una standardizzazione della terminologia specialistica di ambito biomedico. In Zanola, Maria Teresa, Bonadonna, Maria Francesca, eds. *Terminologie specialistiche e prodotti terminologici*. Milan: EdUCatt, pp. 87-99.

Cardillo, Elena, Folino, Antonietta, Iozzi, Francesca, Taverniti, Maria and Oliveri, Elisabetta. 2009. Towards a thesaurus for energy efficiency in building construction: the Italian context. On line proceedings of TIA 2009 conference, 8th international conference on Terminologie et Intelligence Artificielle, Toulouse (France), 18-20th November 2009.

Cardillo, Elena, Folino, Antonietta, Guarasci, Roberto and Taverniti, Maria. 2010. GoldThes: a faceted thesaurus for goldsmith handcraftsmanship in a regional context. In Gnoli, Claudio and Mazzocchi, Fulvio eds. *Paradigms and Conceptual Systems in Knowledge Organization, Proceedings of the 11th International ISKO Conference, 23-- -26 February 2010, Rome, Italy.* Advances in knowledge organization v. 12. Würzburg: Ergon Verlag, 2010, pp. 202-07.

Cavarretta, Giuseppe A., Murzaku, Alexander, Guarasci, Roberto, Rovella, Anna R. 2009. La classificazione automatica dei documenti *Archivi* 4, no. 1: 73-80.

Ministero del lavoro, della salute e delle politiche sociali. 2008. *Classificazione, delle malattie, dei traumatismi, degli interventi chirurgici e delle procedure diagnostiche e terapeutiche, italian version of ICD9-CM 'International Classification of Desease 9th revision - Clinical Modification 2007*, Roma: Istituto poligrafico Zecca dello Stato.

Regenstrief Institute, Inc. 2009. *LOINC® translation users' guide*, October 2009.