A CTS2 Based Terminology Service for Managing Semantic Interoperability 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 crucial for data and document exchanges between and among Healthcare Information Systems, and to assure efficient communication among the different domain experts involved in the patient process of care. In this context, integrated terminology services offer the chance to manage clinical code systems, both standard and local, as well as value sets, through a series of functionalities such as searching, querying, import/export, cross-mapping, etc. This paper describes the approach used for the implementation of an integrated terminology service, namely *Servizio Terminologico Integrato*, in the Italian Federated Electronic Health Record setting. The service is based on Common Terminology Service Release 2 by Health Level 7, which is the main referenced standard in the domain and has the aim to support domain experts and healthcare organizations in managing medical terminologies for ensuring semantic interoperability. Preliminary tests show promising results and the usefulness of such a service, above all for the retrieval of coding systems mapping and updates. The main output of this paper is the implementation in the Italian healthcare setting of an open source, standard-based and bilingual integrated terminology service to support professionals and stakeholders in sharing, querying and maintaining official and up-to-date terminological artifacts.**

Keywords - Code Systems; Semantic Interoperability; Terminology Services; Common Terminology Service Release 2 (CTS2); Healthcare.

I. INTRODUCTION

Nowadays interoperability is the word most commonly associated to e-health services. It 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 dialogue using the same language. Standardized coding systems are the lingua franca of medical data, as they allow to uniquely identify the same concept regardless of the language, synonyms or 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 for 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 objective of this paper, which is an invited extension of [1], is to describe the approach used for designing and developing an integrated terminology service, called STI (acronym of the Italian *Servizio Terminologico Integrato*), to provide support in the use of the clinical coding systems in the Italian Healthcare setting, in particular to ease the main functionalities that users, including both Regions and health facilities, and domain experts, could need in managing terminologies and codes in the documents required in the national federate EHR infrastructure (corresponding to the Italian acronym FSE, meaning “Fascicolo Sanitario Elettronico”), realized by the Italian National Research Council (CNR) in accordance with the Digital Italy Agency (AgID). This infrastructure allows the exchange of clinical documents among the regional EHR systems [2].

In this framework, semantic interoperability is a non-trivial issue, especially because, over time, regional and local coding systems and habits have proliferated. In [3] authors provide a detailed overview of the main issues for the semantic interoperability implementation in Italy. To provide general guidelines on the use of the FSE, the Prime Minister Decree No. 178/2015 [4] was issued, also making some medical terminologies mandatory, and 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.

To develop the STI terminology service, the standard protocol Common Terminology Services Rel. 2 (CTS2) [5] by Health Level 7 (HL7) 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 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" [6].

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 sustainable semantic interoperability of clinical and biomedical knowledge at the European level [7]. Furthermore, the EHR4CR project [8] 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 [9].

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 [10], where HL7 provided the requirements as a Service Functional Model (SFM), and OMG developed the formal specification of the standard. 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 [11], 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 [12], which includes more than 100 biomedical vocabularies integrated on the basis of a common Semantic Network and mapped among them. Researchers use 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 [13] and the LexGrid initiative [14]. The former includes cross lingual multi-terminological mappings on a semantic basis, while the latter 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. On the contrary, PyMedTermino [15] implements a generic model and has a separate database for each terminology to be included into the service. This allows for better preservation of the original architecture of each terminology and the easy integration of post coordinated terminologies. PyMedTermino supports advanced operations, such as derived queries, on the concepts set, which are not usually feasible into other terminology services. Nonetheless, its limits are: exclusive Python implementation, no OWL ontologies link, low quality of the UMLS mapping. Luna et al. [16] describe the implementation of an interinstitutional and transnational Remote Terminology Service, which uses local interface terminologies (in the form of a thesaurus) mapped to reference terminologies (such as the Systematized Nomenclature of Medicine - Clinical Terms, namely SNOMED-CT) to allow users to enter free text diagnosis and procedures and get back coded content. It is also possible to select preferred reference terminology among the standardized clinical classification systems, as the service uses the SNOMED-CT cross matchings. The service reveals its utility in polysemy, synonyms and acronyms management, but it is only available for Spanish.

In the last few years, much effort has been spent on the application of the abovementioned CTS2 standard to develop terminology services, such as that realized by the Mayo Clinic Informatics, which is the most internationally relevant [10]. D2Refine Workbench platform, for example, aimed at standardizing and harmonizing clinical study data dictionaries [17]. 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 [18]. 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 [19]. The main advantages of this service are: i) adaptability, ii) interoperability, because of the numerous standard vocabularies included, iii) usability, since it is focused on users' needs. In the wake of these projects, we propose a multi-layer CTS2 implementation that is not only based on ETL procedures, but also allows mapping (and their validation) between local dictionaries and standardized code systems, including semantic enrichment through external ontological references.

