Towards the Development of a CTS2-based Terminology Service in the Italian Federated Electronic Health Record

Elena Cardillo and Maria Teresa Chiaravalloti

Institute of Informatics and Telematics
National Research Council, IIT - CNR
Rende, Italy
Email: elena.cardillo@iit.cnr.it, maria.chiaravalloti@iit.cnr.it

***Abstract*— Semantic interoperability is essential for advanced Electronic Health Records (EHRs) functionality, and in particular for data exchanges, and efficient communication among clinicians. Integrated terminology services offer the chance to manage clinical code systems, both standard and local, and value sets, through a series of functionalities such as searching, querying, cross mapping, etc. The main standard in the domain is Clinical Terminology Service Release 2 (CTS2) by Health Level 7 (HL7). This paper describes the approach used for designing and developing an integrated terminology service based on the CTS2 standard, namely *Servizio Terminologico Integrato* (STI), which aims to support domain experts and healthcare organizations in ensuring semantic interoperability in the Italian Federated EHR.**

Keywords—Coding Systems; Semantic Interoperability; Terminology Services; CTS2; Healthcare.

I. INTRODUCTION

Interoperability of clinical data essentially means that different systems are able to communicate among each other, exchange data, and, above all, reuse them. The general aim is ensuring a worldwide availability of information at the right time and place, in order to deliver better clinical services and improve healthcare. Interoperability is a required function for the proper use of Electronic Health Record (EHR) systems, which remain simple data containers if they do not have the chance to communicate using the same language. Standardized coding systems are the lingua franca of medical data, as they allow to uniquely identify the same concept despite languages, synonyms and local names that could be used to refer them. The advantages of standards are commonly recognized and their usefulness increases over time as they are employed in numerous health-related Information Technology applications. Nonetheless, their use is not always as easy as it may appear and health professionals often complain about the lack of adequate support systems.

Managing clinical terminologies is not only a matter of making them available to the users, but their management needs to include further functions to offer a complete plethora of services allowing a meaningful use of standards. To pursue this aim, there is the need of a standard protocol to manage terminology standards in the same way across multiple healthcare facilities. This role is covered by integrated terminology services, which offer the possibility to interact with terminologies according to a series of standardized functionalities, such as research, hierarchical tree navigation, structured query, cross mapping.

The Italian National Research Council (CNR) is working in accordance with the Digital Italy Agency (AgID) for realizing the national federate EHR (corresponding to the Italian acronym FSE, meaning "Fascicolo Sanitario Elettronico") infrastructure in order to allow the exchange of clinical documents among the regional EHR systems [1]. In this frame, semantic interoperability is a non-trivial issue [2], especially because, over time, regional and local coding systems and habits have proliferated. The Prime Minister Decree No. 178/2015 [3] disciplines the use of the FSE and makes some medical terminologies mandatory, detailing their use in the two kinds of documents (Patient Summary and Laboratory Report) included in the minimum unit expected in the FSE first implementation phase.

The objective of this paper is to describe the approach used for designing and developing an integrated terminology service, called STI (acronym of the Italian *Servizio Terminologico Integrato*), to support Regions, domain experts and health facilities in the management of the clinical coding systems and terminologies prescribed by the cited Decree, and to ease their use in the documents required in the FSE. To develop this terminology service, the standard protocol HL7 Common Terminology Services Rel. 2 (CTS2) [4] was tested.

The paper is organized as follows. Section II gives an overview of the state of the art on semantic interoperability and the main features of CTS2. Section III describes the material used within the Italian implementation of the STI system and Section IV addresses the content approach. Section V shows some preliminary results. Discussion and conclusions in Section VI close the paper.

II. BACKGROUND

A. *Semantic Interoperability: projects and initiatives*

The adjective *semantic* conveys the deep meaning of interoperability as it overcomes lexical and syntactical issues to deal with the meaning of the exchanged information. The best EHR system would be useless without semantic interoperability, as it could not unambiguously interpret data received from other systems. In fact, Semantic Interoperability in healthcare is defined as "the ability of a

 ISBN: 978-1-61208-600-2

healthcare system to share information and have that information properly interpreted by the receiving system in the same sense as intended by the transmitting system" [5].

Projects and initiatives address the semantic interoperability issue trying to propose effective solutions to solve it. Regarding European Community (EU) initiatives and projects, it is worth mentioning the FP7 project *Semantic Interoperability for Health Network*, whose main aim was the implementation of the necessary infrastructure and governance to allow a sustainable semantic interoperability of clinical and biomedical knowledge at European level [6]. Furthermore, the project EHR4CR [7] dealt with the development of a semantic interoperability service platform, which includes a mediation model for multiple standards integration and harmonization. It was tested in 11 EHR systems of 5 EU Countries. Finally, the Trillium Bridge and Trillium Bridge II projects involve EU Countries and US for the creation of a shared model of an International Patient Summary (IPS), to improve semantic interoperability of e-health systems beyond EU borders [8].

Also, international standards organizations proposed protocols for semantic interoperability. The main one is the CTS2 standard proposed in the Healthcare Service Specification Program (HSSP), a joint HL7 and Object Management Group (OMG) initiative [9]. CTS2 is a cohesive model and specification for representing, accessing, querying, exchanging and updating terminological resources (e.g., Code Systems, Value Sets, Mappings), built on the RESTful (Representational state transfer) Architectural Style. More recently, HL7 proposed the Fast Healthcare Interoperability Resources (FHIR) Specification [10], another standard for exchanging healthcare information electronically, which, compared to the previous HL7 standards, is more consistent and easy to implement, thanks to its built-in extension mechanism to cover the needed content. In fact, specific use cases can be implemented by combining resources together through the use of resource references.

