Available online at www.sciencedirect.com

ScienceDirect

Physics Procedia 62 (2015) 29 – 35

Physics Procedia

3rd International Conference Frontiers in Diagnostic Technologies, ICFDT3 2013

Management and Analysis of Biological and Clinical Data: How Computer Science May Support Biomedical and Clinical Research

Pierangelo Veltri[a]

[a]*Bioinformatics Laboratory Surgical and Medical Science Department*
University Magna Graecia of Catanzaro, Italy

Abstract

The use of computer based solutions for data management in biology and clinical science has contributed to improve life-quality and also to gather research results in shorter time. Indeed, new algorithms and high performance computation have been using in proteomics and genomics studies for curing chronic diseases (e.g.,, drug designing) as well as supporting clinicians both in diagnosis (e.g., images-based diagnosis) and patient curing (e.g., computer based information analysis on information gathered from patient).

In this paper we survey on examples of computer based techniques applied in both biology and clinical contexts. The reported applications are also results of experiences in real case applications at University Medical School of Catanzaro and also part of experiences of the National project Staywell SH 2.0 involving many research centers and companies aiming to study and improve citizen wellness.

Peer-review under responsibility of the ENEA Fusion Technical Unit

Keywords: EPR Electronic Patient Record; Data management; Informative System;

1. Introduction

Life science actors include clinicians, biologists, biomedical engineers, bioinformaticians and many other interested in life-science related topics, and of course patients and (more in general) citizens. Information related to life science includes disease protocols, treatments and rules for early disease detection. Recently, many governments are investing in largely diffusion of knowledge regarding health related topics, starting from making available records on chronic diseases treatments, protocols and results (follow up), trying to boost researchers in finding always more accurate and efficient strategies for chronic disease treatment. Having strategies for containment or prevention of diseases is one of the main targets for improving life quality and also to reduce huge costs related to health management. Making data available from health structures, following the direction of the Open Data paradigm, may allow the distribution of knowledge in terms of disease treatments or early detection; it has been proved that having early

* Corresponding author. Tel.: +39-0961-3694149 ; fax: +39-0961-3694092
E-mail address: veltri@unicz.it

Peer-review under responsibility of the ENEA Fusion Technical Unit
doi:10.1016/j.phpro.2015.02.007

disease identification is often the best way to treat chronic or rare diseases. Governments play crucial role in trying to diffuse knowledge on disease treatments and drugs follow up.

Management of Health Care structures is focusing on appropriateness of care measurement by using computer based methodologies to model the whole process of admission, permanence hospitalization period, resignation and controls of patients, considering both clinical and economic aspects. This challenge is very relevant for health structures (both public and private ones), which must optimize services and monitor expenses in order to obtain refunds for the services offered to patients. Moreover, keeping expenses under control and making them more objective could make the financial distribution of resources among structures (or operative units) much easier, helping decision makers for resource planning. The requirements of using digital format for clinical information has been thus assuming not only the relevance of gathering improvements in quality of service and information diffusion and portability, but also in terms of economical monitor and flow analysis. The clinical infrastructure has to include thus patients and medical actors, as well as relevant information and their flows, such as invoices, employees turnover, dismissing documentations with performed studies, human and automatic procedures, diagnoses and clinical histories, patients moves and transfers among structures and so on. Information related to the aspects need to be appropriately and efficiently acquired, coded, processed, persisted, organized and retrieved. Such objectives could be achieved by using existing standards for data management and computer based information systems, i.e. systems for automatic management of information. It is also undoubtful that IC technologies improve the possibility of applying interventions for patient treatments and monitoring as well as reduce the time of disease detections. This can be simply proved both referencing to more and more accurate diagnostic techniques (e.g., image diagnostic) as well as accurate and high-performance patient status monitoring (e.g. Telemedicine and controls). By using digital and inter-operable systems it should be possible having benefits from both quality of services and cost savings.

We here focus on role of computer science based strategies for managing and querying life-science-related data, focusing on *omics* data and on information related to patient treatment. Indeed, today there is an always increasing necessity of managing and analyzing omics data to extra information useful for disease treatment and prevention (Cannataro et al 2007; Indolfi et al. 2009). Omics data includes genomics, proteomics, and interactomics, and respectively refer to the study of the genome, proteome and interactome of an organism. Omics data is also increasing due to the availability of novel, high-throughput platforms for the investigation of the cell machinery, such as mass spectrometry, microarray, next generation sequencing, that are producing an overwhelming amount of experimental omics data. The increased availability of omics data also poses new challenges for its integration and correlation with clinical information to improve procedures and processes in health informatics (PRIN Gendata2020).