Interesting applications of CTS2 can also be found 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 [20]. 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 [21]; 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 [22]. Finally, in Italy, the existing implementations of CTS2-based terminology services are proprietary solutions, i.e., the Distributed Terminology Assets Management system (DITAM), a software designed by Codices-Noemalife that offers standard terminology services through a federate architecture [23]; and the HQuantum (a spin-off of the University of Genova) with the Health Terminology Service (HTS) [24], 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 the time of the initial efforts on this activity, 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 [5], "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 for significant migration or rewrite". The standard, currently, consists of:

- CTS2 Normative Edition v1.0 and the Service Functional Model (SFM), which 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, which 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. In particular, in the CTS2 Service Functional Model the following semantic entities are distinguished (see [5] for details):

- *Code Systems* – defined as a collection of codes with associated designations and meanings. To meet the requirements of a code system as defined by HL7, a given code must refer to one and only one meaning within the code system. A code system in CTS2 Functional Model can be of different types:
  - Set of codes – a flat list of code/term pairs;
  - List of terms – a set of words (or tokens), each one associated with an unambiguous definition, if used in a specific context;
  - Thesaurus – a controlled vocabulary, i.e., a (semi-) structured set of words or locutions (structured according to semantic relations) whose aim is to allow and facilitate information retrieval and navigation;
  - Classification system – a set of hierarchically organized categories or classes used to organize and aggregate data in a specific domain of knowledge or for a specific aim;
  - Ontology – a formal representation or a conceptualization of a given domain of reality.
- *Value Sets* – defined as uniquely identifiable sets of valid concept representations (Codes), where any concept representation can be tested to determine whether or not it is a member of the value set. More practically, they are subsets of codes, classes or descriptions of entities selected from one or more code systems and grouped for a specific purpose (e.g., a value set of units of measures, a value set of anatomical parts, etc.).
- *Maps* – a set of formal rules that permits concepts (EntityDescription) transcoding from a starting value set or code system to a second one.
- *Concept Domains* – metasyntactic variables in database models, messages or other representative forms.
- *Entity Description* – the definition of a concept in the context of a particular terminological system.
- *Association* – the semantic statement (subject – predicate – object) that associates entities within the same code systems or in different code systems.

The Computational Model specifies the services descriptions and interfaces needed to access and maintain structured terminologies. The main CTS2 profiles and functionalities are:

a) *Search/Query Profile:* i.e., “read-only” functionalities and availability of the resources to the client. It includes: 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*: i.e., “write” administrative functionalities. It includes: import of a resource; creation of mappings between the imported resources; the possibility to use updates and notifications.
c) *Terminology Authoring Profile*: i.e., “read-write” functionalities. They are intended for an application used by specialized users (e.g., translators, or domain experts) to create and maintain terminological resources.

More specifically:

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

The CTS2 Development Framework is a development kit for rapidly creating CTS2 compliant applications. It allows users to create plugins, which may be loaded into the Development Framework 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 Development Framework 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 Development Framework toolkit provided by Mayo Clinic Informatics [10] available from Github, 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 Development Framework 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. More specifically, the standards to be used are:

- ICD-9-CM (*International Classification of Diseases – 9th revision – Clinical Modifications*) [25], developed by the World Health Organization (WHO): its use is required in the Patient Summary for coding relevant and chronic diseases, in Prescriptions and in Discharge Letters for coding diagnoses;
- ATC (*Anatomical Therapeutic Chemical Classification*) [26], developed and maintained by the WHO Collaborating Centre for Drug Statistics Methodology, Norwegian Institute of Public Health: its use is required in the Patient Summary for coding adverse reactions to food and medication, the medication plan, and vaccinations;
- AIC (*Autorizzazione all’Immissione in Commercio*) [27], developed and maintained by the Italian Medicines Agency (AIFA): as for ATC, its use is required in the Patient Summary for coding adverse reactions to medication, the medication plan, and vaccinations;
- LOINC (*Logical Observation Identifiers Names and Codes*) [28], developed by the Regenstrief Institute Inc. in Indianapolis: its use is required in the Laboratory Report for coding performed tests and their specialty or class. LOINC codes are also used for identifying the clinical document type, when it is structured according to the HL7 Clinical Document Architecture rel.2 (CDA2).

As each resource has a different structure, the most suitable solution was integrating them into the STI Knowledge Base 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 the re-structuring of the available files according to the CTS2 concept model, to integrate them in the STI Knowledge Base. In addition to 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. For the second type, local files mapped to the code systems were included in STI because they are useful 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, the service includes:

- 106 local tests from 6 different laboratories in the Umbria Region mapped to LOINC 2.54;
- 154 local tests from a laboratory in the Campania Region mapped to LOINC 2.34;
- 1071 local tests from 6 different laboratories in the Calabria region mapped to LOINC 2.54.

