# Standards for the Reuse of Clinical Data in Research: the Connection of the Liguria HIV Network and the CISAI Cohort

M. Bonetto[1], S. Mora[2], N. Maggi[2], C. Ruggiero[2], M. Giacomini[2]

[1],Healthropy s.r.l., Savona, Italy

[2]Department of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Genoa, Italy

Email: mauro.giacomini@dibris.unige.it

*Abstract* — **An example of clinical data re-use achieved in an Italian region for HIV patient treatment is given. A network including all public hospitals in this region has been setup, using several standards at different levels. Specifically, the indications of Healthcare Service Specification Project (HSSP) jointly developed by HL7 and OMG have been extensively used. The automated system which has been setup has brought about a significant increase in the number of lab-tests that have been collected for clinical data re-use purposes. It is intended to extend the software environment to other regions and ultimately at national level.**

*Keywords* — **Interoperability, clinical data re-use, HIV patient data.**

## I. INTRODUCTION

Nowadays large amounts of clinical data are produced during daily routine in hospitals, and most of it is generated in a native digital format[1]. It would be extremely useful if these large amounts of data, which directly derive from real hospital, could be easily used also for research purposes. Unfortunately, many barriers make it difficult to achieve this important re-use:

- most of these data are recorded using a proprietary format;
- when coded, the codes that are used are local
- in many cases, people in charge are concerned about privacy issues when clinical data are made available to external organizations.

It is possible to summarize these aspects under the poly-semantic term "interoperability". Interoperability can be achieved by a correct use of several standards at different levels. The efforts of the HSSP project assure that clinical data can be adequately located and retrieved and the correct management of identification indexes can provide an appropriate level of pseudo-anonymization[2][3].

This paper describes an example of clinical data re-use achieved in the Liguria Italian region for the Liguria HIV network (LHN)[4], in which all Liguria hospitals that treat HIV patients share their native digital data in a common repository. So, all doctors in the network can use their data either to perform local clinical trial or to participate to national and international observational studies. Therefore, this repository contains well assessed and standardized clinical information[5].

The collaboration between LHN and the Italian Coordinated Study on Allergy and HIV Infection (CISAI)[6] within the Surveillance Cohort Long-Term Toxicity Antiretrovirals (SCOLTA) observational study is the main subject of this paper.

SCOLTA cohot has been used for a large number of national clinical studies, more than 200 titles are returned is you use SCOLTA and HIV keywords in Google Scholar since 2005 [7].

## II. METHODOLOGY

SCOLTA aims to empower the pharma-surveillance systems, as relates to new anti-viral drugs. A focus on the control of unexpected and late drug adverse effects that are commonly present in each anti-retroviral treatment is embedded in the systems.

The Ligurian HIV Network can be regarded as a usual web-platform to collect clinical data, however it has some innovative aspects.

The LHN technical implementation choices are based on the recommendations by the Electronic Source Data Interchange (eSDI) group. The whole technical structure adopted and the general architectural approach of the developed solution are completely described in [8], while the technical solutions are similar to those described in [9], [10]. In order to fulfill the European General Data Protection Regulation (GDPR) and to assure adequate protection to patients' privacy, only indispensable personal information is managed such as birth date, sex and nationality. The use of patient hospital codes, that cannot be directly connected to the identity of the patient, makes possible to track each patient through various hospital departments[3]. If a patient moves to another hospital, explicit medical communication is required.

A highly normalized SQL database, developed in a Microsoft SQL Server environment, allows easy maintenance of the parameter set, collected into categories and organized by event types. Clinical results are saved into highly typified data tables, thus avoiding saving errors [11]. The web platform is developed in Microsoft .NET framework and visualization is automated by Microsoft .ascx templates, through which the web pages are dynamically loaded, based on a specific table structure.

The LHN is connected with the Hospital Information Systems of every public hospital that treat HIV patients in Liguria. This assures a continuous feed of the main parameter values relating the clinical situation of each patient. The

system assures that each relevant result about the situation of the patient is received in an anonymized environment within a week from its measurement. For specific clinical trials, the LHN receives data also from two hospitals in Piedmont region.