In the literature, different initiatives aimed at developing terminology integration platforms or services were launched. Initial studies and applications focused, for example, on the use of the Unified Medical Language System (UMLS) Metathesaurus, developed by the US National Library of Medicine [11], which includes more than 100 biomedical vocabularies integrated on the basis of a common Semantic Network and mapped among them. Researchers used UMLS to create knowledge-based representation for controlled terminologies of clinical information and to extract and validate semantic relationships. Particularly relevant are also the HETOP terminology service, which includes cross lingual multi-terminological mappings on a semantic basis [12]; and the LexGrid initiative [13], which promotes the use of common terminology models to accommodate multiple vocabulary and ontology distribution formats, as well as the support of multiple data storing for federated vocabulary distribution.

In the last few years, much effort has been spent on the application of the abovementioned CTS2 standard to develop terminology services, as that realized by the Mayo Clinic Informatics, which is the most internationally relevant [9]. D2Refine Workbench platform, for example, aimed at standardizing and harmonizing clinical study data dictionaries [14]. Focused on laboratory catalogues, the experience of the Partners HealthCare System of Boston, applies the CTS2 *Upper Level Class Model* to represent and harmonize the structure of both local laboratory order dictionaries and reference terminologies [15]. Peterson et al. presented, instead, a design user-centered approach, based on the use of Extraction, Transformation and Loading (ETL) procedures in CTS2-based terminology services [16]. The main advantages of this service are: i) adaptability, ii) interoperability, because of the numerous standard vocabularies included, iii) usability, since focused on users' needs. In the wake of these projects, we propose a multi-layers CTS2 implementation that is not only based on ETL procedures, but allows also mapping (and their validation) between local dictionaries and standardized code systems, including also semantic enrichment through external ontological references.

Interesting applications of CTS2 can be found also in the European context, where the main implementation is the Standard Terminology Services (STS) provided by the French non-profit development standards and services organization PHAST [17]. Other implementations are the following: the Austrian national patient health record ELGA, where, all relevant clinical terminologies are provided through a CTS2-conformant terminology server [18]; and the Terminology Server, realized by the University of Applied Science of Dortmund, which also offers a collaboration environment to develop terminologies in a team [19]. Finally, in Italy, the existing implementations of CTS2-based terminology services are proprietary solutions, i.e., the Distributed Terminology Assets Management system (DITAM) [20]; and the HQuantum [21], which is especially focused on the management and integration of local laboratory data through the LOINC standard. These two solutions were evaluated as non-fitting for the purpose of our project because they are subject to license, while the FSE project required an open and reusable solution. Furthermore, they were, at that time, only partially developed and tested, and this would have implied a lot of customization effort.

B. CTS2 HL7 overview

As stated in the ANSI/HL7 V3 CTS R2-2015 standard [4], "the HL7 Common Terminology Services (HL7 CTS) is an API specification that is intended to describe the basic functionalities that will be needed by HL7 Version 3 software implementations to query and access terminological content. It is specified as an Application Programming Interface (API) rather than a set of data structures to enable a wide variety of terminological content to be integrated within the HL7 Version 3 messaging framework without the need

 ISBN: 978-1-61208-600-2

for significant migration or rewrite". The standard, currently, consists of:

- CTS2 Normative Edition v1.0 and the Service Functional Model (SFM), that serve as Functional Requirements Documents, defining the capabilities, responsibilities, inputs, outputs, expected behavior and a set of core functionalities to support the management, maintenance, and interaction with ontologies and medical vocabulary systems.
- CTS2 Technical Specification, that serves as a technical specification document to define the precise API interface specifications for CTS2 implementation compliance in Simple Object Access Protocol (SOAP).

The CTS2 Information Model specifies the structural definition, attributes and associations of Resources common to structured terminologies such as Code Systems, Binding Domains and Value Sets. The Computational Model specifies the service descriptions and interfaces needed to access and maintain structured terminologies. The main CTS2 profiles and functionalities are:

a) *Search/Query Profile*, including: reading of a resource, a code or a concept; browsing or visualization of the tree of a resource; the download of a resource.
b) *Terminology Administration Profile*, including administration functionalities: import of a resource; creation of mappings between the imported resources; the possibility to use updates and notifications.
c) *Terminology Authoring Profile*, including "read-write" functionalities intended for an application used by specialized users (e.g., translators) to create and maintain terminological resources.

More specifically:

- *Read* – the direct access to the resource content via URI, local identifier or a combination of an abstract resource identifier and version tag (e.g., LOINC/Current version).
- *Query* – the ability to access, combine and filter lists of resources based on their content and user context.
- *Import and Export* – the ability to import external content and/or export the contents of the service in different formats.
- *Update* – the ability to validate load sets of changes into the service that updates its content.
- *History* – the ability to determine what changes occurred over stated periods of time.
- *Maintenance* – the ability to create and commit sets of changes.
- *Temporal* – the ability to query on the state of the service at a given point in the past (or in the future).
- *Specialized* – service specific functions such as the association reasoning services, the map entry services and the resolved value set services.

The CTS2 Development Framework (DF) is a development kit for rapidly creating CTS2 compliant applications. It allows users to create plugins, which may be loaded into the DF to provide REST Web Services that use CTS2 compliant paths and model objects. Since it is plugin based, users are only required to implement the functionality that is exclusive to their environment. Thus, CTS2 DF provides all the infrastructures and utilities to help users create plugins. Given the short time available to develop the service, in this work, we reused the mentioned CTS2 DF toolkit provided by Mayo Clinic Informatics [9], which is useful for rapidly creating CTS2 compliant applications, and, at the moment, it is recognized as the most complete and documented. Furthermore, the community that uses the DF is wide and quite reactive.

III. MATERIAL

The terminology service was designed taking into account both the CTS2 main functionalities requirements and the structures of the medical coding systems required by the law (described in Table I).