IV. Approach

The proposed solution consists of a standard-based and web-based distributed software infrastructure, which was required to be designed open and extendable. It 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 (Figure 1) was designed and realized by using Full Open Source integrated components:

- Liferay 6.2. CE [29] 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) [30], 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 Knowledge Base.
- Virtuoso Open Source Edition [31], 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 [32], for the management and publication of Open Data. This open source software allows for cataloging datasets and describes 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.

Strengths of this architecture and implementation are: i) all the components are open source; ii) it is scalable, modular and easy to maintain; iii) it is installable on open environment without the need for a license.

B. The STI Knowledge Base

The implementation of the STI Knowledge Base 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 (i.e., medical concepts) was made through Porting on the Database.

The definition of the STI Knowledge Base was based on four application layers: 1) the Data layer; 2) the Integration layer; 3) the Semantic layer (or Interoperability 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 database, 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 Knowledge Base, the main code systems and terminology standards, described in Section III, are included in different versions after a readaptation of their structure to the CTS2 model, but at the same time, maintaining their specifications. In particular, versions of the code systems included in the STI Knowledge Base are:

- ICD-9-CM Italian 2007 version, counting more than 16,000 codes. Since the official CSV file distributed by the Ministry of Health is incomplete (it includes only the hierarchical structure, thus the codes and the label of the diagnoses), it was necessary to integrate it. To this aim, we reused an ontological version of the system that was built for another project, where each ICD-9-CM code has multiple associated information: i) official description, ii) alternative descriptions, i.e., synonyms, iii) inclusions criteria, iv) exclusions criteria, v) information about the coding of a primary diagnosis vi) information about the coding of additional diagnoses, vii) further notes. In order to provide access to this additional information, a customized form for ICD-9-CM code details was built.

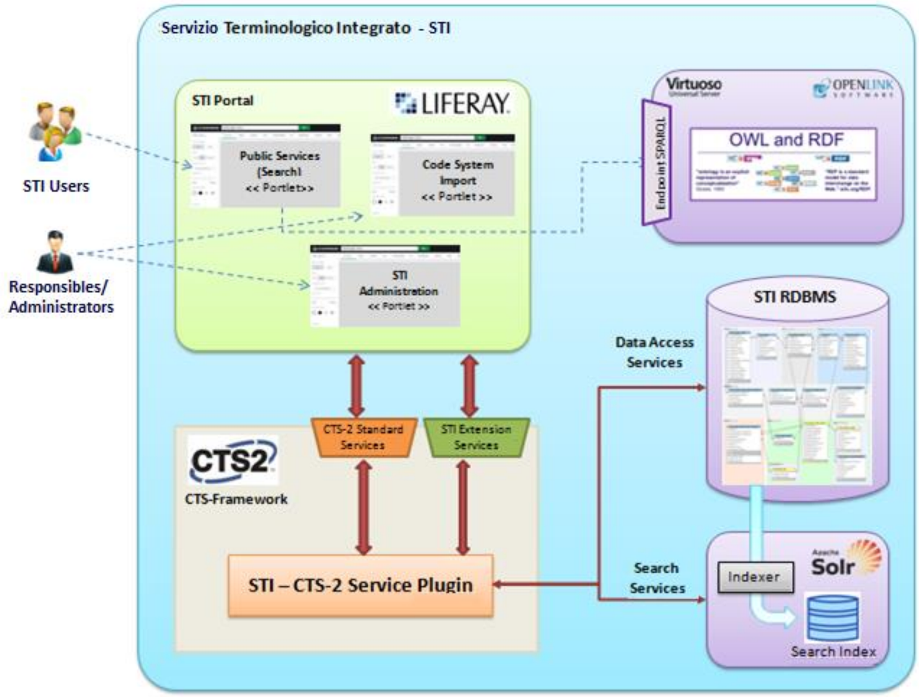

Figure 1. STI Architecture – Deployment Diagram

- LOINC Italian and English versions 2.34 (required by the cited Decree), 2.52, 2.54, 2.56, 2.58, 2.61 and 2.63 (it includes the latest updates of the system, released in December 2017). See Table I for details about LOINC terms. In order to correctly integrate LOINC in the STI Knowledge Base, we needed to upload, for each version, three CSV files: i) the Italian DB, LOINC_IT (it 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 over successive versions), and iii) the file including the changes of the mapping codes from one version to another, named *Map_to*. Regarding the LOINC_DB, it was necessary to make all the versions compliant with the structure of the last updates (v. 2.63) 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 [27] at this time, 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 also appear 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 also appear in the equivalent medicines file); and iv) *Class_C_medicines*, including drugs that are not licensed by the Italian National Health Service and are therefore to be paid by patients (all AIC codes in this file are mapped to the corresponding ATC code). These four files were separately integrated into the STI Knowledge Base 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 this time), 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 Knowledge Base. Figure 2 shows the ETL process related to the import of AIC and ATC and the extraction of their mappings.