During this migration a mapping from local codes used by each hospital and international codes (LOINC, ICD 10, ATC, etc.) is performed according to the principles of the Healthcare Service Specification Project (HSSP) jointly developed by HL7 and OMG[2]. Specifically, the mapping is performed according to the definitions collected into the specifications of the Common Terminology Service – Release 2[12]. In this way, the data present in the LHN repository is ready to be transferred to external observational studies, since they can be interpreted in a homogeneous way.

Specifically, RLH data are semi-automatically transferred into the SCOLTA repository using the following procedure, again strongly based on the HSSP requirements:

- within the RLH web interface, the medical doctor responsible of the local site selects patients that are suitable to be inserted into a specific anti-retroviral monitoring class
- RLH and SCOLTA servers exchange a set of messages according to the Cross-Reference Service (IXS) requirements in order to set up a link between the RLH index for a specific patient and the SCOLTA index for the same patient
- SCOLTA, using the indication of Retrieve, Locate and Update Services (RLUS), asks for detailed clinical data of the selected patient and RLH answers with a Clinical Document Architecture (CDA-R2) containing these details
- The follow-up of each patient is performed every six months.

The idea to use the CDA-R2 structure to import clinical data into SCOLTA repository is due to the fact that an Italian rule[13] obliges each Italian hospital to make clinical data available in this format for the Electronic Health Record national system.

## III. RESULTS

The Liguria Ethics Committee has approved this architecture and its functionalities.

The LHN currently involves the Infectious Diseases Departments of eight hospitals (two in Genoa and one each in Alessandria, La Spezia, Pietra Ligure, Sanremo, Savona and Torino). Three hospitals of this group (IRCCS AOU San Martino IST, Galliera and Sanremo) use the automatic connection between their Hospital Information System (HIS) and the RLH repository and from RLH to SCOLTA.

Local automatic collection of data started in 2013 and since than the three hospitals with direct and standardized connection stored a number of lab tests per year five times higher than the other hospitals, in which manual registration was still used. On average, for automatized centers more than 10500 lab-tests per year were recorded, whereas for non-automated centers about 1900 lab-tests per year were recorded.

Automatic and standardized connection between LHN and SCOLTA started in 2016 and similar improvements have been experienced: the average number of patients enrolled per year has been increased by 5 times (from 12 to 64) and the average number of follow up forms has been increased by 6 times (from 52 to 312).

## IV. DISCUSSION AND CONCLUSION

The LHN was initially set up to provide physicians with a web based tool to manage data from HIV-positive patients in primary-care and to re-use the collected clinical information to perform local clinical trials. In this paper, an extension of the former architecture is presented, which provides a system that automatically transfers clinical information from the hospitals' EHR into the LHN database. A further extension enables direct communication between the developed system and one of the main HIV national databases (SCOLTA).

The data can be automatically extracted in a standardized way, and conveyed towards the central repository for their re-use in clinical trials. Normal clinical trials, usually, copy manually data either from paper collected clinical notes or from original Electronic Medical Record systems, but stored in a proprietary format.

In order to manually collect a complete report of the laboratory activity for all the patients involved, a strenuous human effort and the allocation of great resources would have been necessary. Moreover, the risk of errors in a manual system is very high. Instead, a tool able to automatically transfer all data, such as the one described in this paper, is essential to rapidly collect all laboratory information in a reliable way.

The use of medical standards ensures possible future integration with additional health care structures, enhancing the expansion of the project. Currently this system manages and processes data coming from regional hospitals, but in the near future, it would be possible to expand it, involving extra-regional facilities, thanks to interoperability efforts. Extremely important in this respect will be a wide adoption of the HSSP indications, mainly as regarding to the efforts for international terminology vocabularies adoptions[14]. The inclusion of such terminology system could solve one of the biggest problems that we faced during the expansion of the automatic connection to the Laboratory Information System (LIS) of different hospitals. Almost all LISs we encountered contained local coding systems that vary quite frequently during the life cycle of the project. The correct adoption of a CTS2 based terminology system can significantly help both

during the phase of the connection adaptation and in the maintenance of the mapping system between the local code system and the adopted international code system (LOINC in Italy).