Indeed, omics data produced by laboratories may be used for improving clinical procedures and patient treatments. In both cases, i.e. omics data management and clinical data management and integration, computer science techniques are required to efficiently: managing data, integrating data, querying data, and making available knowledge to main actors of life science.

The paper presents some experiences of applying computer based know-how on both omics and patient/citizen-oriented information management. We present experiences on omics data manipulation to extract (unveil) information from (i) mass spectrometry data, by using data manipulation techniques as well as time series based techniques; and from (ii) microRNA data by using ad hoc designed plug ins, and ontology based techniques (Cannataro et al 2007; Indolfi et al. 2009). Also, we present application of bioinformatics technologies for supporting clinicians while treating patients both in an (i) offline cases, allowing early disease detections as well as supporting diagnosis validations and (ii) on-line cases where injuries treatments can be improved by using computer aided techniques. An example in hemodynamic surgery room is reported. In the first case we discuss the advantages of using computer based solutions for monitoring and managing clinical data, as well as to support screening on large population. In the latter we discuss how computer science research may aid in studying solutions to support physicians (Amato et al. 2009; Guzzi et al 2010).

2. Experiences of software tools developed for patient-bed use

Here we report on experiences of software tools developed starting from requirements and necessities we gathered from University School Hospital. We started from a mechanism of control data flow related to requirements of patient flow control and report about an Electronic Patient Record system (EPR) for monitoring data in surgery room and present a software on stent analysis measurement in surgery room. We define both patient-bed based tool since the decision related to patient treatment is performed on line while patient is hosted in the hospital.

EPR management and Flow Control in Surgery Room. The advantages of using electronic patient records (EPRs) in health-care systems have been studied and proved in the literature. Similarly, the problem of data and procedure integration to allow health structure cooperation has been studied and represents today a challenge for interoperability and cooperation among health structures (Cannataro et al 2008). The main problem seems to be related on implementation and application of system information solutions. We here focus on the application of EPR to control data flow in hemodynamics room i.e., to check the status of patients arriving into the surgery rooms. The system allows to manage data-flow of patient in cardiological unit. Patients may arrive in emergency from remote hospitals, thus the EPR management needs to include data from data management coming from external systems (often in hard-copy form).

Patients coming from day hospital (for follow up controls or for periodical controls) are also included in the system. The system allows to manage patient flow in hemodynamics room for angioplastic procedures. In this case an ad hoc panel shows on the wall of the protected emergency rooms the information about incoming patient, type of interventional, medical doctors and all the historical information necessary to give full information. Moreover, the system has been integrated with the internal PACS that stores and manages DICOM images produced during intervention and that must be stored in a RAID-based image data management. Panel system is shown in Figure 1 (see also (Cannataro et al. sighit2012)).

Fig. 1. Surgical Room EPR System

Stent designing system. Angioplatic is the primary tool to asses procedural outcomes after a number of cardiac for cornoary stent implantation. Nevertheless, in order to avoid stent thrombosis, drug eluding stents, as well as bare metal ones, stent implantation should be perfectly deployed. Thus, the estimation of diameter and length of the coronary vessels as well as overlap of stents at the origin of large collateral branches, is critical. Among the experiences, we report on *Cartesio*, an innovative software tool to be used by physicians working in emodynamics surgery rooms. It helps in making a pre-implant analysis for the estimation of the dimensions of the stent to be implanted. The tool interacts with virtually any angiographic equipment by acquiring its high-resolution video signal and offering a set of functions to play with images and to draw a virtual stent over the acquired video frames. It allows the operator to calibrate the stent before implanting it. Measurements help physicians to evaluate the exact dimension of the stenosis. Once having determined such a dimension, the physician can set the parameters of the virtual stent and visualize it over the vessel structure.

The system is currently used as a prototype for testing in the emodynamic surgery room at University of Catanzaro (see also (Indolfi et al. 2009)).