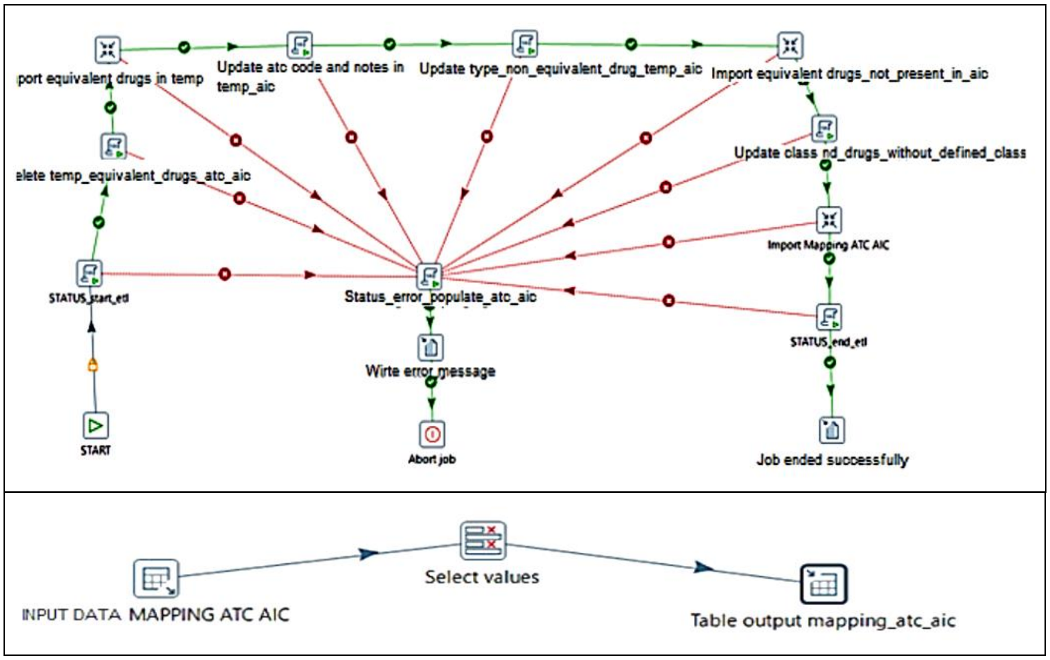

Figure 2. Mappings extraction via ETL: AIC – ATC example

As can be observed, the main job is MAPPING ATC-AIC. From the previous ETL job, expanding the transformation “Import Mapping ATC AIC” we obtain the transformation to import mapping data extracted from the Database (4 files of AIC and ATC). 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 the latest version of the system in our Knowledge Base). 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 included in the code systems is enriched by semantic content and correlations derived from the Integration layer. Table I gives some statistics about the content integrated in STI Knowledge Base. As observable, not all the code systems are mapped, except for the AIC – ATC, and some local catalogues mapped to LOINC. Since we did not find 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 administration users.

TABLE I. STI KNOWLEDGE BASE 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
LOINC	2.61	~ 85,000	63,367
LOINC	2.63	+ 86,500	66,204
ICD-9-CM	2007	16,100	16,100
ATC	2014	-	5,530
AIC	January 2017	-	18,309
Umbria_laboratories tests Catalogues - LOINC 2.54 Mapping	2016	-	106
Campania_laboratory _tests Catalogue - LOINC 2.34 Mapping	2012	-	154
Calabria_laboratories tests Catalogues – LOINC 2.54 Mapping	2016	-	1,071
AIC – ATC Mapping	2017	-	18,309
Total		+ 571,500	475,970

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 Knowledge Base. 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 successive releases) 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 forth, so that migration does not cause data loss.

The population of the Knowledge Base 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 Knowledge Base;
- through Rest services. The system will allow to import resources within the Knowledge Base;
- 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 [33]). 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 Knowledge Base, characterized by the conceptual entities and enriched by a series of information. In particular, each concept in the Knowledge Base 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 Knowledge Base 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 the external resource Bioportal.

An important aspect of the STI Knowledge Base 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 (being them code systems, value sets or mappings) into the Knowledge Base; the dataset within the CKAN component; RDF/OWL graphs into the Virtuoso triple store;
- *Export*: allows to export complete resources or the results of specific queries in CSV or JSON formats;
- *Update*: allows the editing of the Knowledge Base 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 Knowledge Base, as well as the validation of new mappings created by users with special permissions. The

validation can be carried out only by administrative users.