TABLE I CODE SYSTEMS REQUIRED IN FSE

Code Systems	**Maintaining Organizations**	**Use in FSE**
Internatioanl Classification of Disease 9th Revision, Clinical Modifications (ICD-9-CM)	World Health Organization	Required in the Patient Summary for coding relevant and chronic diseases, in Prescriptions and in Discharge Letters for coding Diagnoses
Anatomical Therapeutic Chemical Classification (ATC)	WHO Collaborating Centre for Drug Statistics Methodology Norwegian Institute of Public Health	Required in the Patient Summary for coding adverse reactions to food and medication, medication plan, and vaccinations
Autorizzazione all'Immissione in Commercio (AIC)	Italian Medicines Agency	Required in the Patient Summary for coding medications
Logical Observation Identifiers Names and Codes (LOINC)	Regenstrief Institute	Required in the Laboratory Report for coding performed tests and their specialty or class

As each resource has a different structure, the most suitable solution was integrating them into the STI Knowledge Base (KB) allowing the correct visualization and searching into each of them. ICD-9-CM, for example, is a classification, which has a hierarchical tree structure so in the visualization it needed the use of indentations and expandable/collapsible branches for navigating the tree; LOINC, instead, is more like a nomenclature, without any hierarchical structure (codes are progressive and not informative). Furthermore, each LOINC code has associated numerous information to be visualized, which are discriminative in choosing a code rather than another, so it was necessary to realize a personalized form to access LOINC code details (e.g., system, scale, method, etc.).

As further explained in Section IV.B, a great deal of effort was spent on the collection of the different versions of each standard and on re-structuring the available files according to the CTS2 concept model, to integrate them in the KB. Other than the standard terminologies required by the law, some other resources, such as value sets and local files mapped to the code systems, were integrated into STI. The first type includes files of synonyms of LOINC and ICD-9-CM terms, which could be used as further research items to find a specific code. Local files mapped to the code systems are useful in STI as both basis of comparison for who is working on the same type of mapping and collector of local synonyms of the official terms used in the standard clinical terminologies. At the moment, a file of LOINC mapped local tests of some laboratories of the Umbria Region is in service.

IV. Approach

The proposed solution consists of a standard-based and web-based distributed software infrastructure, which is open and extendable, and aims to support the production, integration, maintenance, and use of the terminological resources according to the CTS2 protocol. To design and develop the terminology service, an *Agile* methodology was applied. This led to an iterative development of the system functionalities, starting from the core ones and continuing with further iterations in the process of analysis and development. Each iteration and progress in the design and development of the functionalities were submitted to tests by terminology and domain experts.

A. STI Architecture

The STI Architecture (see Figure 1), was designed and realized by using Full Open Source integrated components:

- The Content Management System Liferay 6.2. CE [22] as environment to create the Web Application. It manages simultaneous user accesses, content versioning and classification. The platform functionalities were realized through the development of appropriate portlet allowing the management of: i) search and visualization of code systems; ii) administration management of import and elimination of code systems.
- Kettle (Pentaho Data Integration) [23], used to realize ETL procedures for data integration during data migration from different Database Management Systems. ETL procedures include: (i) heterogeneous data aggregation; (ii) data transport and transformation, by performing data cleaning operations, or scheduled-based data storing, on the destination database. ETL procedures are mostly used in the construction and population of the KB.
- Virtuoso Open Source Edition [24], developed by OpenLink, used for the management of ontologies and data in RDF. RDF data can be queried through SPARQL endpoints, to facilitate the connection with structured dataset derived from other sources.
- CKAN [25], for the management and publication of Open Data. This open source software allows cataloging datasets and describe them across a range of metadata that, on the one hand, help users to navigate through information, and on the other hand, facilitate indexing of the same datasets on search engines. In the present work CKAN is useful to export data (i.e., resources in STI) in the Open Data format and to publish them on open data platforms.

The strengths of this architecture and implementation are various: all the components are open source; it is scalable, modular and easy to maintain; it is installable on open environment without the need for a license.

B. The STI Knowledge Base

The implementation of the STI KB started with the integration of the basic elements, represented by the code systems. They were processed by ETL procedures, in order to enrich them with knowledge derived from external services. The modeling of the basic entities contained in these code systems was made through Porting on the Database. The definition of the STI KB was based on four application layers: 1) the Data layer; 2) the Integration layer; 3) the Semantic layer; and 4) the Presentation layer.

1) The Data layer: it is represented by the CTS2 Conceptual Model, for the representation of the different types of resources and semantic relationships, and by a relational DB, containing the useful facilities to integrate the resources, in particular code systems, in compliance with CTS2. Each concept of the code systems represents the basic entity that composes the atomic information of the conceptual model and is classified according to the structure defined in the HL7 standard, in XML format. In the STI KB, the code systems described in Section III are included in different versions after a readaptation of their structure to the CTS2 model, but maintaining, at the same time, their specifications. In particular, the collected versions are:

- ICD-9-CM Italian 2007 version, counting more than 16,000 codes. Since the official CSV file distributed on the Ministry of Health website is incomplete (it includes only code/description pairs), it was necessary to integrate it. To this aim, we reused an ontological version built in another project, where each ICD-9-CM code has different attributes: i) description, ii) alternative descriptions, i.e., synonyms, iii) inclusions, iv) exclusions, v) information on primary diagnoses, vi) information on additional diagnoses, vii) further notes.
- LOINC Italian and English versions 2.34 (required by the cited Decree, which counts more than 43,152 terms in the Italian version), 2.52, 2.54, 2.56 and 2.58 (which includes the latest updates of the system, released in December 2016, see Table II for details about LOINC terms). In order to correctly integrate LOINC in the STI KB, we needed to upload, for each version, three CSV files:

