Ten Years of Medical Informatics and Standards Support for Clinical Research in an Infectious Diseases Network

Sara Mora[1]   Barbara Giannini[1]   Antonio Di Biagio[2,3]   Giovanni Cenderello[4]   Laura Ambra Nicolini[2]
Lucia Taramasso[2]   Chiara Dentone[2]   Matteo Bassetti[2,3]   Mauro Giacomini[1]

[1] Department of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS), University of Genoa, Genoa, Italy
[2] Infectious Diseases Unit, Policlinico San Martino Hospital, IRCCS for Oncology and Neuroscience, Genoa, Italy
[3] Department of Infectious Disease, IRCCS AOU San Martino IST, (DISSAL), University of Genoa, Italy
[4] Infectious Diseases Unit, ASL-1 Imperiese, Sanremo, Imperia, Italy

Address for correspondence Sara Mora, Eng, Department of Informatics, Bioengineering, Robotics and System Engineering, (DIBRIS), University of Genoa, Italy (e-mail: sara.mora@edu.unige.it).

Abstract

Background   It is 30 years since evidence-based medicine became a great support for individual clinical expertise in daily practice and scientific research. Electronic systems can be used to achieve the goal of collecting data from heterogeneous datasets and to support multicenter clinical trials. The *Ligurian Infectious Diseases Network* (LIDN) is a web-based platform for data collection and reuse originating from a regional effort and involving many professionals from different fields.

Objectives   The objective of this work is to present an integrated system of ad hoc interfaces and tools that we use to perform pseudonymous clinical data collection, both manually and automatically, to support clinical trials.

Methods   The project comprehends different scenarios of data collection systems, according to the degree of *information technology* of the involved centers. To be compliant with national regulations, the last developed connection is based on the standard *Clinical Document Architecture Release 2* by *Health Level 7* guidelines, interoperability is supported by the involvement of a terminology service.

Results   Since 2011, the LIDN platform has involved more than 8,000 patients from eight different hospitals, treated or under treatment for at least one infectious disease among *human immunodeficiency virus* (HIV), *hepatitis C virus*, *severe acute respiratory syndrome coronavirus 2*, and tuberculosis. Since 2013, systems for the automatic transfer of laboratory data have been updating patients' information for three centers, daily. Direct communication was set up between the LIDN architecture and three of the main national cohorts of HIV-infected patients.

Conclusion   The LIDN was originally developed to support clinicians involved in the project in the management of data from HIV-infected patients through a web-based tool that could be easily used in primary-care units. Then, the developed system grew modularly to respond to the specific needs that arose over a time span of more than 10 years.

Keywords
► data sharing
► EHRs and systems
► clinical data reuse
► HL7 standards
► clinical data management

Background and Significance

It is 30 years since evidence-based medicine became a great support for individual clinical expertise in daily practice and scientific research.[1] However, this required the need to conduct several clinical trials[2] and to collect a huge amount of data. To achieve this objective, the development of systems able to support first the coordination of clinical trials and then the interoperability of heterogeneous data originating from different medical centers,[3] was necessary. These kinds of systems are indispensable not only to control the huge amount of data produced by health facilities but also to build clinical pathways, research systems, and effective public health management policies.

Since the beginning of the 2000s, the ability to describe heterogeneous data through the use of different kinds of models and connect them to the available web forms has led to the setup of web applications that facilitate the exchange of research data in many fields[4–8]. Multicentric research networks and clinical data management systems (CDMSs) have played an important role in data storing and management within a varying range of medical domains.[9] During the last decades of the 21st century, a large number of health-related networks have been established, such as Cancer Biomedical Informatics Grid in the United States which is a computer-based network that connects scientists and institutions with the objective of empowering the sharing of data concerning cancer.[10] At a higher level, the European Clinical Research Infrastructures Network supports multinational trials in Europe, by connecting and coordinating a great number of European national centers and networks.[11] Then, the i2b2 (Informatics for Integrating Biology and the Bedside) initiative developed a tool for data sharing and integration in precision medicine.[12] In more recent years, two other research projects, ELIXIR[13] and FAIR4Health,[14] have arisen with the aim of coordinating, integrating, and supporting the sharing of the huge amount of data generated by European research organizations during publicly funded research initiatives.

This is the scenario where, in 2011, some of the authors informally created a new research network at a regional level called the Ligurian Infectious Diseases Network (LIDN)[15] during a scientific collaboration between medical groups and bioengineers. At the very beginning, the LIDN was a web platform aimed at enabling the easy collection of human immunodeficiency virus (HIV)-infected patient data in order to conduct multicenter clinical trials at a regional level,[16] the first one involved the drug Maraviroc.[17] Medical experts manually copied data from the electronic health records (EHRs) to the LIDN through a web user interface.[a] During the first year, the platform supported several regional and national studies.[18] However, frequent use of the platform led to the rise of well-known problems caused by manual data input.[19–21] First, the huge amount of time required to insert even a minimal set of data necessary to conduct a clinical trial. Second, the randomness of errors induced by manual

data imputation. Therefore, the evolution of the system was highly required. In particular, structured data such as laboratory test results were already stored in the laboratory information system (LIS) in digital format and so they could be easily read and automatically transferred to the LIDN database. Thus, the paradigm of data reuse allowed the daily update of patients' clinical data without human intervention. The authors decided to start from laboratory data since they are the largest dataset used in scientific research about infectious diseases. Once the LIDN obtained a completely updated database, we planned to exploit this system to avoid manual data entry in other databases such as Antiretroviral Resistance Cohort Analysis (ARCA),[22] Italian Cohort Naive Antiretrovirals (ICONA),[23] and Italian Coordination for the Study of Allergies and HIV Infections (CISAI).[24] These are the most important cohorts for clinical studies concerning HIV at a national level.

These first progresses in the automatization process led the authors to ask for a new approval by the Ligurian Ethics Committee (LEC), specifically for the automatic connection, which was accepted with approval number: 2/2013. This also made the architecture appealing to other diseases. In fact, in the following years, some of the authors conducted other research studies on two of the main infectious diseases present in Italy. One project by the Istituto Superiore di Sanità focused on monitoring new drugs against hepatitis C virus (HCV) and, in parallel, the continuity of care in tuberculosis (TB) patients. Finally, when the COVID-19 pandemic arose, the architecture was rapidly expanded to store data for this new class of patients as well. Specific approval by the LEC was requested and obtained (163/2020–10475).

In addition to the expansion of the group of monitored infectious diseases, the system has also evolved toward the automatic collection of clinical data from other sources. In particular in 2016, the LEC approved the connection between the LIDN and the drug delivery systems of the hospitals involved in the project. Specifically, three of the involved hospitals, the ones directly connected to the LIDN (as detailed below in section "Methods"), adopt an electronic drug delivery system. One of them adopts an individual dose-packaging system, while the other two adopt a computerized physician order entry system.