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 among 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, such as the Bioportal one.

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.
- Use of ontological resources and graphs visualization by means of Lodlive.
- Browsing of the RDF resources with 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 the most technical and expert 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. It has to be noticed that when looking for a specific LOINC code, the service returns not only the exact matched code but also all the LOINC codes that have been redirected to it for guiding the users' mapping choices. This is particularly useful as it shows at the same time and for the same concept deprecated or discouraged terms and their correspondent new reference code, facilitating also double check and updates of the already performed mappings. The interface of the navigation functionality was built taking inspiration from the cited *Terminology Server*, the CTS2 implementation provided by the abovementioned University of Applied Sciences and Arts of 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 2.63 – December 2017), 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 Knowledge Base. Regarding the interoperability services, as said in Section IV, STI allows 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 ICD-9-CM and LOINC:

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

Parameters:
* matchvalue= a string for fulltext search or a query in the Lucene syntax for more complex search on indexed fields
* 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: 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

Example 2: List of mappings between the code systems LOINC v. 2.54 and ATC v. 2014

- http://sti.iit.cnr.it/cts2framework/associations?list=true&changesetcontext=LOINC (2.54) - ATC (2014)&format=json

5. Entity Description Query Service

Example: List of mappings between a local code system (e.g., Umbria) and LOINC v. 2.54:

- 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 in JSON format the Italian version of LOINC v. 2.58 and of AIC v. January 2017:

- http://sti.iit.cnr.it/cts2framework/exporter?codesystem=LOINC:2.58&language=it&format=json
- http://sti.iit.cnr.it/cts2framework/exporter?codesystem=AIC:16.01.2017&aictype=classe_h

The developed terminology service also has a FAQ section including the most common questions about code systems used in the FSE setting, questions about the Service and CTS2 itself, and finally, a Documentation section where it is possible to consult or download technical documentation about the service, such as the User Manual, documents related to the STI design and development, and a tutorial explaining its use and functionalities.

In order 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 those 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 3 and Figure 4 show respectively the search and cross-mapping activities performed by a domain expert (more specifically a Laboratory technician) for the concept *Glucose.*

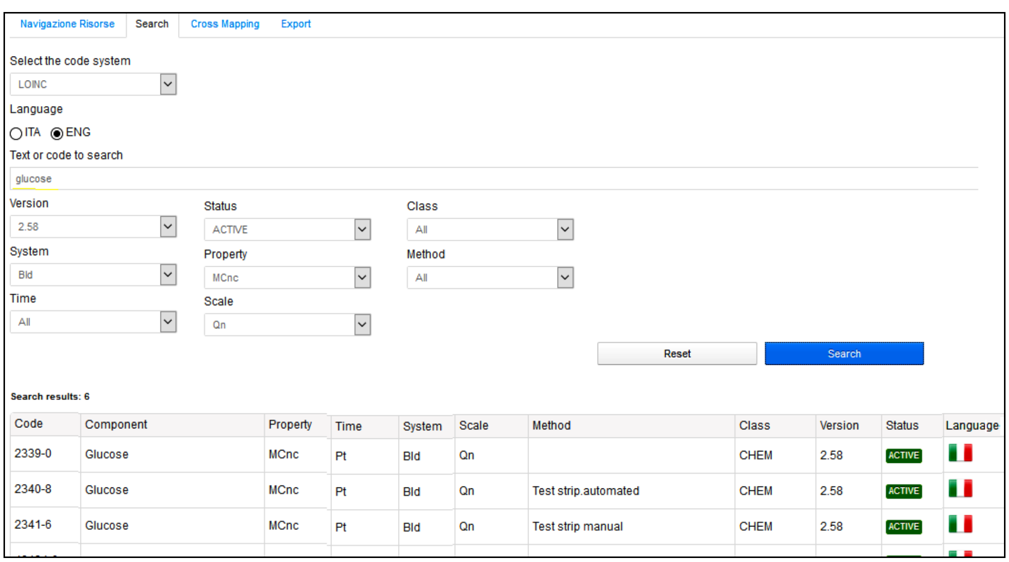

Figure 3. STI Web Application screenshots showing the Search functionality for the concept "Glucose" in LOINC