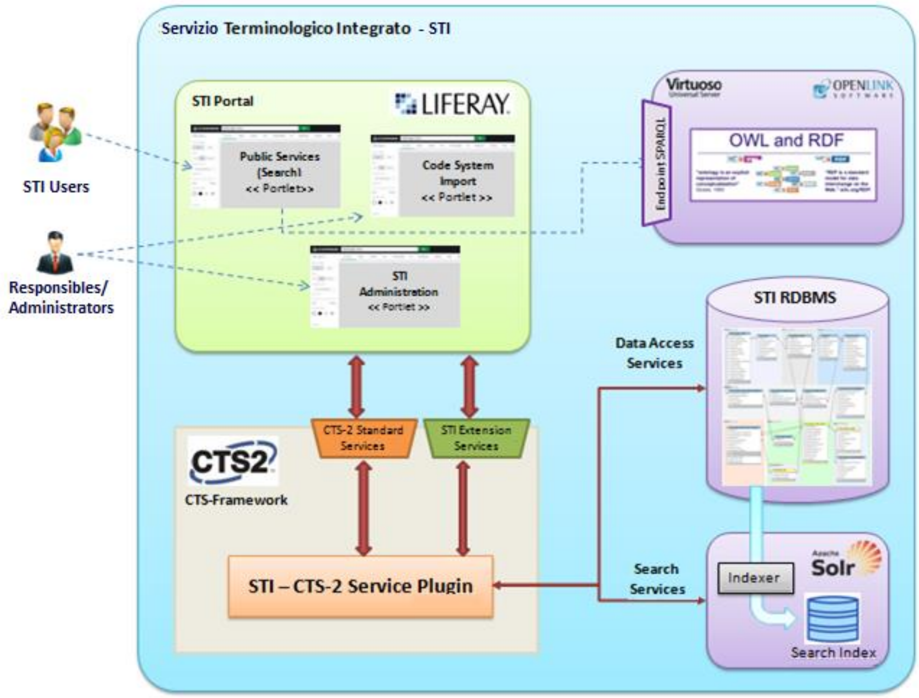

Figure 1. STI Architecture – Deployment Diagram

i) the Italian DB, LOINC_IT (which has limited number of fields associated to each code, with respect to the English version); ii) the English database, *LOINC_DB* (whose structure and fields changed several times during the updates), and iii) the file including the changes of the mapping codes from one version to another, named *Map_to*. For the LOINC_DB, it was necessary to make all the versions compliant with the structure of the last updates (v. 2.58) and to align the CSV of the Italian version to the same structure.

- AIC January 2017 version (the latest available updates on the AIFA website [26] at that moment, including more than 18,000 medicines codes). More specifically, AIC related files are published on the AIFA website as separate files according to the type of drugs. In particular, there are four different csv files: i) *equivalent_medicines* file, which includes for each AIC code the mapping to the active ingredient and thus the corresponding ATC code; ii) *Class_A_medicines*, including essential medicines and those for chronic illnesses (some of them can appear also in the equivalent medicines file); iii) *Class_H_medicines*, including medicines used only in hospital facilities, which can therefore not be sold in public pharmacies (some of them can appear also in the equivalent medicines file); and iv) *Class_C_medicines*, including drugs that are not licensed by the National Health Service (namely SSN) and are therefore paid by the citizen (all AIC codes in this file are mapped to the corresponding ATC code). These four files were separately integrated into the STI KB and ETL processes were used to clean and normalize data, in order to avoid concepts overlap.
- ATC Italian 2014 version (the latest one freely available at that moment), which counts about 5,000 codes. As for ICD-9-CM, access and navigation of the ATC classification tree was provided.

As AIC and ATC cover the same semantic area, a cross mapping file is constantly updated and available, but not in the form of a unique CSV file with 1-1 mappings. As mentioned above, in fact, only two of the four collected files contain mappings to ATC. In order to create a complete 1-1 mappings file, ETL processes were trained to perform mapping extraction processes from the data into the KB. In particular, where an explicit mapping in the different CSV files was not available, it was created by matching the active ingredient of the medicine and by querying external resources (i.e., the AIFA drugs database) in case of ambiguities (i.e., multiple AIC codes associated to one active ingredient), the status changes in AIC, and ATC code updates (considered that we did not use in our KB the latest version of the system). The final mapping file establishes mappings from one ATC code to multiple AIC codes, in fact AIC uniquely identifies branded medications while ATC encodes the medication active ingredient. This file was stored as a Mapping resource into the STI in order to give the chance to have a cross reference between the two code systems. In the Semantic layer, the basic information

 ISBN: 978-1-61208-600-2

included in the code systems is enriched by semantic content and correlations derived from the Integration layer.

Table II gives some statistics about the content integrated in STI Knowledge Base.

TABLE II STI KB STATISTICS

Resources	*Version*	*N. of Concepts (En)*	*N. of Concepts (It)*
LOINC	2.34	+ 60,000	43,152
LOINC	2.52	+ 72,000	58,045
LOINC	2.54	+ 73,000	61,419
LOINC	2.56	~ 80,000	60,837
LOINC	2.58	+ 80,000	63,367
ICD-9-CM	2007	16,100	16,100
ATC	2014	-	5,530
AIC	January 2017	-	18,309
LOINC 2.54 – Umbria_laboratory tests Catalogue Mapping	2016	-	111
AIC – ATC Mapping	2017	-	18,309
Total		+ 400,000	345,179

As can be seen, not all the code systems are mapped, except for *AIC - ATC*, and the *Umbria laboratory catalogue - LOINC*. Since we have not found official and available mappings between the other resources, we provided, as explained in Section V, a functionality to edit mappings directly on the STI platform, under specific permissions and subjected to validation by special users.