Objectives

The objective of this work is to present an integrated system of ad hoc interfaces and tools that allows pseudonymous clinical data collection, both manually and automatically, to support clinical trials. First our system drastically reduces the time and human effort necessary to collect the data, and moreover, the level of completeness is higher. Specifically, the system retrieves from the LIS always the most complete set of available information, approved by the LEC. All the data stored in the database can be easily accessed through queries. Furthermore, our system improves the quality of data collected because unlike a human that manually copies data, it does not introduce completely random errors.[25] In addition, it is possible to improve the accuracy of a

---

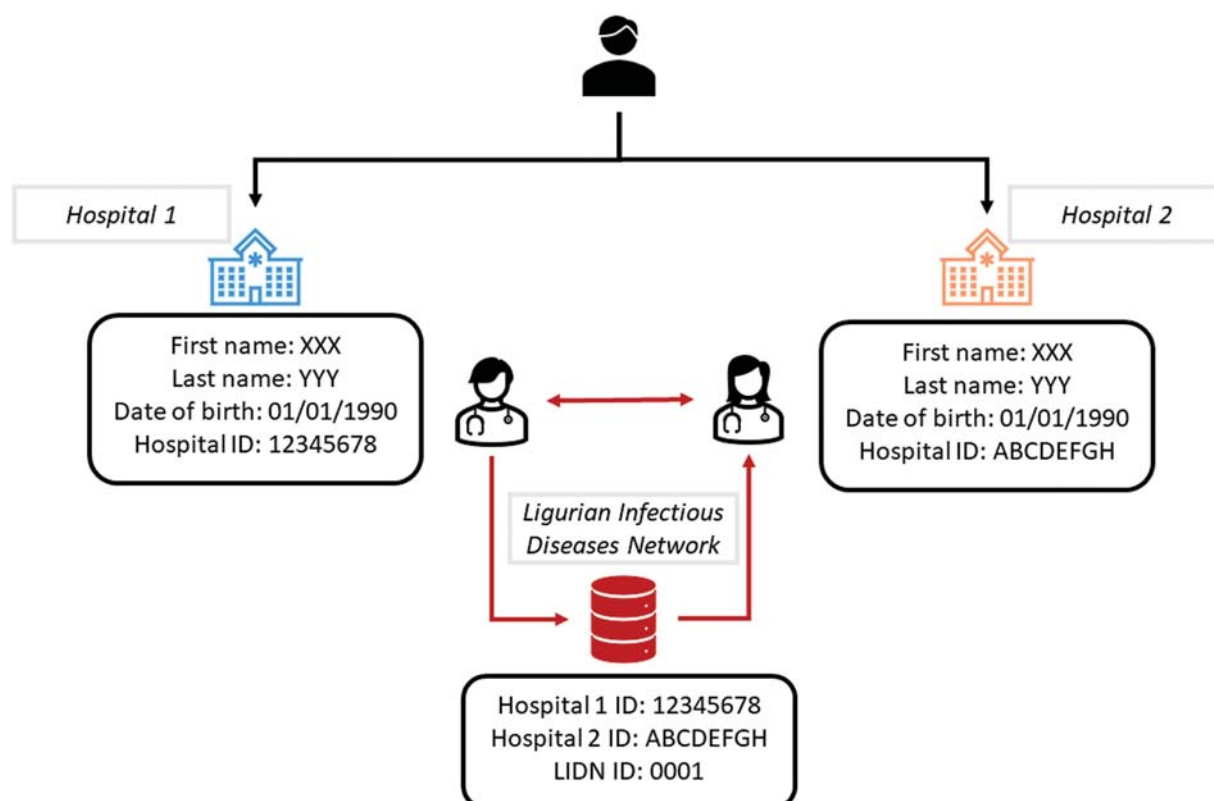[a] LIDN homepage can be reached at: https://www.reteligurehiv.it/.

Fig. 1 Patient's follow-up tracking system. Patient first goes to hospital 1 where the patient is identified by his personal information and a unique hospital ID. Clinician 1 decides to enroll the patient into the LIDN and the patient is assigned a unique ID specific of the LIDN. Then, the patient moves, and the patient starts going to hospital 2 where the patient is identified by personal information and another unique hospital ID. The second clinician decides to enroll the patient into the LIDN. Without any specific coordination with clinician 1 the patient would be duplicated into the LIDN because no information that could uniquely identify the patient are present into the LIDN. While, if the two clinicians coordinate themselves then into the LIDN only one record is generated, and the two hospital IDs are assigned to the same LIDN entity. LIDN, Ligurian Infectious Diseases Network.

computerized system and minimize errors, while it is more difficult to work on the level of attention of a human being. In a more recent but similar platform,[26] we performed a systematic review of data extracted exploiting the same systems described here and results proving the high quality of the final content of the database will be published soon. In the near future, a similar validation will also be performed on data collected in this platform.

Methods

The Ligurian Infectious Diseases Network
According to the project's aim and the previously mentioned approvals received from the LEC, we collected data in a central common database in pseudonymous form and then we made them available to the medical experts, participating in the project, to conduct clinical studies.

The choices that guided the architectural development of the LIDN are described extensively.[27] From a technical perspective, the LIDN is compliant with the recommendations of the electronic source data interchange group and its approach is analog to those introduced.[28,29]

To accomplish the General Data Protection Regulation requirements,[30] we only collected minimal sensitive information, for example, year of birth, gender, and nationality.

Then, considering that most of the involved infectious diseases are chronic ones, this platform needs to identify patient events over long periods of time. Therefore, we decided to collect the patient hospital code, which does not directly identify the patient, but it helps to keep track throughout follow-ups. However, nowadays it easily occurs that a patient moves from one place to another, so it becomes necessary to track him/her through the different hospital departments. Specifically, when a patient with a preexisting diagnosis of a chronic infectious disease, for example, HIV, starts to grasp to another hospital within the one involved in the LIDN, then explicit coordination between clinicians is needed. This coordination is required as the LIDN does not contain any personal information that uniquely identifies the patient. ►**Fig. 1** graphically explains how a patient can be tracked across multiple hospitals. This suboptimal solution has been adopted temporarily while waiting for an institutional solution to patient ID exchange. In Liguria, a platform based on the cross-reference service of the Healthcare Services Specification Project (HSSP) has been announced some time ago but is not yet available.

The platform is designed to allow the storage of different types of clinical data through specific entities called events, such as blood collections, therapy prescriptions, microbiological tests, follow-ups, diagnosis, pregnancies, immunizations,

adverse events, therapeutic failures, hospitalizations, and discharges. Each event consists of different categories, for example, the blood test event includes T-cell subsets CD4 and CD8 panels, a lipid metabolism panel, a liver function panel, a metabolic panel, and a cell blood count differential panel. Each category in turn consists of several parameters, for example, the lipid panel category includes, among others, cholesterol, triglycerides, arteriosus systolic pressure, and arterial diastolic pressure. For each laboratory test result, its unit of measure and normality reference range is also stored so that in the case of multicentric clinical studies, the normalization of the laboratory test results can be achieved through the z-score method, therefore assuring data harmonization.

We developed the platform's database on a Microsoft SQL Server and thanks to the highly normalized structure it allows the easy addition of new parameters, categories, and event types. Highly typified data tables stored clinical results to avoid errors while saving. This means that each type of result (integer, float, string, categorical, and date) is saved into a specific database data table. Results will only be saved into the data table if they belong to the correct type. The web user-interface is developed in Microsoft .NET Framework and page visualization is standardized using Microsoft .ascx templates. These templates allow the dynamical loading of the web pages based on a specific structure defined in database tables achieving a fast personalization of the interface.