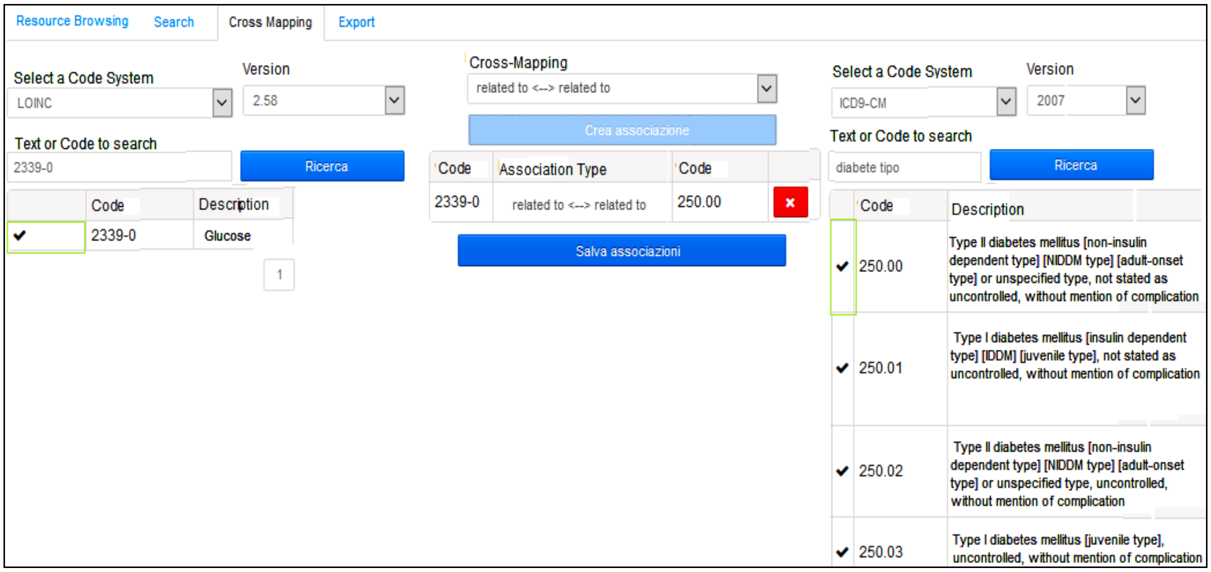

Figure 4. STI Web Application screenshots showing the Cross-mapping between the LOINC code 2339-0 “Glucose” and ICD-9-CM

In a Linked Data perspective, the STI users with appropriate permissions are able to manage graphs and, in the future, will be able to import data in RDF format within the STI Virtuoso platform. To support Virtuoso, OpenRefine is proposed [34], for guaranteeing the connection to ontology models, and thus the connection of the CSV fields in the STI DBMS to the corresponding RDF. Figure 5 shows a scheme that clarifies this workflow. This will permit to access and reuse STI resource in additional and innovative ways:

- through the browsing of RDF resources using LodView, which will be connected to the SPARQL endpoint of STI Virtuoso platform;
- through LodLive, allowing the user to navigate and explore a graph within the Linked Data Cloud, as shown in Figure 6.

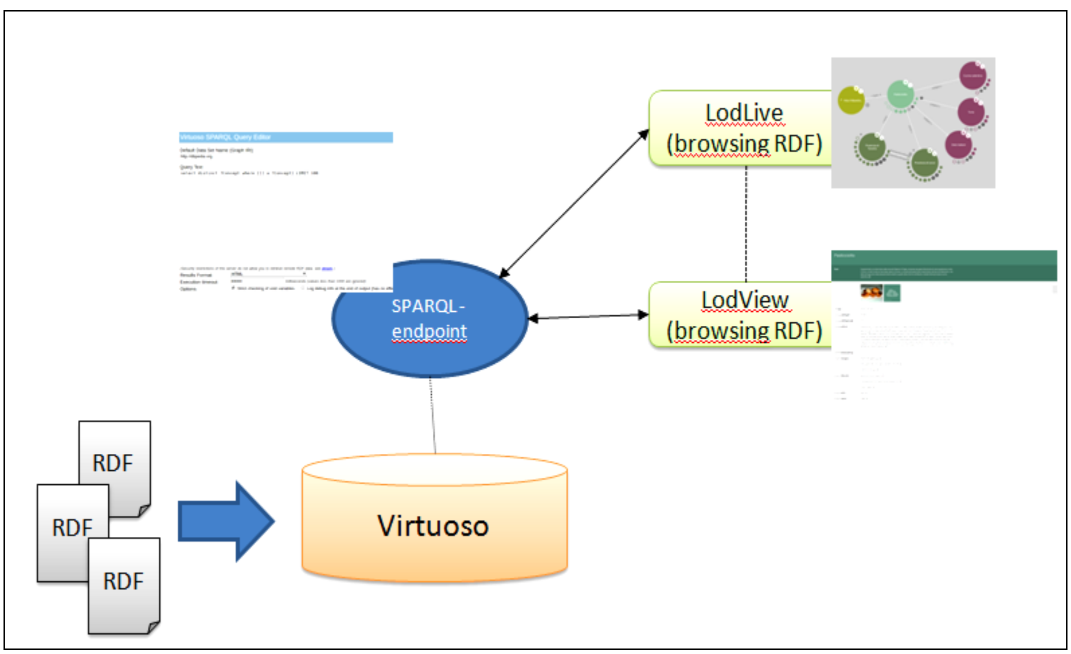

Figure 5. Data import scheme within the STI Virtuoso platform and graphs browsing