2) The Integration layer: it is used for the phase of design and modelling of the data transformation process, semantic enrichment by means of external endpoints, and for the internal organization of data in the KB. To this aim, we used the Kettle component and ETL procedures. In STI, ETL is a key process to bring all the data derived from code systems, which are heterogeneous in their structures and formats, in a standard, homogeneous environment. In particular, during the transformation phase, imported code systems are manipulated to be compatible with the target system (CTS2 model). In some cases, the necessary transformation rules are trivial, but in other situations (as happened for AIC and LOINC files, which changed database structures over the different updates) it may be necessary to sort, unite, and aggregate data. Pentaho Kettle provides a wizard that guides in the migration process, defining the source database server, the destination one, the mapping of the data types, and so much, so that migration does not cause data loss.

The population of the KB can be:

- manual, through the Web Application interface, by means of the compilation of specific forms and the selection of the relationships and classifications in the KB;
- through Rest services. The system will allow to import resources within the KB;
- semi-automatic, through the use of specific ETL systems that guarantee the information extraction and enrichment by querying external services, and finally adding the data to modelled knowledge.

3) The Semantic layer: it is based on the use of ontologies to extend data related to the atomic units (concepts in STI) with external components that have the related knowledge (e.g., the ontologies related to LOINC or ICD-9-CM concepts available in Bioportal [27]). To this aim, the Virtuoso platform was used. In particular, for each LOINC, ATC, or ICD-9-CM concept, it is possibile to query the Virtuoso platform in order to retrieve additional semantic/ontological information (e.g. the semantic type class of the LOINC concept *Hemoglobin A*, code "45208-6" is *Amino Acid, Peptide, or Protein*, or for example the LOINC concept *Aciclovir*, code "1-8", is a pharmaceutical substance whose semantic type is *Nucleic Acid, Nucleoside, or Nucleotide*).

4) The Presentation layer: it is the interface that uses the KB, characterized by the conceptual entities and enriched by a series of information. In particular, each concept in the KB was enriched by the following information provided by using different panels in the interface:

- code details, including information derived from the structure of the code system;
- status and versioning, including information on the changes of the status of a code and on the different versions available in the system;
- relationships, including the relationships of a precise concept to other concepts in the code systems (e.g., the hierarchical relationships of the three digit ICD-9-CM code 282 *Hereditary hemolytic anemias* to its leaf codes, etc.);
- mapping, where all the mappings of the selected code/concept to other resources available in the STI KB are visualized, if present;
- HL7 Specifications, including information needed to exchange data according to HL7 standard, i.e., Code, Code System OID, Code System Name, Code System Version, and concept Display Name;
- ontology, which gives access to the components LodView, to visualize the RDF data of a concept, and LodLive, to navigate the graph of a concept in the ontology derived from Bioportal.

An important aspect of the STI KB is that all the resources, where applicable, were imported in bilingual versions. In particular, LOINC, ATC and ICD-9-CM are available in English and in their official Italian version. The Italian version is in most cases aligned to the corresponding English one. Exceptions are LOINC, where the Italian translation, as all other international LOINC translations, is always aligned to the previous LOINC English version; and the Italian translation of ATC, available in the system as version 2014, since it was not possible to collect the last Italian updates by the responsible government agencies.

C. Web services development

Considering the CTS2 functionalities described in Section II.A, we selected and implemented the following services in STI:

- *Reading*: reads the list of resources in STI and shows complete information on a single resource;
- *Search*: allows to search the resources for keywords or in particular fields thanks to the application of personalized filters;
- *Import*: allows to import the resources into the KB; the dataset within the CKAN component; rdf/owl graphs into the Virtuoso triple store;
- *Export*: allows to export complete resources in CSV or JSON formats;
- *Update*: allows the editing of the KB content;
- *Mapping*: allows the visualization of the existing mappings or the editing of new cross-mappings between the resources in STI;
- *Editorial workflow*: allows the approval of a particular resource or change in the KB, as well as the validation of new mappings created by users with special permissions.

The listed services are to be used through specific Rest Services useful for the reuse of the STI functionalities and for the interoperability with other systems. They will allow interoperability between the systems Liferay-CKAN-Virtuoso. For the development of these services, the Spring Web MVC framework was used. It provides Model-View-Controller (MVC) architecture and ready components that can be used to develop flexible and loosely coupled Web applications. The MVC pattern results in separating the different aspects of the application (input logic, business logic, and UI logic), while providing a loose coupling between these elements.

In a Linked Data perspective, STI service allows semantic enrichment of a resource, thus obtaining relationships with related resources, by querying exposed SPARQL endpoint, as the Bioportal ones.

D. Web Application Development

Regarding the Web Application, the system covers the following functionalities:

- User registration, authentication, roles and permissions management.
- CMS management (platform browsing management, content and versioning management, etc.).
- Multilingualism/bilingualism management (possibility to switch from Italian to English language when browsing a resource).
- Resource utilization by means of the Reading Web service.
- Search of one or more resources or concepts by means of the Search Web service.
- Resource and workflow management.
- Import and Export.
- Mapping between resources.
- Browsing of ontological resources by means of LodLive and LodView, linked to the Virtuoso SPARQL endpoint.
- Use of SPARQL endpoint for the resources stored in Virtuoso. Virtuoso will expose the resources in the RDF/TTL format via the SPARQL endpoint, allowing users to make more sophisticated queries.
- Management of the STI dataset imported in CKAN.

The Web Application allows to navigate the available resources according to the type (i.e., Code System, Value Set, Mapping). After the selection of a specific resource, depending on the original structure, it is possible to navigate the hierarchical tree, to directly select a code and visualize its details; or to search on the selected code system by using filters or full text search. The interface of the navigation functionality was built taking inspiration by the cited *Terminology Server*, the CTS2 implementation provided by the abovementioned University of Applied Sciences and Arts Dortmund.