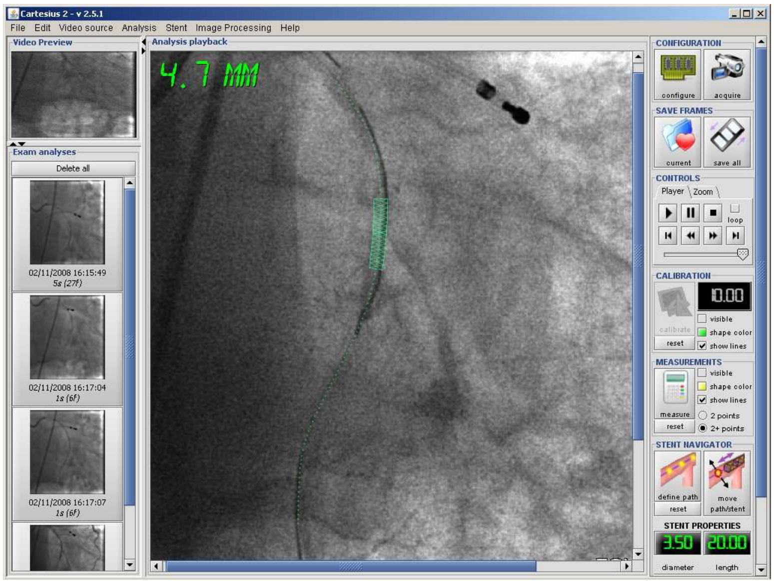

Fig. 2. A screenshot of the *Cartesio* tool

Managing patients in the Surgical Operative Units. We report on an experiences system for the management of data about surgical operative units. It is an integrated information system able to manage the cycle of management of patients into the surgical operative unit as depicted in Figure 1. Systems collect all the data of patient into an integrated EPR and support medical doctors for: (i) Patient Registration; (ii) Planning of Surgical Intervention; (iii) Patient Management during the Follow-Up and data anlysis for performance and clinical trial definitions.

System also provides statistical report about the number of patients, the geographical distribution of patients, the mean time needed for the execution of a surgical intervention, the number of patients for each medical doctor for administrative management.

3. Using EPR information for clinical protocol studies

A system for sharing oncological patient records among different regional hospital (Cannataro et al 2008) has been defined to wrap data from heterogeneous patient records (both hard-copy and digital) and to load data into XML-based EPR. A single simplified EPR allows operator of single hospital to query about the results of using adopted healthy protocols. Figure 3 shows a snapshot of query results performed from an hospital node looking for patients using protocol P in a lung cancer.

The system is composed by: (*i*) a data wrapper which is in charge of collecting data from singl EPR-data sources; (*ii*) a module able to map information into an XML-database storing subset of EPRs; (*iii*) a query engine able to compose and distribute queries among nodes of a peer-to-peer network; (*iv*) a security and update management module able to guaratee privacy and data updates (i.e., keeping updated information in XML database). The system allows to retrieve information on clinical protocol application and results (such as follow up information) improving cooperation among oncological health structures.

The system has been based on the definition of an ad-hoc EPR simplified data structure (Cannataro et al 2008), with the advantage of avoiding to impose standard to clinical structures that have currently their own systems, while solving the problem of sharing the results and feedbacks of using clinical drugs and protocols.

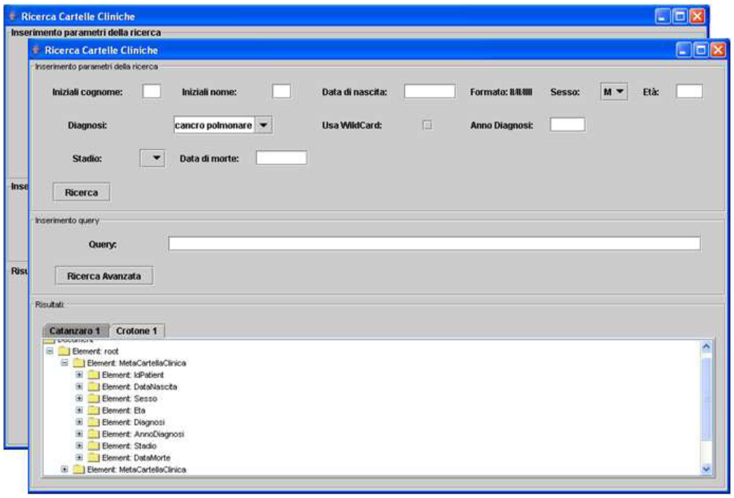

Fig. 3. GUI of SIGMCC System

4. Remote control and early-detection systems

Two developed projects, namely ReVA and Hearing Project, have been developed to support clinical analysis of the status of hearing human tract and vocal tract. Moreover, patients may monitor their status linking via Web pages and loading feedback on system directly from home. The rationale of both projects is the use of a distributed architecture (client-server) enabling the performing of medical test by the patients themselves. Such tests can be used for a first consult or for monitoring patients during the follow-up of surgical interventions. The results of the analysis, integrated with the patients clinical information, can be used for early detection and to monitor parameters.