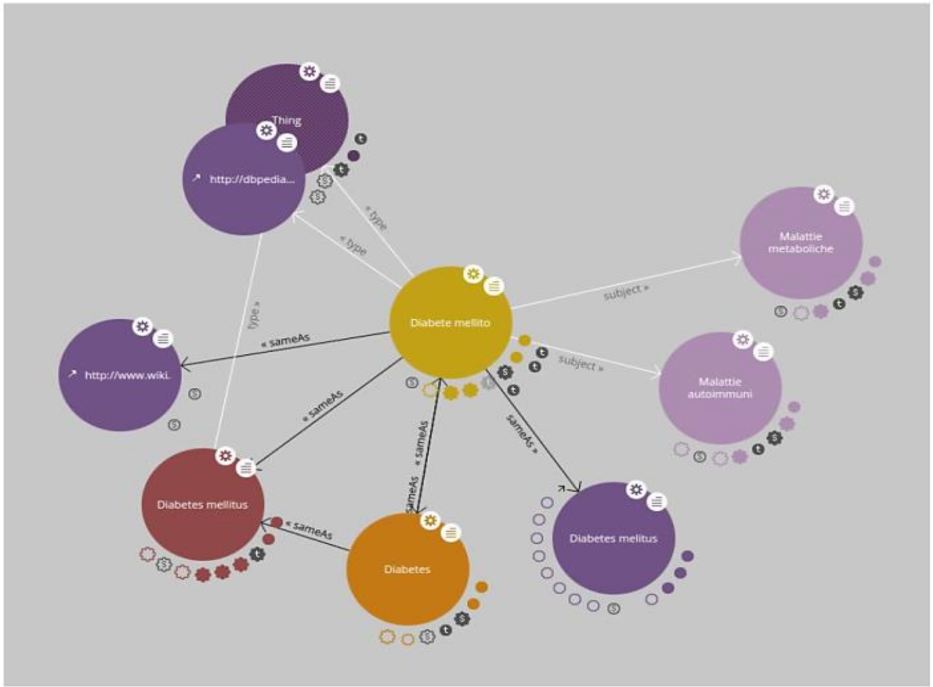

Figure 6. Navigation of the graph related to the concept “Diabete Mellito” in the resources exposed by STI Virtuoso.

VI. Discussion and Conclusions

This paper describes 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 Knowledge Base. 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 evolve. So, a preliminary normalization step was carried out in order to uniform names and values of the fields of the different versions.

Moreover, when dealing with AIC, as the system is released in four separate files, ETL procedures needed to be trained to import each of them every time there is an update, and check 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 services offered (e.g., searching, querying, and cross mapping) are particularly useful when national or local code systems need to be linked to standard classification systems. 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 has 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 that coordinates the entire process.

Among the several advantages provided by STI use in the Italian FSE framework 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 a 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.

Among the improvements currently underway on the STI, we can mention: i) an extension of the service data model, and of the Knowledge Base, in order to guarantee modeling, semantic integration and managing also of other types of code systems used at the national level but also local code systems (regional laboratory catalogues, regional catalogue for prescriptions and services, etc.); ii) the definition of a general structure for importing and mapping in order to make the service more flexible and scalable; iii) the definition of new ETL processes rules for the code systems integration and mapping; iv) the development of new functionalities supporting the maintenance of the system like the CTS2 *History* (i.e., the ability to determine which changes occurred over stated periods of time) and the *Update* services (the ability to validate load sets of changes into the service that updates its content); v) the improvement of the cross-mapping functionality to allow mapping editing by applying new association types; vi) the extension of the export functionality also to query results and mapping tables (currently the service allows exporting available code systems, in their different versions, but does not allow to export extracts of them or query results tables).

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. The use of STI as an open source Terminology Service in the FSE context, whose functioning is directly linked to the level of interoperability and the degree of security of the sensitive data processed among the different Regional systems, gives the chance to Regions and local healthcare facilities to comply with a series of regulations and, in particular, with the modifications to art. 68 of the Legislative Decree 82/2005 - “Digital Administration Code” (CAD) [35], all aimed at giving a preferential road to the use of free software.