V. RESULTS AND EVALUATION

STI service was released in its beta version in April 2017, as both Web service and Web application. It contains four standard code systems, which are those prescribed by the Law Decree regarding FSE, in their Italian and English versions, and also some mapping resources (local mapping to LOINC and AIC-ATC mappings). They can be accessed through the CTS2 main functionalities, such as searching, querying, navigation. Versions available are both those fixed by the cited law (i.e., LOINC 2.34 – December 2010) and the most recent ones (i.e., LOINC 2.58 – December 2016), so users can choose which one best fits their needs.

The service is open to the possibility of uploading additional code systems, mapping and value sets. They will be integrated, as it was for the four standards already available in the STI, taking into account their peculiar structure so to ensure a proper use of them. Furthermore, more local files mapped to the standard code systems can be uploaded by the system administrators after validation of their correctness. Regarding the mapping, there is also the chance, for users with special permissions (e.g. physicians, laboratory technicians, etc.), to create mappings between the available resources directly through the STI platform by using the *Cross-Mapping* functionality. During the cross-mapping, users have to qualify the mapping that they are creating between two concepts belonging to two different code systems, by selecting the type of association between the two selected concepts (e.g., choosing if two concepts are synonyms, clinically correlated, or if one is the hypernym of the other, etc. These cross-mappings, in any case, will be validated by the system administrators before becoming effective and saved in the STI KB. Regarding the interoperability services, as said in Section IV, STI allows

 ISBN: 978-1-61208-600-2

external applications, e.g., other terminology services installed at a regional level, to make requests to the Web service, which are those provided by the CTS2. In particular, the following examples are given:

1. **Entity Description Query Service**

Example: Search the entity *Immunoglobulina* in the code systems ICD9-CM:

- http://sti.iit.cnr.it/cts2framework/entities?matchvalue=immunoglobulina&page=0&maxtoreturn=20&codesystem=ICD9-CM

Parameters:
* matchvalue= a string for fulltext search or a query in Lucene syntax
* page= page number (starting from 0)
* maxtoreturn= number of elements per page
* codesystem= code system to query (mandatory)
* codesystemversion= code system version to query (optional)
* format= required format (e.g., "json")

2. **Code System Version**

Example: Entity *Immunoglobulina* in LOINC version 2.56:

- http://sti.iit.cnr.it/cts2framework/codesystem/LOINC/version/2.56/entities?matchvalue=immunoglobulina&page=0&maxtoreturn=20&format=json

3. **Entity Description Read Service**

Example: Read the detailed information of AIC code 19227038:

- http://sti.iit.cnr.it/cts2framework/codesystem/AIC/version/16.01.2017/entity/AIC:19227038

4. **Association Query Service**

Example 1: Existing cross-mapping associated to the ATC v. 2014 code "B02AA01".

- http://sti.iit.cnr.it/cts2framework/associations?list=true&codesystemversion=2014&sourceortargetentity=B02AA01&format=json

5. **Entity Description Query Service**

Example: List of LOINC codes (version 2.54) mapped to a local code system (e.g., Umbria Region):

- http://sti.iit.cnr.it/cts2framework/codesystem/LOINC/version/2.54/entities?page=0&maxtoreturn=250&matchvalue=LOCAL_CODE_LIST:Umbria&format=json

6. **Export Service**

Example: Export of AIC csv format, version January 2017:

- http://sti.iit.cnr.it/cts2framework/exporter?codesystem=AIC:16.01.2017&aictype=classe_h

To test the functionalities and suitability of STI, we recruited a sample of test users, belonging to some of the Italian Regions that already implemented the FSE infrastructure. On one hand, we provided special permissions to Domain Experts (e.g., General Practitioners and Laboratory technicians) in order to let them use both free functionalities (e.g., concept search, navigation of the resources, download) and the Cross-Mapping functionality, to create clinical/semantic mappings directly through STI. On the other hand, we asked regional technical referent users to query the Web service from their local application to make requests such as the ones provided above (e.g., to have the list of all the *map_to* codes in order to verify if some of their mappings changed the LOINC reference code). Figure 2 shows the cross-mapping performed by a Laboratory technician for the concept *Glucosio*.

VI. DISCUSSION AND CONCLUSIONS

This paper described the design and development of a bilingual (Italian – English) integrated terminology service, named STI, based on the CTS2 HL7 standard. The service includes for now the four code systems required by the FSE Law Decree, but it is open to the possibility to integrate further terminologies in the future.

Designing a terminology service is a non-trivial pursuit, especially when resources with different structures need to be integrated and available for different uses. This was the first issue of this work, as it required a personalized design and implementation for each code system uploaded into the STI KB. For example, LOINC has multiple informative axes, which were reported into both the main visualization screen (the six fundamental axes) and an openable window tagged with different labels. Nonetheless, importing LOINC into the service was challenging because its database structure changes as versions evolved. So, a preliminary normalization step was carried out to uniform names and values of the fields of the different versions. Moreover, dealing with AIC, as the system is released in four separate files, ETL procedures needed to be trained for importing each of them every time there is an update and checking if mappings to ATC are present in the new AIC files or if they need to be extracted by following the procedure described in Section IV.B.2). All the above mentioned issues are an obstacle to the flexibility and scalability of the service. Furthermore, it was not always easy finding updated versions of the four code systems, especially in computable format, such as csv files, and for some of them both master English and translated Italian files are not available (i.e., ATC). The chance to visualize ontology representation of the clinical terminologies is not usable for all the versions of the systems. This is an interesting possibility offered by the STI that needs to be improved in the future releases of the service. Efficiency and effectiveness of an EHR also depend on the possibility of unambiguously exchanging and understanding incoming information.

Semantic interoperability improves significantly thanks to the implementation of a terminology service, especially if it is compliant to a standard such as HL7 CTS2, which is widely adopted. The offered services (e.g., searching, querying, and cross mapping) are particularly useful when national or local code systems need to be linked to standard classification systems.

