An ecosystem for Precision Medicine: Closing the Cycle from Doctor to Patient*

Irini Fundulaki[1,2], Tzanina Saveta[1], Vassilis Papakonstantinou[1], and Yannis Roussakis[1]

Institute of Computer Science-FORTH
fundul@ics.forth.gr
National Infrastructures for Research and Technology (GRnet)
fundul@admin.grnet.gr

1 Extended Abstract

According to the Precision Medicine Initiative[1], precision medicine is *"an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person"*. Precision Medicine allows treating doctors to decide on the appropriate treatment for a patient who is suffering from a specific disease and who has a specific genotype. The Precision Medicine approach differs from the *traditional* approach in which the treatment of diseases follows a "one-size-fits-all" approach focusing on the average person without considering the differences exhibited by the various patients. Precision medicine is a model in which the individuals phenotype and genes, environment, and lifestyle are used as additional layers of patient data in disease treatment and prevention plans.

All Precision Medicine strategies include the use of decision-making processes based on the use of *biomarkers*. Biomarkers can be used to augment and refine traditional medical data to provide a higher level of specificity for disease prevention and patient treatment. The knowledge of the patient's genotype is crucial in providing the appropriate treatment to her. From this perspective the field of *'omics'* has played a very important role in the development of precision medicine. There are different types of *'-omics'* data: *(i) genomics* that study the genome, or sets of genes (panels), *(ii) epigenomics* that focus on genome-wide characterization of reversible modifications of DNA or DNA-associated proteins *(iii) transcriptomics* where RNA genome-wide levels are studied, both from a qualitatively and quantitatively perspective (how much of each transcript is expressed), *(iv) proteomics* used to quantify peptide abundance, modification, and interaction *(*v) *metabolomics* that simultaneously quantify multiple small molecule types *(vi) microbiomics* that investigate all the microorganisms of a given community together and *(vii) radiogenomics (imaging genomics)* that focus on the relationship between imaging phenotypes and genomics.

* Supported by the National Network For Genetic Cardiovascular Diseases Study And Prevention Of Sudden Death In The Young On The Basis Of Precision Medicine (GR ıCARDIACNET)

[1] *https://obamawhitehouse.archives.gov/precision-medicine*

Genomics are used for diagnosis in many medical fields such as oncology, cardiology, nephrology among others, and are explored in the field of neurology and most importantly in the case of neurodegenerative diseases. Proteomics and metabolomics, are now often incorporated into the everyday methodology of medical researchers. The *integration* of different *'-omics'* data (*multi-omics*) is used to elucidate potential causative changes that lead to disease, or the treatment targets, that can be then tested in further molecular studies.

Producing solely *'-omics'* data and storing them in a repository *does not suffice for either clinical or research purposes*. Such data are useful only when they are associated with the patient's *clinical information*. For example, in the case of *inherited cardiac diseases*, the geneticist can determine whether a mutation is *pathogenic*, *non pathogenic*, of *unknown significance (VUS)* only if she is aware of the patient's clinical data such as her symptoms, comorbidities, and family history of heart disease or sudden death in the family or other related diseases. In order to serve in an optimal way the researchers (clinicians, geneticists) it is absolutely necessary to *interconnect '-omics'* data with the *patient's clinical data*. *To support such interconnection, one needs to follow a patient-centric design for the ecosystem of technologies and tools that will be used to serve the objectives of Precision Medicine.*

In Greece currently there are three initiatives for Precision Medicine: (a) The Hellenic Precision Medicine Network on Cancer (August 2018 - August 2020, extended to May 2021) (b) The National Network For Genetic Cardiovascular Diseases Study And Prevention Of Sudden Death In The Young On The Basis Of Precision Medicine (July 2019 - July 2021) (GR iCARDIACNET) and (c) the National Research Network for Genetic Neurodegenerative Diseases (May 2020 - May 2022) (GR DeGeNEURONET). The ICT objectives of the last two initiatives are:

- the creation of *registries* for the *Inherited Diseases of the Heart* (Cardiomyopathies, Aortopathies and Channelopathies) and *Neurodegenerative Diseases* (Parkinson's, Huntington Disease, Amyotrophic lateral sclerosis (ALS), Alzheimer's, Multiple Sclerosis) that store *clinical patient data*
- the creation of a *Genomic Data Repository*
- the creation of an *Epidemiological Map* that will lead to the identification of populations with pathogenic mutations.