Automatic Laboratory Information System Data Export

In multicenter clinical trials, the use of CDMS has become essential to handle the huge amount of data. Several international projects started working on this topic. In the present work, the authors describe their proposal, the LIDN that will be compared to other existing solutions in the discussion section.

Until 2013, physicians manually inserted data through a dedicated web user interface. Then, the platform evolved into its current form when the convenience of data reuse became evident as a consequence of the development of several parallel clinical trials. In particular, we implemented console applications for the automatic extraction of laboratory data directly from the LIS toward the LIDN. We decided to invest first on laboratory data because in the scenario of clinical trials on infectious diseases they are the largest dataset used.

At that time the following two possible scenarios arose,[31] based on the specific hospital level of informatization:

1. In the hospital, there was not an LIS which stored data in a digital format or there was a LIS but external agents could not obtain access to data.
2. In the hospital, there was a LIS and external authorized agents could access a specific subset of data.

Therefore, we considered two different approaches to connect new hospitals to a CDMS, such as the LIDN. Hospitals belonging to scenario 1 could take part in the LIDN only by manually inserting data through the web user interface. While those belonging to scenario 2 could exploit the data reuse paradigm. So, we set up and performed the automatic transfer of clinical data from the specific hospital LIS toward the LIDN in different ways according to the availability of services exposing patients' data and to the national regulation scenario. This choice of organizing our research platform on a regional base is guided by the Italian health care infrastructure. In particular, health care is mainly managed at a regional level and each region also developed its own health information infrastructure.[32] Some of the clinical trials supported by our platform also included data coming from other Italian regions (a complete list of clinical trials involving multiple centers is available in ►Table 1). However, these extraregional involvements are organized by medical experts that designed the specific trial depending on the scientific competencies that each center may bring to enhance it. With regard to the most recent connection, we organized laboratory data in standard documents, in accordance with the Health Level 7 Clinical Document Architecture (HL7 v3 CDA r2) guidelines. Our decision to use the CDA r2 standard was guided by national regulations that all regions must be compliant with. In particular, the Decree of the President of the Council of Ministers (DPCM) of September 2015, titled "Regulation on National Electronic Health Records (NEHR),"[33] identifies the CDA r2 as the standard schema that defines the structure of documents and messages stored in the NEHR. So, from a future perspective, if a hospital is compliant with the regulation, data could be directly retrieved from the NEHR. Since the NEHR was also founded with an explicit research aim, this new scenario will first eliminate the fundamental need to implement ad-hoc systems that read from the databases/service interfaces of each specific hospital and then it will also lead to easier connection of new centers. In particular, after the well-known COVID-19 pandemics, an instance of the NEHR has been automatically opened for everyone, and it is actually extensively used in Ligurian hospitals to store laboratory test results. In the NEHR, all data are stored in CDA r2 standard documents, so for features connections this will facilitate the automatic transfer of data. In the near future, this innovation process will also be supported by the investments allocated by the National Recovery and Resilience Plan (NRRP). Available institutional platforms will be strengthened from the standards point of view, and this will be reflected in a more satisfactory level of security and privacy. For this reason, we think that research platforms, like the one we present here, in the near future, can improve and increase their robustness and usage.

Moreover, another requirement of the 2015 DPCM, cited above, is the univocal interpretation of clinical data. This is guaranteed by the presence of the CDA r2 element "Translation," which maps each code identifying a laboratory result from the local coding system to the international vocabulary Logical Observation Identifiers Names and Codes, as requested in the specifications of the Implementation Guidelines of the CDA r2.[b] The involvement of a terminology service supports this process of mapping between coding

---

[b] Implementation Guidelines of CDA R2 "Laboratory medicine report" can be available at: https://www.fascicolosanitario.gov.it/sites/default/files/public/media/Specifiche%20HL7%20CDA%20R2_RefertoMedicinaLab-v1.1-S.pdf.

Table 1 Clinical trials available on the LIDN

Studies	Description	Patients	Centers
MARHIV (Maraviroc HIV)	Study on Maraviroc therapy	120	The Infectious Diseases Departments of Pietra Ligure, Sanremo, La Spezia, Torino, Alessandria, Savona, Galliera and San Martino hospitals. The Clinical Immunology department of San Martino hospital.
Long-Term	Study on HIV long term non-progressor patients	35	The Infectious Diseases Departments of San Martino and Sanremo hospitals.
ACTeA-I (Cost analysis of patients in Antiretroviral Therapy - Italy)	Analysis of the costs for patients in antiretroviral therapy	208	The Infectious Diseases Departments of Galliera, Sanremo, San Martino, La Spezia hospitals.
Vertical transmission	Study on mother-to-child transmission of HIV	51	The Infectious Diseases Departments of San Martino and Sanremo hospitals.
IANUA (Investigation on the prescribing appropriateness of antiretrovirals used in patients with HIV infection)	Study on antiretroviral prescription appropriateness in HIV	2,245	The Infectious Diseases Departments of Galliera, Sanremo, San Martino, La Spezia hospitals.
T reg (Regulatory T Lymphocytes)	Determination of regulatory T lymphocytes as a new biomarker for HIV monitoring	93	The Infectious Diseases Departments of Albenga, Galliera, Sanremo, San Martino, La Spezia hospitals.
VELA (Value of Ligurian Outcomes for the Appropriateness of HIV Therapies)	Effectiveness and sustainability strategies in the management of long-term HIV patient quality of life	275	The Infectious Diseases Departments of Albenga, Galliera, Sanremo, San Martino, La Spezia and Savona hospitals.
AIFA HCV (Italian Drug Agency pharmacovigilance project for HCV)	AIFA pharmacovigilance project for HCV therapy prescription in monoinfected and coinfected patients	1,504	The Infectious Diseases Departments of Albenga, Galliera, Sanremo, San Martino, La Spezia and Savona hospitals. The Departments of Gastroenterology and Diagnosis and Therapy of Hepatopathies of San Martino hospital.
SCUDI (Study on the costs attributable to the modification of an antiretroviral therapy)	Study on therapeutic failure in HIV patients	389	The Infectious Diseases Departments of Galliera and San Martino hospitals.
SELFY MPI (Self-Administered Multidimensional Prognostic Index)	Complete evaluation of fragility in senior HIV+ patients	157	The Infectious Diseases Departments of Galliera and San Martino hospitals.

Abbreviations: HCV, hepatitis C virus; LIDN, Ligurian Infectious Diseases Network.

systems, and it was built compliant to the standard Common Terminology Service Release 2 developed within the HSSP.[c] Many health care authorities highlighted the deep necessity of using automatic systems for terminology management in clinical-medical fields, among them some are settled in Italian regions.[34–36]

Results

Currently, the LIDN involves the infectious diseases departments of six hospitals (two in Genoa and one each in La Spezia, Pietra Ligure, Sanremo, and Savona). The network is also open to extraregional collaborations on specific proj-

ects, and currently data coming from Alessandria and Turin are also present.