The Hearing system (see Figure 4) has been developed on indication of the Unit of Audiology of University Hospital and aims to perform a pre-screening of hearing status of patients that may give feedbacks by using keyboards on a web application by listening to a contionusly increasing audio signal. The architecture of the system is based on the following modules:

- Presentation Module: It is based on a web interface between the user and the analysis module;
- Acquisition Module: It is responsible for collecting clinical data of patients;
- Screening Module: It implements the hearing test;
- Data Analysis Module: It elaborates the analysis results provided by the Screening Module;
- Data Management Module: It is responsible for the management of clinical data;
- Database: it stores all the data of the application.

Currently the system allows to use an external system for improving their hearing, to test and monitor the status of the system to give indication to medical doctor for parameter tuning.

The ReVA system aims to check the status of the vocal tract of the patient by using a web based remote system. The system is able to measure parameters associated to vocal signal and to measure voice quality, e.g. the frequency of fundamental tone or the pitch, and to associate data quality to presence of anomalies. ReVA (Amato et al. 2009) is a distributed architecture for collecting voices of different patients showing the ability to correctly discriminate among voices from healthy patients and voices from diseased patients. ReVA in particular enables patients to access the system through a web based interface, to register the voice and then to have the result of classification. Moreover, the loading of the voices of healthy patients could lead to the accumulation of a large amount of voices wasting computational resources.

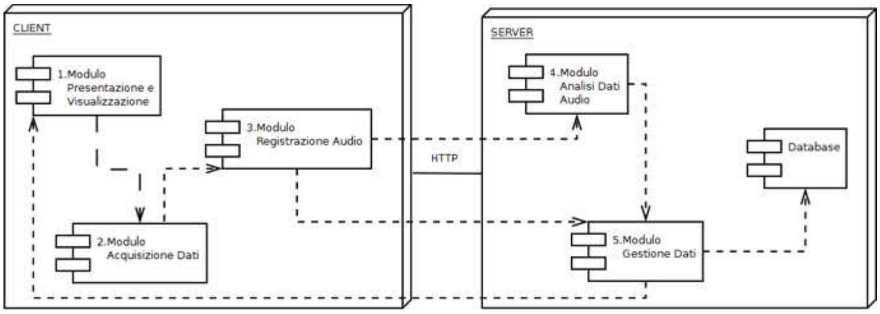

Fig. 4. The Hearing Project Distributed Architecture

- Presentation Module: it represents the Web-based Graphical User Interface as depicted in Figure 5. Through this module users finds all the ReVA services;
- Data Acquisition Module: it is responsible for the collection of clinical data of patients into an EPR.
- Voice Registration Module: it is responsible for the registration of voices of patients;
- Voice Analysis Module: it implements the algorithm of analysis;
- Data Management Module: it it responsible for the actions that are generated from the previous module (e.g. alerting medical doctors in case of the individuation of a voice problem);
- Database: it is responsible for the collection of all the data of the system.

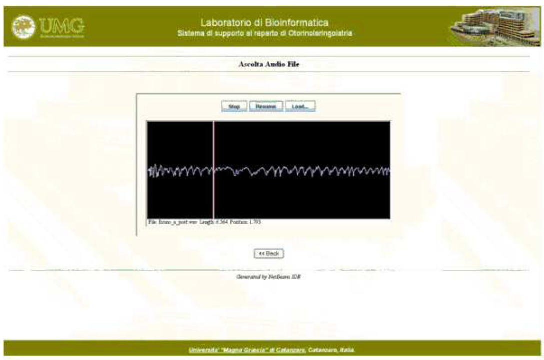

Fig. 5. Vocal system web Interface

5. Epidemiological models for data analysis

For statistics but also for correlating diseases with regional and environment data, we developed a web based geographical databases system, called Geomedica to associate information about disease diffusion with layers land information (such as landfills). The system (Tradigo et al 2010) is a framework for applying GIS technologies to the analysis of clinical data containing health information about large populations. By using the Geomedica module, existing clinical data can be geocoded, by associating tuples related to some geographical position with the coordinates of a map, and analyzed or queried using both SQL-like languages and web-based graphical user interfaces. The interfaces have been designed to allow non-expert users to easily query data with a *query-by-example* GUI.

Experiments were performed on the system for discoverying clinical and diseases indicators. Test queries performed on available datasets were able to correctly correlate health data about patients with geographical features (e.g. points of interest, boundaries, coastlines vectors) and to visualize the geographical distribution of diseases and health-related phenomena on a cartographic map. Moreover, administrations responsible of managing reimbursement processes are interest in Geomedica system to study association among regional hospitals services with respect to patient residence (e.g., correlation between service requirements and distance from hospital).