The utlimate goal of both networks is to *use the data stored in the systems to find populations who carry a specific genetic mutation and break the chain of the inherited disease.*

The design of all ICT systems is based on the *requirements* provided by expert clinicians and geneticists who perform the samples' analysis, as well as epidemiologists who want to be aware of the epidemics of a disease in order to decide on actions that must be undertaken in order to address the challenges presented. To be able to *integrate clinical* and *genomic* data, we follow a *patient-centric* approach in which each patient is *uniquely identified by her social security number* known only to the treating doctor. The systems that we have built to address the requirements of the two initiatives are the following:

- a *registry* for inherited diseases of the heart for GR iCARDIACNET for clinical patient data. The registries retrieve patient information from the National Social Security Registry (ATLAS) and support the *secure stotage*, and *controlled access* to sensitive health data.
- an *e-prescription* system that allows the treating physician to *order genetic tests* for the patient
- an *e-lab_book* system used by the lab personnel in the place of the traditional lab book to support the *complete life cycle* of *sample analysis*. It supports
  1. the sample recipient in recording the sample's arrival date and its appropriateness (b) the technical supervisor to assign the analysis to a geneticist (c) the scientific supervisor to upload and sign the genetic report and (d) the treating doctor to follow the progress of the analysis. The system follows the ISO15189 standard that specifies the quality management system requirements particular to medical laboratories.
  2. a *genomic data repository* that stores for each analysed sample the produced genomic primary data (FASTQ, BAM files), and secondary data (VCF and annotated VCF files), the genetic report and metadata thereof. The metadata support the efficient retrieval of primary and secondary data in the cases in which re-analysis is necessary.

The ecosystem of tools shown in Figure 1 is currently used by cardiologists to perform clinical diagnosis and research for hereditary diseases of the heart in the context of GR iCARDIACNET. A cardiologist stores the clinical data of the patient in the registry of inherited diseases of the heart (`iCARDIAC Registry`). Once sufficient information is gathered, the treating physician can *prescribe genomic tests* for the patient using the `E-prescription system` where he chooses the appropriate *panel (set of genes)* associated with the patient's diagnosis. The prescription is then sent to the *research lab*, responsible for performing the analysis along with the necessary patient's clinical information so that the lab's scientific supervisor can provide the genetic report. Note that the patient's personal data (name, social security number, address, date of birth) is not sent to the lab: the lab personnel know only the system's internal identifier provided for the patient. The scientific supervisor can then upload the primary and secondary data produced by the analysis of the sample to the `genomic data repository` along with their metadata and the genetic report which is subsequently sent to the treating physician. The same ecosystem of tools will be used in GR DeGeNEURONET(tailored for neurodegenerative diseases). A more detailed presentation of the aforementioned systems can be found at *https://sites.google.com/view/icardiacnet-services/* (in Greek).

In order to protect the sensitive health data we designed and implemented always in collaboration with the MDs, the Data Protection Officer of the Network and the Legal Team, a *role-based access control framework* that implements rules for *roles* (MD, Lab Personnel), *group members* (Member of Supervising Committee, Scientific Committee, Genetic Tests Approval Committee, ...) and *individuals*. A user can be assigned more than one roles, and can be a member of more than one groups. In addition to the access control framework, we

anonymize the patient's data: we assign to the patient a unique identifier used in the complete data lifecycle that cannot be used to trace back the patient. For instance, the treating doctor has access to all personal and clinical data of the patient, but the lab personnel does not have access to personal data (i.e., she has access to the patient identifier and to a subset of patient data that cannot be used to identify her such as date of birth, place of origin).