Architectures for Clinical Data Sharing

The three hospitals IRCCS AOU San Martino IST, Galliera, and Sanremo belong to the second scenario described previously in section 2.2, while the others belong to the first one. These three hospitals have different types of information systems, so we developed three different solutions to enable the sharing of clinical data. Galliera hospital offers multiple services that expose patients' clinical data, in particular, a data management service allows access to laboratory test results. After an authentication phase, data can be retrieved in two formats: XML or JSON. The request can contain some specific parameters, for example, the hospital code of the patient and starting date for the research.

[c] https://www.omg.org/spec/company/healthcare-services-specification-project-(hssp)/About-healthcare-services-specification-project-(hssp)/.
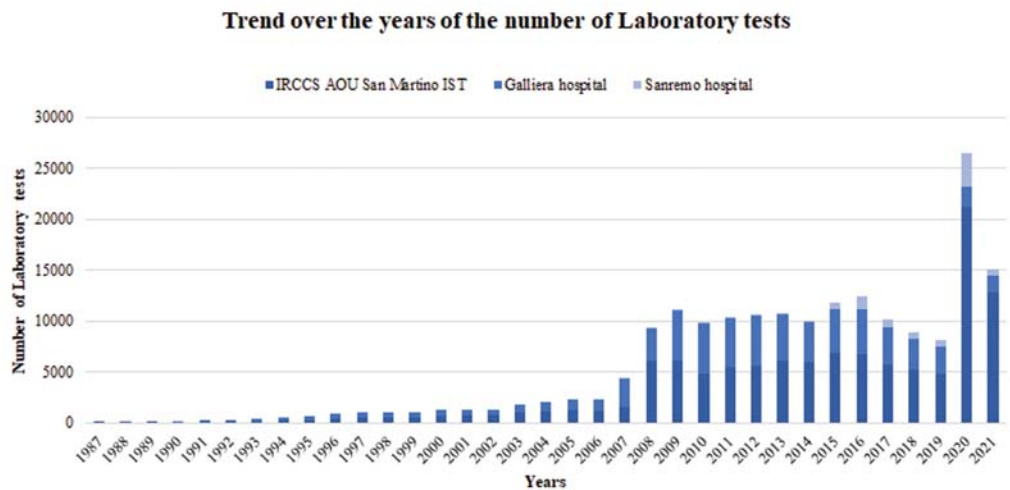
Fig. 2 Cumulative number of tests results stored in the LIDN per year. Only centers with an automatic data connection are included in this figure so that it is possible to appreciate the increased number of data stored after 2008 (IRCCS AOU San Martino IST and Galliera hospital) and 2015 (Sanremo hospital). In 2020 the number of stored tests results is huge due to the COVID-19 pandemic, since patients have been hospitalized and so a great number of tests were done in a short period of time. COVID-19, coronavirus disease 2019; LIDN, Ligurian Infectious Diseases Network.

On the contrary, IRCCS AOU San Martino IST and Sanremo hospitals do not share data through any services, but the information systems can give authorized external agents a read-only access to database views, built ad hoc. This way it is possible to extract the laboratory data of interest for clinical trials through the execution of simple queries.

In both cases, specific Windows Console Applications (WCAs) were developed and, to guarantee the secure read and transfer of data, we installed them within the hospital firewall. A trigger executes the console application program once each night, searching for new available laboratory tests results done during the day by registered patients. When a new patient is enrolled by clinicians into the platform, the program retrieves all historical data according to the characteristics of the disease.

As the connection between Sanremo hospital and the LIDN is the most recent one, it was built compliant with the DPCM, mentioned in section 3.2, which identifies the CDA r2 as the standard schema for clinical data sharing. Two actors are involved to make the standard connection possible: a (1) WCA which differs from the others because it organizes laboratory test results in a CDA r2 document and then sends it to a listening (2) Windows Communication Foundation service which validates the CDA r2 structure and stores the data in the LIDN database.

Numbers of the Ligurian Infectious Diseases Network
Currently, more than 8,000 patients that have been or are still under treatment for at least one of the aforementioned infectious diseases in a center within the eight hospitals are involved in the LIDN. Considering HIV, data from about 1,900 patients from IRCCS AOU San Martino IST, about 1,200 from Galliera, and about 200 from Sanremo have been automatically monitored and updated daily using our system. The numbers are related to the hospital's catchment area, which is smaller for Sanremo hospital as it reflects a less populous area. Patients' data concerning SARS-CoV-2, since

it is an acute infection, are automatically collected starting from the date of the hospital access that resulted in the first positive swab until the date on which the patient is considered cured or dies. Similarly, for data concerning HCV-infected patients, since this infectious disease evolved over time and is currently curable thanks to new treatments, clinicians are asked to insert the date of negativization or death which stops the automatic data read and transfer from the hospital database. On the contrary, since HIV is a chronic condition, patients' data are retrieved from the first positive antibodies test against HIV until the date of death or the date the patient is transferred to another hospital. Data collection rules depend on the approval by the LEC for the specific studies. So, if a study on the long-term effects of acute diseases is presented and approved, our system will be able to collect all necessary data from the day of negativization to the current date (if the patients are still in follow-up in one of the hospitals of the LIDN). In this process of automatic data retrieving, clinicians play an important role in enrolling the patients and storing the precise date of diagnosis.

►**Fig. 2** shows the great impact that automatic data transfer had on the presence of laboratory tests available for clinical studies at Galliera hospital and IRCCS AOU San Martino IST in the LIDN, from 2008, while Sanremo hospital made laboratory test data available in the ad hoc view starting from 2015.

So, the differences between automatic and manual storing of data can be quantitatively evaluated considering the cumulative number of laboratory tests stored per year and the mean number of parameters collected per single laboratory test. ►**Fig. 3** shows these two quantities, calculated for the three automatically connected centers: IRCCS AOU San Martino IST, Galliera hospital, and Sanremo hospital.

To obtain a report of laboratory activities without automatic data transfer for all patients involved in a clinical trial, with a number of collected parameters as high as the one produced by the LIDN, a huge effort would be necessary from
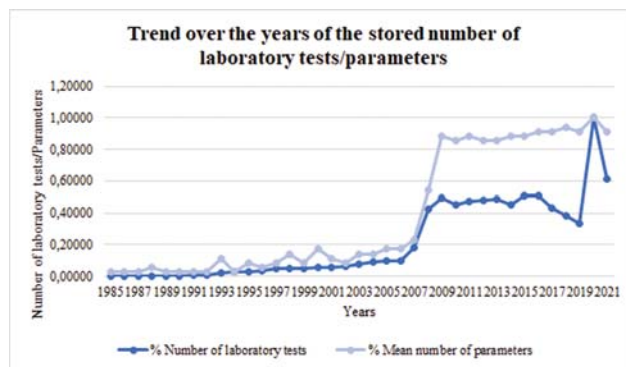
Fig. 3 Trend over the years of cumulative number of tests results and mean number of parameters stored for each test considering IRCCS AOU San Martino IST, Galliera hospital, and Sanremo hospital. In 2020 there is a peak considering the curve of laboratory tests and the other curve reaches its maximum level, as a consequence of the COVID-19 pandemic. Patients were hospitalized so it meant a higher number of tests and new parameters in addition to those of HIV were stored, for example, emogas analysis. COVID-19, coronavirus disease 2019.