6. Conclusion

The need of computer-based tools in clinical applications is broadly recognized. Nevertheless, existing systems are still immature for what concern usability and popularity among interested audiences often due to data integration problem and interoperabilty. Our experiences is that many efforts need to be done to integrate data and information in commonly adopted platforms to share information about diseases cure and preventions.

Acknowledgment

We thank Ciro Indolfi, Leonzio Fortunato, Aldo Garozzo, Nicola Lombardo, Alessandro Puziello of Faculty of Medicine for specification and testing. We also thanks Mario Cannataro and Pietro Guzzi for collaboration and Giuseppe Tradigo for all his support.

References

[Cannataro et al 2007] M. Cannataro, P. H. Guzzi, T. Mazza, P. Veltri *Using Ontologies for Preprocessing and Mining Spectra Data on the Grid*, Future Generation Computer Systems Special Issue on Data Mining in Grid Computing Environments, Vol. 23, n. 1, pp. 55-60, January 2007.

[PRIN Gendata2020] Data-Centric Genomic Computing (GenData 2020) Prin Project.
Ministero dellUniversit e della Ricerca PRIN 2010-2011.

[Cannataro et al 2008] Cannataro M. , Talia D. , Tradigo G. , Trunfio P. , Veltri P. , *SIGMCC: a System for Sharing Meta Patient Records in a Peer-to-peer Environment*. Future Generation Computer Systems, 2008, Vol. 24, n. 3, pp. 222-234.

[Amato et al. 2009] F. Amato, M. Cannataro, C. Cosentino, A. Garozzo, N. Lombardo, C. Manfredi, F. Montefusco, G. Tradigo, P. Veltri, *Early Detection of Voice Diseases via a WebBased* System, J. Biomedical Processing and Control, 4 (2009) 206211, 2009.

[Indolfi et al. 2009] C. Indolfi, M. Cannataro, P. Veltri, G. Tradigo *Cartesio: A Software Tool for Pre-implant Stent Analyses*, ICCS (1) 2009: 810-818

[Cannataro et al. 2007] M. Cannataro, G. Cuda, M. Gaspari, S. Greco, G. Tradigo, P. Veltri. *The EIPeptiDi tool: enhancing peptide discovery in ICAT-based LC MS/MS experiments*, BMC BIOINFORMATICS 2007 vol. 8:255 ISSN: 1471-2105.

[Quaresima et al. 2008] Barbara Quaresima, Telma Crugliano, Marco Gaspari, Maria Concetta Faniello, Paola Cosimo, Rosa Valanzano, Maurizio Genuardi, Mario Cannataro, Pierangelo Veltri, Francesco Baudi, Patrizia Doldo, Giovanni Cuda, Salvatore Venuta, Francesco Costanzo, *A proteomics approach to identify changes in protein profiles in serum of familial adenomatous polyposis patients*, Cancer Letters 272(1):40-52. 2008.

[Cannataro et al 2007b] Cannataro, A. Barla, R. Flor, G. Jurman, S. Merler, S. Paoli, G. Tradigo, P. Veltri, C. Furlanello, *A grid environment for high-throughput proteomics*, IEEE Transaction on NanoBiosciences, 6(2):117-123. 2007.

[Cannataro et al 2010] M. Cannataro, Pietro H. Guzzi and P. Veltri, *IMPRECO: Distributed prediction of protein complexes*, Future Generation Computer Systems 26(3):434-440. 2010.

[Guzzi et al 2010] Mario Cannataro, Pietro Hiram Guzzi, Pierangelo Veltri *Using ontologies for querying and analysing protein-protein interaction data* Procedia CS 1(1): 997-1004 (2010). International Conference on Computational Science (ICCS 2010), Amsterdam, The Netherland, May 31-June 2, 2010.

[Cannataro et al. sighit2012] Mario Cannataro, Pietro Hiram Guzzi, Giuseppe Tradigo, Pierangelo Veltri
On the choice of centralized vs decentralized systems for EPR in hospitals. SIGHIT Record 2(1): 19 (2012)

[Vizza et al] Patrizia Vizza, Veltri Pierangelo, Giuseppe Lucio Cascini, *A ROI-based algorithm for the classification of PET neuroimages* Workshop on Biomedical and Health Informatics, in conjuction with IEEE International Conference on Bioinformatics and Biomedicine

[Tradigo et al 2010] Giuseppe Tradigo, Pierangelo Veltri, Sergio Greco
Geomedica: managing and querying clinical data distributions on geographical database systems
ICCS, 2010 Proc of the International Conference on Computational Science, 2010