REFERENCES

[1] B. AL and K. IS, "Translating artificial intelligence into clinical care," *JAMA*, vol. 316, no. 22, pp. 2368–2369, 2016.

[2] K. Kawamoto, A. Honey, and K. Rubin, "The HL7-OMG Healthcare Services Specification Project: motivation, methodology, and deliverables for enabling a semantically interoperable service-oriented architecture for healthcare," *J. Am. Med. Informatics Assoc.*, vol. 16, no. 6, pp. 874–881, 2009.

[3] G. Gazzarata, R. Gazzarata, and M. Giacomini, "A Standardized SOA Based Solution to Guarantee the Secure Access to EHR," in *Procedia Computer Science*, 2015, vol. 64.

[4] P. Fraccaro and M. Giacomini, "A web-based tool for patients cohorts and Clinical Trials management," *Stud Heal. Technol Inf.*, vol. 180, pp. 554–558, 2012.

[5] P. Fraccaro, V. Pupella, R. Gazzarata, and M. Giacomini, "Reuse of Clinical information: Integrating primary care and Clinical research through a Bidirectional Standard Interface," in *IFMBE Proceedings*, 2014, vol. 41.

[6] "CISAI Coordinamento Italiano Studio Allergie e Infezioni da HIV," 2016. .

[7] P. Bonfanti, C. Martinelli, E. Ricci, S. Carradori, G. Parruti, O. Armignacco, C. Magnani, T. Quirino, and for the C. G. (Italian C. for the S. of A. H. I. V Infection), "An Italian Approach to Postmarketing Monitoring: Preliminary Results From the SCOLTA (Surveillance Cohort Long-Term Toxicity Antiretrovirals) Project on the Safety of Lopinavir/Ritonavir," *JAIDS J. Acquir. Immune Defic. Syndr.*, vol. 39, no. 3, 2005.

[8] R. Gazzarata, B. Giannini, and M. Giacomini, "A SOA-Based Platform to Support Clinical Data Sharing," *J. Healthc. Eng.*, vol. 2017, 2017.

[9] M. Bonetto, M. Nicolò, R. Gazzarata, P. Fraccaro, R. Rosa, D. Musetti, M. Musolino, C. E. Traverso, and M. Giacomini, "I-Maculaweb: A Tool to Support Data Reuse in Ophthalmology," *IEEE J. Transl. Eng. Heal. Med.*, vol. 4, 2016.

[10] R. Gazzarata, F. Vergari, T. Salmon Cinotti, and M. Giacomini, "A standardized SOA for clinical data interchange in a cardiac telemonitoring environment," *IEEE J Biomed Heal. Inf.*, vol. 18, no. 6, pp. 1764–1774, 2014.

[11] P. Fraccaro, V. Pupella, R. Gazzarata, C. Dentone, G. Cenderello, P. De Leo, F. Bozzano, G. C. Finocchio, A. De Maria, and D. Fenoglio, "The Ligurian human immunodeficiency virus clinical network: A Web tool to manage patients with human immunodeficiency virus in primary care and multicenter clinical trials," *Med. 2.0*, vol. 2, no. 2, 2013.

[12] 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," *Eur. J. Biomed. Informatics*, vol. 13, no. 1, pp. 43–50, 2017.

[13] "Decreto del Presidente del Consiglio dei Ministri," *Gazzetta Ufficiale della Repubblica Italiana - 29/09*, p. 178, 2015.

[14] R. Gazzarata, M. E. Monteverde, E. Vio, C. Saccavini, L. Gubian, and M. Giacomini, "A CTS2 compliant solution for semantics management in laboratory reports at regional level," in *Informatics for Health*, 2017.