a timing point of view in order to have a wide range of data. For example, when compiling an electronic case report form, some of the parameters to be collected can be flagged as optional. Using an automatic system, the effort necessary to transfer the complete set of all the available data that are required or just the minimum set of mandatory ones is the same, but this is not true when the data imputation is performed manually. So, when an automatic transfer system is used, the set of available data is always as wide as possible. Our system, which automatically extracts and transfers pseudonymous data from the EHRs toward the CDMS, remarkably reduced the time necessary to collect a considerable subset of the clinical data necessary to conduct research activities. In order to guarantee the high quality of our data, we periodically perform a manual validation by comparing the information stored in our database directly with those present in the LISs. Two of the main criticalities that arose during the 10 years of activity and that led to a sudden lack of information are (1) the changes in the names and codes that identify the laboratory analysis within the hospital database and this problem further highlighted the need of using standard vocabularies; (2) the transition from decimal notation to scientific notation when reporting the results of some laboratory analyses. The presented platform supported the 10 local studies listed in ►**Table 1–**[17,37–41] each of them with its specific interface.

The increased speed of collecting high-quality data resulted in an increased frequency of publication by supporting the comparison between the different antiretroviral treatments and the periodical data update between the LIDN and the main cohorts at the national level. Two examples of important and innovative studies conducted on the amount of data available in our platforms are shown.[42,43] The first one investigates the comparison between "no treatment" disease progression costs and the economic impact of regimens based on Sofosbuvir (SOF) on a population of HIV/HCV-coinfected patients. The second one provides one of the first experiences studying the changes in lipid profile,

CD4, and creatinine after switching from a combination therapy with tenofovir disoproxil fumarate to a combination therapy with tenofovir alafenamide.

Connections to National Databases

Once a completely updated database is obtained, it could be reused to export reliable information toward national databases: ARCA, ICONA, and CISAI. At present, almost 1,300 patients involved in the LIDN are also enrolled in ARCA, 700 in CISAI, and more than 700 in ICONA. Considering ARCA and CISAI, authors set up a direct collaboration that allows an automatic update of data. On the contrary, ICONA is connected to an offline procedure based on Excel files derived from shared predefined templates.

Regarding the connection between the LIDN and the ARCA and CISAI databases, it is established at an analytical level. So, it is possible to create a complete description of the patient because the semantics of the arrival database is known. Thanks to this, all information about the single patient can be translated from the local vocabulary to the arrival database semantics in a stable way. Specifically, a WCA was developed to fill the ARCA database with the records of the LIDN patients, in particular with CD4+ cell counts, HIV RNA copies/mL, and genetic sequence data. The CISAI cohort, together with HIV, also involves HCV-infected subjects and collects data about both these infectious diseases.

Additional Tools of the Web User Interface

Data stored in the database can be used to produce statistical reports and charts to visualize continuously updated information. Pie and point-line charts, histograms, and tables display absolute, relative, and percentage frequencies of data for epidemiological studies. It is also useful to monitor complicated cases because clear and synthetic information is displayed, for example, in ►**Fig. 4** all CD4 and HIVRNA values for a randomly selected patient are displayed over a time span of 9 years. To export a selected set of information to predefined spreadsheet files, a flexible and optimized extraction tool was set up. The user can customize the following information: patient, time range, and parameters of interest.

The web interface also allows clinicians to calculate some important clinical scores, for example, Glasgow coma scale, glomerular filtrate, etc., but we decided not to store results in the database to reduce data redundancy. Other more complex scores, such as viremia copy years, can be requested by clinicians and provided in the process of data extraction for specific clinical trials.

Discussion

In the present manuscript, we described the results of our work which involved many professionals from different fields for more than 10 years and whose main aim was to support infectious diseases clinical trials. We developed a modular system which grew in order to respond to the specific needs that appeared in these 10 years, for example, COVID-19 pandemics, as shown in ►**Fig. 5**.
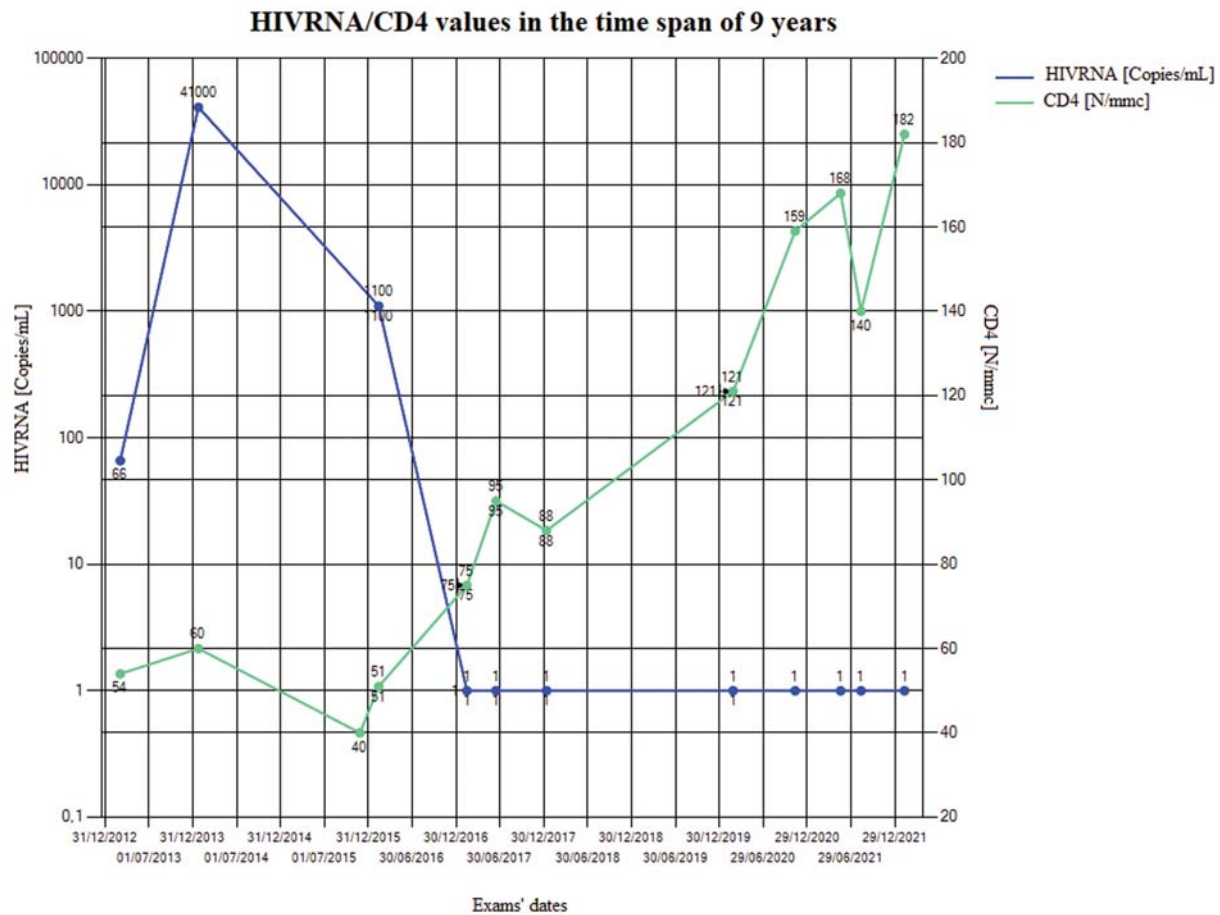
Fig. 4 Trend over the years of HIVRNA and CD4 values for a randomly selected patient within the LIDN. The two scales have different dimension orders, the HIVRNA scale is logarithmic and values corresponding to 1 are below the threshold, while the CD4 scale is linear. LIDN, Ligurian Infectious Diseases Network.
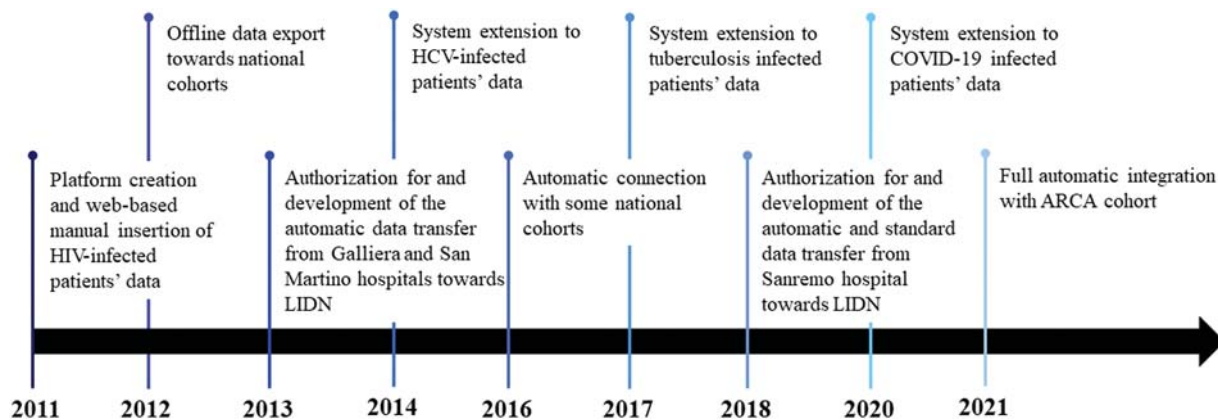