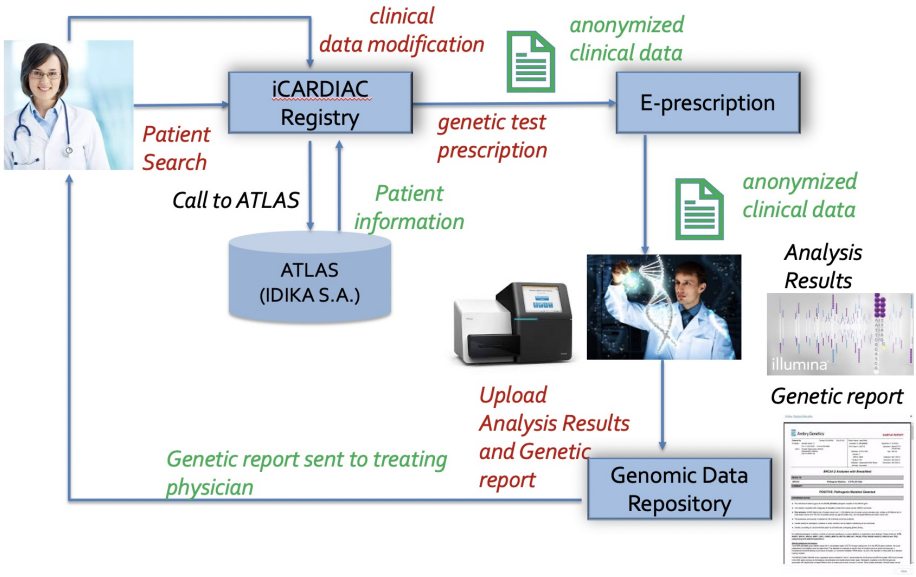

Fig. 1. ICT Ecosystem for GR iCARDIACNET

In addition to the above, we installed all the systems in the secure infrastructure of GRnet (National Infrastructures for Research and Technology) to which only authorized users have access. We use GRnet virtual infrastructure and we have deployed the aforementioned systems on different virtual machines (see Figure 2); data level encryption is also supported supported. Users are authenticated with Keycloak rather than individual applications. Keycloak provides Single sign-on (SSO) and single log-out (SLO) that allows a user to log in with a single ID and password to any individual application and the user passwords are encrypted using SHA512 encoding.

2 Building an ICT Health Ecosystem: Lessons Learned

The objective of an *ICT Health Ecosystem* is to create a *holistic view of patients data*; in order to achieve this, we follow a *patient-centric design* that supports the integration with *National Registries* (National Registry of Blood Donors,

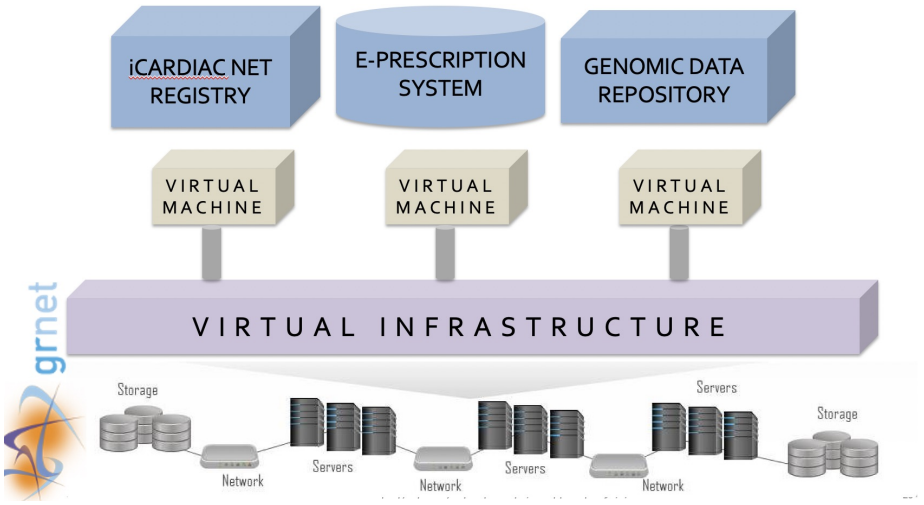

Fig. 2. Virtual Infrastructure of GRnet

Registry for the Chronically Ill, Registry for Diabetic Patients, ...), and *e-health* systems (e-Prescription, Private Health Record, Hospital Information Systems). Following such a patient centric design there are benefits to the patient who does not need to carry her health record around (extremely important for people with comorbidities), to the healthcare system since such integrated data can be used for *population genetics*, for building *epidemiological maps* that can be subsequently used for *disease prevention* and *diagnosis*, and to the researcher (clinician, bioinformatician, geneticist) who has a tremendous amount of data available.

There are different phases one should go through when building an ICT health ecosystem. Gathering user requirements is not an easy task since MDs are extremely hard to find and gathering their needs takes a significant amount of time. The advice is not to wait for them, but to try and understand what they need and propose them how their needs can be addressed by the systems that will be developed. Furthermore, it is utopian to think that the "waterfall model" for system development would do: MDs want to see actual systems and use them in order to understand if they address their needs. A report describing in detail the system functionality will not do. Hence the advise here is to continuously build prototypes and make them available to the users. Another important issue that must be addressed is the issue of consent that dictates the extent to which patient data (clinical and genomic data) can be accessed and reused and by whom, inside and outside the context of the project, and for which purpose. Consent should set out potential risks and benefits to participants, as well the data anonymization procedures to be undertaken. The consent forms should

be specified by the Data Protection Officer (DPO) and they Legal Team. To do this, they should know the data flow and the processes implemented in the ICT Ecosystem. A result of this effort is the specification of the access control framework, i.e., who has access, and for each purpose and to what type of data. This task is not easy since the MDs strongly believe that their patients' data is their data and should have total control over them (and unfortunately not only they). According to the Greek Legislation, a treating doctor is obliged to share for *diagnostic purposes*, their patients' data. The consent form should specify clearly the potential use of the patients' data for research purposes by members of the academic and research community.