Finally, the creation of a terminology management service, such as STI, to be used 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* (prot. IIT 0005723, 01/09/2015) and *Realizzazione dei servizi e strumenti a favore delle Pubbliche Amministrazioni per l’attuazione del Fascicolo Sanitario Elettronico* (agid.AOO-AgID.REGISTRO UFFICIALE (I).0021033.31-12-2015) funded by the Agency for Digital Italy. A special thanks goes to LINK Management and Technology S.p.A. for the support in the development of the STI Web Service.

REFERENCES

[1] E. Cardillo and M. T. Chiaravalloti, “Towards the Development of a CTS2-based Terminology Service in the Italian Federated Electronic Health Record”, SEMAPRO

2017 : The Eleventh International Conference on Advances in Semantic Processing, 2017, pp. 12-21.

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

[3] 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.

[4] 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: 05-15-2018].

[5] 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: 05-15-2018].

[6] 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].

[7] Semantic Interoperability for Health Network FP7-ICT-2011-7, URL: http://www.semantichealthnet.eu/ [05-15-2018].

[8] C. Daniel, D. Ouagne, E. Sadou, K. Forsberg, M. Mc Gilchrist, E. Zapletal, et al., "Cross border semantic interoperability for clinical research: the EHR4CR semantic resources and services", AMIA Jt Summits Transl Sci Proc. 2016, pp. 51–59.

[9] 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: 05-15-2018].

[10] CTS2 Development Framework, Mayo Clinic Informatics, URL: https://github.com/cts2/cts2-framework [accessed: 05-15-2018].

[11] Fast Healthcare Interoperability Resources (FHIR), Health Level 7 Specification, URL: https://www.hl7.org/fhir/ [accessed: 05-15-2018].

[12] 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.

[13] 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: 05-15-2018].

[14] 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.

[15] J. B. Lamy, A. Venot, and C. Duclos, "PyMedTermino: an open-source generic API for advanced terminology services", Studies in Health Technology and Informatics, 2015, vol. 210, pp. 924-928.

[16] D. Luna, G. Lopez, C. Otero, A. Mauro, C. Torres Casanelli, and F. González Bernaldo de Quirós, "Implementation of Interinstitutional and Transnational Remote Terminology Services", AMIA Annu Symp Proc, 2010; pp. 482–486.

[17] 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.

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

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

[20] Standard Terminology Services (STS), PHAST, URL: http://www.phast-association.fr/accueil_doc_sts/ [accessed: 05-15-2018].

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

[22] Terminology Server, University of Applied Sciences and Arts, Dortmund, URL: http://www.wiki.mi.fh-dortmund.de/cts2/index.php?title=Hauptseite [accessed: 05-15-2018].

[23] DiTAM, Codices S.r.l.. 2014. URL: http://www.codices.com/prodotti/ditam.html. [accessed: 05-15-2018].

[24] R. Gazzarata, M. E. Monteverde, E.Vio, C. Saccavini, L. Gubian, I. Borgo, and M. Giacomini, "A Terminology Service compliant to CTS2 to manage semantics within the regional HIE", EJBI – Vol. 13(1), pp. 43–50, 2017.

[25] International Classification of Diseases, 9th Revision Clinical Modification (ICD-9-CM), Italian version, Italian Ministry of Health URL: http://www.salute.gov.it/portale/documentazione/p6_2_2_1.jsp?lingua=italiano&id=2251 [accessed: 05-15-2018].

[26] Anatomical Chemical Therapeutical (ATC), WHO Collaborating Centre for Drug Statistics Methodology Norwegian Institute of Public Health URL: https://www.whocc.no/atc_ddd_index/ [accessed: 05-15-2018].

[27] Autorizzazione all'Immissione in Commercio (AIC), Agenzia Italiana del Farmaco. URL: http://www.agenziafarmaco.gov.it/en . [accessed: 05-15-2018].

[28] Logical Observation Identifiers Names and Codes (LOINC), Regenstrief Institute, URL: https://loinc.org/ [accessed: 05-15-2018].

[29] Liferay 6.2., URL: https://www.liferay.com/downloads . [accessed: 05-15-2018].

[30] Pentaho Data Integration (ETL) a.k.a Kettle, URL: https://github.com/pentaho/pentaho-kettle [accessed: 05-15-2018].

[31] Virtuoso, OpenLink Software, URL: https://virtuoso.openlinksw.com/ . [accessed: 05-15-2018]

[32] CKAN, URL: https://ckan.org/ . [accessed: 05-15-2018].

[33] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, et al., "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: 05-15-2018].

[34] Open Refine, URL: http://openrefine.org/ . [accessed: 05-15-2018].

[35] Legislative Decree 7 March (2005), n. 82. Codice dell'amministrazione digitale. In GU Serie Generale n.112, 16-05-2005 - Suppl. Ordinario n. 93 [available from: http://www.agid.gov.it/sites/default/files/leggi_decreti_direttive/dl-7-marzo-2005-82_0.pdf].