Fig. 5 Timeline summarizing the 10-years evolution of our system, from 2011 to current.

By describing the history of our platform development, we take the opportunity to highlight the critical issue of lack of funding for infrastructures, like ours, which are fundamental in establishing efficient clinical data management. The lack of funding was recognized by Italian government who, in the NRRP, identified as of primary importance the objective of professional management of health care data.[44]

During the 10-year period, the platform was mainly developed and managed by PhD students with minimal funding and to date, the platform has survived on residual funds from the clinical trials themselves. We hope that our findings will encourage more clinical trial groups to include data management plans alongside proposal submissions. The platform staff (medical experts and one bioengineer) played an important role in the supervision and quality

assessment of the interventions performed by the PhD and postgraduate students. The natural transition from one student to another produced periodical checkups of the system, and students and staff formed a deep understanding of the platform, allowing them to look for and identify possible criticalities, for example, due to the introduction of new technologies and the development of new branches.

When we started in 2011, there were no known worldwide systems with this purpose of supporting clinical trials, but at present we know that in the following years, systems with this aim grew and nowadays some of them are internationally adopted. In this discussion section, we want to compare our work with two worldwide well-known systems, Research Electronic Data Capture (REDCap) and i2b2, and we want to motivate the decision to continue using our system instead of adopting a new one. REDCap is a web-based system first released in 2004 by the Vanderbilt University with the aim of creating databases and projects for clinical trials. The main advantage is that users can quickly develop and customize databases and projects for study requirements, even though they are not programmers. Users can access projects from any device, they only need an internet connection, and they can export all data (all records and fields) in many different formats, one of which is the standard format "CDISC ODM". The main disadvantages are that changes in the database or project structure should be approved by the informatics core and certain revisions may not be allowed; data should be collected separately for each project, and even though the REDCap system also made available some API's for the automatic data transfer from a specific EHR to the REDCap project, this feature requires more in-depth technical competences.[45,46] Recent papers, found in the literature, reported that other groups stepped back from the REDCap application to custom solutions due to specific limitations they found.[46,47]

The i2b2 is a clinical data warehousing and analytics research platform that was first released in 2007, and its main aim is to create a cost-effective and efficient system for clinical and translational research.[48] It uses a simple data model that is similar to the *Event-Attribute-Value* approach and ontologies to translate concepts to local codes. The i2b2 query tool provides users with an instrument for easy cohort design and exports data in various formats compatible with many analytics tools. The main advantage of the use of i2b2 in multicenter clinical trials is that it provides an information model that once mapped to the local models can be used for several studies on different topics. In this way, it is not necessary to map data to a new information model each time we want to take part with a different cohort.[49] The main limitation is that i2b2 only allows the importation of bulk data through CSV files, many ad-hoc services and APIs have been developed within separate research projects[50,51] while others suggest the use of medical standard formats, for example, CDISC ODM.[52]

The system we present is similar to REDCap, even regarding the presence of an automatic data transfer system, but on a smaller scale. In addition, the presence of a multidisciplinary team relieves clinicians from several technical issues. We

think that it is not worth stopping the use of our system to build everything from scratch again. During the COVID-19 pandemic, the authors collaborated on a cardiological clinical study organized with the REDCap platform. Taking into consideration the emergency situation and the need to collect data in the shortest time possible, the authors used a predefined template to transfer data, but a future objective is to make this connection more efficient, also from a timing point of view, setting up an automatic system. To improve the semantic connection between the two systems, we expect to use the new standard ISO: 23903:2021,[d] in addition to the HL7 standards and the above-mentioned terminology system. This new standard is devoted to the definition of data interchange through a more profound knowledge of the data models of the sending and receiving systems.

Similar to i2b2, we exploited the data reuse concept, so that once retrieved, data can be used for several clinical trials, it is only necessary to change the queries. Then, we provide clinicians with a set of tools to evaluate patients' follow-up, for example, through the graphical visualization of the trend over time of some laboratory test results. An advantage of an "in house" system is the more efficient building of specific rules for data extraction due to the deep knowledge of the database structure. So, we have a direct access to data instead of always using the system interfaces.