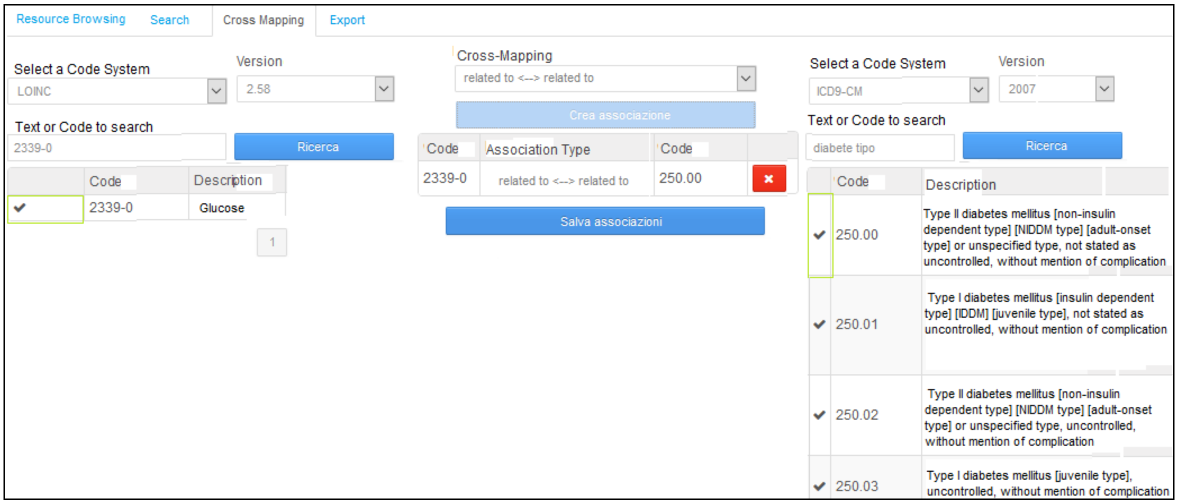

Figure 2. STI Web Application screenshots showing the Cross-mapping between the LOINC code 2339-0 "Glucosio" and ICD-9-CM.

This interoperability also strongly depends on the alignment between terminologies and their quality. This work shows the path that has been taken, also thanks to the recent advancements promoted by the law and by the AgID and CNR collaboration within the context of FSE projects, to align Italian FSE with international initiatives that promote the use of integrated management services of medical terminologies. Nonetheless, it is to be considered that the implementation of integrated terminology services is just the beginning of the process. In fact, the most important aspect in managing medical terminologies is the maintenance over time to update resources and coordinate processes such as transcoding, translation, and licensing. In fact, maintenance of such a system is the real challenge: systems change, errors are made, and the lifecycle of mappings and data must be considered. Sometimes, mappings can be contextual and absolute consistency is very hard to achieve. That evidences the need for a dedicated governmental authority to coordinate the entire process.

Among the several advantages provided by STI use in the Italian FSE frame there are:

- the possibility to share official terminologies, and their updates between the FSE central node and the services used by the local/regional FSE nodes;
- the possibility to configure policies (roles and authorizations) and to model the organization of the system (concerning production/editing of the terminological resources) through terminology management roles;
- the compliance of the data model and application services with the CTS2 standard (Normative Standard CTS2 Version 1.2);
- the delivering services of terminological resources with standard protocols and formats (json, CSV);
- the possibility to make advanced searches with personalized filters according to the code system selected, and to find additional semantic information by navigating their ontological graphs;
- the distribution as open source tool, with a GNU GPL license.

Some of these advantages and functionalities characterize STI if compared to existing CTS2 implementations, especially at national level. In fact, the terminology services cited in Section II.A, even if more sophisticated from a technical and architectural point of view (e.g., in the cited DiTAM service, the possibility to have many local terminology service nodes connected to the central DiTAM node in a federated network), are: proprietary, thus more difficult to be used by a Public Administration; less precise in the structuring and visualization of the code systems; and, to our knowledge, do not allow the access to the resources as Linked Data, or their ontological graphs; and, finally, they do not provide bilingual access as provided in STI.

Possible improvements that could be made on the STI in the future include: i) an extension of the service for including and managing also local code systems; ii) the definition of a general structure for importing and mapping in order to make the service more flexible. The ability to share, query and maintain official and up-to-date terminological artifacts using an accepted standard terminology service interface, such as STI will allow standard terminology content to be readily disseminated and validated, and becomes more useful as organizations (healthcare facilities, Regions, Ministry of Health, and national Standard Development Organizations) in the FSE context begin to undertake the enhancement and maintenance of terminologies to support language translations, jurisdictional extensions to standard code systems, or maintenance and development of local terminologies, avoiding the proliferation of heterogeneous resources, and local tools and technologies to manage terminologies.

Finally, the creation of STI in the context of the Italian FSE, is not only a way to reach semantic interoperability, but it represents a better support to healthcare professionals for improving the quality of clinical data ensuring maximum benefits along the healthcare process and the cooperation among different healthcare providers.

ACKNOWLEDGMENT

This work is supported by the projects *Realizzazione di Servizi della Infrastruttura Nazionale per l'Interoperabilità per il Fascicolo Sanitario Elettronico* and *Realizzazione dei servizi e strumenti a favore delle Pubbliche Amministrazioni per l'attuazione del Fascicolo Sanitario Elettronico* funded by the Agency for Digital Italy. A special thanks goes to LINK Management and Technology for the support in the development of STI.

REFERENCES

[1] M. Ciampi, A. Esposito, R. Guarasci, and G. De Pietro, "Towards Interoperability of EHR Systems: The Case of Italy" ICT4AgeingWell, 2016, pp. 133-138.