The major limitation of our system is that a technician must manage it. The interface is very simple from a graphical point of view, easy to use, and specific for regular users (who always do the same operation). If, on the contrary, clinicians want to carry out more complex processing, they should request that a technician performs the changes because they cannot autonomously perform operations such as modify a page and consequently the collected parameters. The procedures illustrated in this paper do not all belong to a routine process, so they should be followed by a person with a deep knowledge of database management and whose background allows them to interpret medical requests. Therefore, we believe that the most focused figure that could assume this role is the bioengineer with a good knowledge of medical informatics. In recent years, our group also developed a university spin-off which, in addition to other projects centered mainly on the use of standards in medical informatics and planned a reengineering process of the platform in order to insert it at least in the national market. However, until now, no relevant economic evidence was found. This is due to the presence of several barriers that should be overcome in terms of finance, dissemination of standards, misperception of privacy by hospital staff, etc. On the contrary, the main favorable factors for the implementation, development, and diffusion of this kind of platform for scientific research are the laws related to the NEHR and the growing propension of medical experts toward the digitalization of health.

In our laboratory PhD students and other trainees are now applying machine learning and natural language processing.[53–55] It will be quite interesting and relatively easy to

---

[d] *https://www.iso.org/standard/77337.html.*

extract and organize these data to verify clinical and diagnostic hypotheses given a large number of data and taking inspiration from similar works already present in the literature.[56–58]

Conclusion

In conclusion, we think that we should keep our system, but in the following years, we will arrange a set of interfaces and services (hopefully standard based) to be ready to collaborate with other systems.

The original objective of the LIDN was to develop a web-based tool that allowed clinicians in primary-care units and specific infectious diseases wards to collect and manage data of HIV-infected patients both for patient care and scientific research. In this paper, the authors want to present the evolution over time of the original plan with the development of an automatic data transfer system first from the hospitals EHRs toward the LIDN and then from the LIDN toward the main HIV national cohorts. This process of data reuse is essential to avoid errors due to manual data input, and it remarkably reduces human effort and the time necessary to collect clinical data for research purposes. It also guarantees the high quality of the collected data. Moreover, the paper presents the evolution of the platform that at present also includes data from patients affected by other infectious diseases like HCV, TB, and more recently SARS-CoV-2.

The implemented solution, thanks to the use of medical standards, could easily be further expanded in the future by adding new regional or extraregional hospitals.

Clinical Relevance Statement

The LIDN platform simplifies the management of clinical trials thanks to its web interface that allows smart visualization of a patient's clinical information through statistical elaborations, charts, and a user-friendly customizable query tool. Moreover, the addition of the described tool that automatically extracts data from hospital EHRs allows the management of a greater number of patients; in fact, after the enrollment of a new patient, all his clinical data are available on the platform without human effort, with continuous updating.

Multiple-Choice Questions

1. Which is the appropriate standard for document definition and exchange according to the DPCM "Regulation on National Electronic Health Records"?
   a. Clinical Document Architecture Release 2
   b. Common Terminology Service Release 2
   c. LOINC
   d. Clinical Data Management System

   Correct Answer: The correct answer is option a. The CDA R2 is a standard architecture suitable for clinical document exchange and it is based on the Reference Information Model (RIM) of HL7 V3.

2. Which of the following is a possible connection to the presented system?
   a. Only by CDA
   b. Only manually through the interface
   c. Only by not standard web services
   d. Either by CDA or by manual data insert or by web services

   Correct Answer: The correct answer is option d. The authors developed specific interfaces for all three possible connections based on the different IT levels of the involved centers. The most appropriate one is the standard connection through CDA R2.

References

1 Sackett DL. Evidence-based medicine. In: Seminars in Perinatology. Vol 21. The Netherlands: Elsevier; 1997:3–5
2 Fahey T, Griffiths S, Peters TJ. Evidence based purchasing: understanding results of clinical trials and systematic reviews. BMJ 1995;311(7012):1056–1059, discussion 1059–1060
3 Kush RD, Warzel D, Kush MA, et al. FAIR data sharing: the roles of common data elements and harmonization. J Biomed Inform 2020;107:103421
4 Giacomini M, Pastorino L, Soumetz FC, et al. Data modeling for tools and technologies for the analysis and synthesis of NANO-structures (TASNANO) Project. J Inf Technol Res 2009;2(03):49–70
5 Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 2007;35(Database issue):D198–D201
6 Patel S, Lyons-Weiler J. caGEDA: a web application for the integrated analysis of global gene expression patterns in cancer. Appl Bioinformatics 2004;3(01):49–62
7 Izzo M, Mortola F, Arnulfo G, Fato MM, Varesio L. A digital repository with an extensible data model for biobanking and genomic analysis management. BMC Genomics 2014;15(03):1–15
8 Corradi L, Porro I, Schenone A, et al. A repository based on a dynamically extensible data model supporting multidisciplinary

research in neuroscience. BMC Med Inform Decis Mak 2012;12 (01):115

9 El Emam K, Jonker E, Sampson M, Krleza-Jerić K, Neisa A. The use of electronic data capture tools in clinical trials: web-survey of 259 Canadian trials. J Med Internet Res 2009;11(01):e8

10 Fenstermacher D, Street C, McSherry T, Nayak V, Overby C, Feldman M. The cancer biomedical informatics grid (caBIG TM). In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. IEEE; 2006:743–746

11 Demotes-Mainard J, Kubiak C. A European perspective—the European clinical research infrastructures network. Ann Oncol 2011;22(Suppl 7):vii44–vii49

12 Dagliati A, Sacchi L, Bucalo M, et al. A data gathering framework to collect type 2 diabetes patients data. In: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). IEEE; 2014:244–247

13 Crosswell LC, Thornton JM. ELIXIR: a distributed infrastructure for European biological data. Trends Biotechnol 2012;30(05): 241–242

14 Gencturk M, Teoman A, Alvarez-Romero C, et al. End user evaluation of the FAIR4Health data curation tool. Stud Health Technol Inform 2021;281:8–12

15 Giannini B, Riccardi N, Cenderello G, Di Biagio A, Dentone C, Giacomini M. From Liguria HIV web to Liguria infectious diseases network: how a digital platform improved doctors' work and patients' care. AIDS Res Hum Retroviruses 2018;34(03):239–240

16 Fraccaro P, Pupella V, Gazzarata R, et al. 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 2013;2(02):e5

17 Rusconi S, Vitiello P, Adorni F, et al. Maraviroc as intensification strategy in HIV-1 positive patients with deficient immunological response: an Italian randomized clinical trial. PLoS One 2013;8 (11):e80157

18 Fraccaro P, Pupella V, Gazzarata R, Giacomini M. Reuse of clinical information: integrating primary care and clinical research through a bidirectional standard interface. In: IFMBE Proceedings. Vol 41. Springer Verlag; 2014:1274–1277

19 Hammond KW, Helbig ST, Benson CC, Brathwaite-Sketoe BM. Are electronic medical records trustworthy? Observations on copying, pasting and duplication. AMIA Annu Symp Proc 2003; 2003:269–273

20 Kozak M, Krzanowski W, Cichocka I, Hartley J. The effects of data input errors on subsequent statistical inference. J Appl Stat 2015; 42(09):2030–2037

21 Lai KH, Topaz M, Goss FR, Zhou L. Automated misspelling detection and correction in clinical free-text records. J Biomed Inform 2015;55:188–195

22 ARCA (Antiretroviral Resistance Cohort Analysis). Accessed December 15, 2022 at: https://www.fondazioneicona.org/

23 Fondazione ICONA Italian Cohort Naive Antiretrovirals.

24 CISAI Coordinamento Italiano Studio Allergie e Infezioni da HIV. Accessed December 15, 2022 at: https://www.cisai.it/

25 Cong G, Fan W, Geerts F, Jia X, Ma S. Improving data quality: consistency and accuracy. In: Proceedings of the 33rd International Conference on Very Large Data Bases, Vol. 7. Austria: University of Vienna; 2007:315–326

26 Mora S, Giacobbe DR, Russo C, et al. A wide database for future studies aimed at improving early recognition of Candidemia. Stud Health Technol Inform 2021;281:1081–1082

27 Gazzarata R, Giannini B, Giacomini M. A SOA-based platform to support clinical data sharing. J Healthc Eng 2017. Doi: 10.1155/2017/2190679

28 Gazzarata R, Vergari F, Cinotti TS, Giacomini M. A standardized SOA for clinical data interchange in a cardiac telemonitoring environment. IEEE J Biomed Health Inform 2014;18(06): 1764–1774

29 Bonetto M, Nicolo M, Gazzarata R, et al. I-Maculaweb: a tool to support data reuse in ophthalmology. IEEE J Transl Eng Health Med 2016;4:1–10

30 General Regulation on Data Protection. Published 2016. Accessed December 5, 2022 at: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/6264597

31 Matsumura Y, Hattori A, Manabe S, et al. Interconnection of electronic medical record with clinical data management system by CDISC ODM. Stud Health Technol Inform 2014;205:868–872

32 Lazarova E, Mora S, Maggi N, et al. An interoperable electronic health record system for clinical cardiology. Informatics (MDPI) 2022;9(02):47

33 President of the council of Ministers. Regulation on Electronic Health Records. Published 2015. Accessed December 05, 2022 at: https://www.agid.gov.it/sites/default/files/repository_files/linee_guida/dpcm_178_2015.pdf

34 Gazzarata R, Eugenia Monteverde M, Vio E, et al. A terminology service compliant to CTS2 to manage semantics within the regional HIE. Eur J Biomed Inform (Praha) 2017;13(01):43–50

35 Iannacchero R, Mastrandrea C, Conforti D. Digital health and clinical decision support: the HealthSOAF project and the Calabria Headache Network. Confinia Cephalalgica 2018;28:25–31

36 Ciampi M, Marangio F, Schmid G, Desira M, Esposito A, Sicuranza M A Blockchain architecture for the Italian EHR system proof-of-familiarity: a privacy-preserved blockchain scheme for collaborative medical decision MD Mehedi Hassan Onik Dwarna: a blockchain solution for dynamic consent in biobanking A Blockchain Architecture for the Italian EHR System

37 Ameri M, Cassola G, Cenderello G, et al. Quality of life among HIV patients: results from the Ianua Clinical Trial. Value Heal J Int Soc Pharmacoeconomics Outcomes Res 2015;18(07):A592

38 Giannini B, Gazzarata R, Orcamo P, et al. IANUA: a regional project for the determination of costs in HIV-infected patients. Stud Health Technol Inform 2015;210:241–245

39 Venturini A, Giannini B, Montefiori M, et al. Quality of life of people living with HIV, preliminary results from IANUA (Investigation on Antiretroviral Therapy) study. J Int AIDS Soc 2014;17(4, suppl 3):19581

40 Pupella V, Gazzarata R, Macrì M, et al. A tool for the evaluation of the cost effectiveness of AIDS therapies. In: IV Congresso Nazionale di Bioingegneria–GNB; 2014:2

41 Mora S, Venturini A, Cenderello G, Fiorellino D, Pilotto A, Giacomini M. A web-based tool for a complete evaluation of fragility in senior HIV+ patients. Stud Health Technol Inform 2019; 261:299–302

42 Cenderello G, Artioli S, Viscoli C, et al. Budget impact analysis of sofosbuvir-based regimens for the treatment of HIV/HCV-coinfected patients in northern Italy: a multicenter regional simulation. Clinicoecon Outcomes Res 2015;8:15–21

43 Taramasso L, Di Biagio A, Riccardi N, et al. Lipid profile changings after switching from rilpivirine/tenofovir disoproxil fumarate/emtricitabine to rilpivirine/tenofovir alafenamide/emtricitabine: different effects in patients with or without baseline hypercholesterolemia. PLoS One 2019;14(10):e0223181

44 PNRR - Salute. Accessed December 5, 2022 at: https://www.pnrr.salute.gov.it/portale/pnrrsalute/dettaglioContenutiPNRRSalute.jsp?lingua=italiano&id=5879&area=PNRR-Salute&menu=investimenti

45 Garcia KKS, Abrahão AA. Research development using REDCap software. Healthc Inform Res 2021;27(04):341–349

46 Blumenberg C, Barros AJD. Electronic data collection in epidemiological research. The use of REDCap in the Pelotas birth cohorts. Appl Clin Inform 2016;7(03):672–681

47 Zhang H, Dimitrov D, Simpson L, et al. A web-based, mobile-responsive application to screen health care workers for COVID-19 symptoms: rapid design, deployment, and usage. JMIR Form Res 2020;4(10):e19533

48 Murphy SN, Mendis M, Hackett K, et al. Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside. AMIA Annu Symp Proc 2007; 2007:548–552

49 Klann JG, Abend A, Raghavan VA, Mandl KD, Murphy SN. Data interchange using i2b2. J Amer Med Inform Assoc 2016;23(05): 909–915

50 Abend A, Housman D, Johnson B. Integrating clinical data into the i2b2 repository. Summit On Translat Bioinforma 2009;2009:1–5

51 Maier C, Christoph J, Schmidt D, et al. Experiences of transforming a complex nephrologic care and research database into i2b2 using the IDRT tools. J Healthc Eng 2019;2019:5640685

52 Ganslandt T, Mate S, Helbing K, Sax U, Prokosch HU. Unlocking data for clinical research—the German i2b2 experience. Appl Clin Inform 2011;2(01):116–127

53 Mora S, Attene J, Gazzarata R, et al. A NLP pipeline for the automatic extraction of a complete microorganism's picture from microbiological notes. J Pers Med 2022;12(09):1424

54 Giannini B, Gazzarata R, Sticchi L, Giacomini M. A SOA-based solution to monitor vaccination coverage among HIV-infected patients in Liguria. Stud Health Technol Inform 2016;228:327–331

55 Giannini B, Riccardi N, Di Biagio A, Cenderello G, Giacomini M. A web based tool to enhance monitoring and retention in care for tuberculosis affected patients. Stud Health Technol Inform 2017; 237:204–208

56 Nagy M, Radakovich N, Nazha A. Machine learning in oncology: what should clinicians know? JCO Clin Cancer Inform 2020; 4:799–810

57 Peiffer-Smadja N, Rawson TM, Ahmad R, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clin Microbiol Infect 2020;26(05): 584–595

58 Fraccaro P, Nicolo M, Bonetto M, et al. Combining macula clinical signs and patient characteristics for age-related macular degeneration diagnosis: a machine learning approach. BMC Ophthalmol 2015;15(01):10