[2] E. Cardillo, M. T. Chiaravalloti, and E. Pasceri, "Healthcare Terminology Management and Integration in Italy: Where we are and What we need for Semantic Interoperability" European Journal of Biomedical Informatics, vol. 12 (1): pp.en84-en89, 2016.

[3] Prime Ministerial Decree [Law (general)] n. 178, 29 September 2015, "Regolamento in materia di fascicolo sanitario elettronico. (15G00192)" GU Serie Generale n.263, 11-11-2015.URL: http://www.gazzettaufficiale.it/eli/id/2015/11/11/15G00192/sg [accessed: 09-15-2017].

[4] ANSI/HL7 V3 CTS R2-2015, HL7 Common Terminology Services - Service Functional Model Specification, Release 2, February 2015, URL: https://hssp.wikispaces.com/specs-cts2 [accessed: 09-18-2017].

[5] European Community, "Directive 2011/24/EU of the European Parliament and of the Council of 9th March 2011 on the application of patients' rights in cross-border healthcare" [Available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:088:0045:0065:EN:PDF].

[6] Semantic Interoperability for Health Network FP7-ICT-2011-7, URL: http://www.semantichealthnet.eu/ [accessed: 09-15-2017].

[7] C. Daniel, D. Ouagne, E. Sadou, K. Forsberg, M. Mc Gilchrist, E. Zapletal, N. Paris, S. Hussain, M. Jaulent, D. Kalra, "Cross border semantic interoperability for clinical research: the EHR4CR semantic resources and services," AMIA Jt Summits Transl Sci Proc. 2016; 2016, pp.51–59.

[8] Trillium Bridge II - Reinforcing the Bridges and Scaling up EU/US Cooperation on Patient Summary, H2020-EU.3.1.5. - Methods and data, URL: http://www.trilliumbridge.eu/ [accessed: 09-18-2017].

[9] CTS2 Development Framework, Mayo Clinic Informatics, URL: https://github.com/cts2/cts2-framework [accessed: 09-18-2017].

[10] Fast Healthcare Interoperability Resources (FHIR), Health Level 7 Specification, URL: https://www.hl7.org/fhir/ [accessed: 09-18-2017].

[11] O. Bodenreider, "The Unified Medical Language System (UMLS): integrating biomedical terminology" Nucleic Acids Res: vol. 1; 32(Database issue); pp. D267–D270, 2004 DOI: 10.1093/nar/gkh061.

[12] J. Grosjean, G. Kerdelhué, T. Merabti, and S. J. Darmoni, "The HeTOP: indexing Health resources in a multi-terminology/ontology and cross-lingual world," 13th EAHIL Conference, 2012, ceur-workshop, vol.952, 2012. URL: http://ceur-ws.org/Vol-952/paper_17.pdf [accessed: 09-15-2017].

[13] J. Pathak, H. R. Solbrig, J. D. Buntrock, T. M. Johnson, and C. G. Chute, "LexGrid: A Framework for Representing, Storing, and Querying Biomedical Terminologies from Simple to Sublime" Journal of the American Medical Informatics Association: JAMIA. vol. 16(3), pp.305–315, 2009, DOI:10.1197/jamia.M3006.

[14] D. K. Sharma, H. R. Solbrig, E. Prud'hommeaux, K. Lee, J. Pathak, and G. Jiang, "D2Refine: A Platform for Clinical Research Study Data Element Harmonization and Standardization," AMIA Jt Summits Transl Sci Proc. 2017; 2017, pp. 259–267.

[15] L. Zhou, H. Goldberg, D. Pabbathi, A. Wright, D.S. Goldman, C. Van Putten, A. Barley, and R. A. Rocha, "Terminology modeling for an enterprise laboratory orders catalog", AMIA Annu Symp Proc., 2009, pp.735-739.

[16] K. J. Peterson, G. Jiang, S.M. Brue, and H. Liu, "Leveraging Terminology Services for Extract-Transform-Load Processes: A User-Centered Approach," AMIA Annu Symp Proc. 2017 Feb 10;2016, pp.1010-1019.

[17] Standard Terminology Services (STS), PHAST, URL: http://www.phast-association.fr/accueil_doc_sts/ [accessed: 09-15-2017].

[18] C. Seerainer and S. W. Sabutsch, "eHealth Terminology Management in Austria" Studies in health technology and informatics., vol.228, pp-426-30, 2016.

[19] Terminology Server, University of Applied Sciences and Arts, Dortmund, URL: http://www.wiki.mi.fh-dortmund.de/cts2/index.php?title=Hauptseite [accessed: 09-15-2017].

[20] DiTAM, Codices S.r.l.. 2014. URL: http://www.codices.com/prodotti/ditam.html. [accessed: 09-18-2017].

[21] S. Canepa, S. Roggerone, V. Pupella, R. Gazzarata, and M. Giacomini, "A Semantically Enriched Architecture for an Italian Laboratory Terminology System," IFMBE Proceedings, XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013. Berlin: Springer; vol. 41, 2013, pp. 1314–1317.

[22] Liferay 6.2., URL: https://www.liferay.com/downloads . [accessed: 03-11-2017].

[23] Pentaho Data Integration (ETL) a.k.a Kettle, URL: https://github.com/pentaho/pentaho-kettle . [accessed: 03-11-2017].

[24] Virtuoso, OpenLink Software, URL: https://virtuoso.openlinksw.com/ . [accessed: 03-11-2017]

[25] CKAN, URL: https://ckan.org/ . [accessed: 03-11-2017].

[26] Autorizzazione all'Immissione in Commercio (AIC), Agenzia Italiana del Farmaco. URL: http://www.agenziafarmaco.gov.it/en . [accessed: 09-18-2017].

[27] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, M. A. Musen. "BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications". Nucleic Acids Res. 2011 Jul;39, pp.W541-5. Epub 2011 Jun 14. URL: https://bioportal.bioontology.org/ [accessed: 09-15-2017].