City Research Online

City St George's, University of London

Citation: Van Gorp, P. & Comuzzi, M. (2014). Lifelong personal health data and application software via virtual machines in the cloud. IEEE Journal of Biomedical and Health Informatics, 18(1), pp. 36-45. doi: 10.1109/jbhi.2013.2257821

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/4046/

Link to published version: https://doi.org/10.1109/jbhi.2013.2257821

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Lifelong Personal Health Data and Application Software via Virtual Machines in the Cloud

Pieter Van Gorp, Marco Comuzzi

***Abstract*—Personal Health Records (PHRs) should remain the lifelong property of patients, who should be enabled to show them conveniently and securely to selected caregivers and institutions. In this paper we present MyPHRMachines, a cloud-based PHR system taking a radically new architectural solution to health record portability. In MyPHRMachines, health-related data and the application software to view and/or analyze it are separately deployed in the PHR system. After uploading their medical data to MyPHRMachines, patients can access them again from remote virtual machines that contain the right software to visualize and analyze them without any need for conversion. Patients can share their remote virtual machine session with selected caregivers, who will need only a Web browser to access the pre-loaded fragments of their lifelong PHR. We discuss a prototype of MyPHRMachines applied to two use cases, i.e. radiology image sharing and personalized medicine.**

***Index Terms*—Personal Health Record, Cloud Computing, Electronic Health Record, Radiology, Personalized Medicine.**

I. Introduction

In a recent review paper, Kaelber et al. define a Personal Health Record (PHR) as *"a set of computer-based tools that allow people to access and coordinate their lifelong health information and make appropriate parts of it available to those who need it"* [1]. PHRs should be *portable*, i.e. remain with the patient, contain lifelong information, and should not be restricted by file formats or other local issues [2]. In other words, they are Electronic Health Records (EHRs) that are owned by patients. These are usually opposed to hospitals' Electronic Medical records (EMRs), which only contain medical data generated within one specific care institution.

The research question addressed by this paper is *"How can we design a sustainable and privacy-compliant IT infrastructure that facilitates at least for a patient lifetime and across the boundaries of care institutions and medical specialisms (1) the storage of raw PHR data and (2) the use of this data with specialized software?"*

Sustainability in this context refers to the financial and political aspects of the health care and software industries. Point (1) focuses on raw PHR data since care institutions may not be able or willing to provide their EHR data in "one" standardized PHR format. Tang et al. mention in their PHR adoption barrier analysis that *"(US) Government can play a number of important roles in increasing PHR use. At the infrastructure level, the federal government could catalyze development and adoption of data and interchange standards for key PHR content areas."* [3]. Such standards are useful and slowly emerging, but we argue that regardless of such evolution patients should already be empowered with the ability to manage their own (potentially raw) data. With point (2) we aim at so-called *functional* interoperability (i.e., *"the ability of two or more systems to exchange information so that it is human readable by the receiver"* [4]). Concretely, we aim at providing patients (and their trusted caregivers) remote desktop or tablet computer access to all their PHR data, and support this access by the software that matches the data format. Since we do not tackle semantic data integration in this paper, one can more specifically label this as health record mobility and portability.

Cloud computing offers unique opportunities for supporting long-term record preservation [5]. In this paper we present MyPHRMachines, a cloud-based PHR system that answers our research question. One of the agreed key requirements for shareability of the EHR is to break the nexus between the EHR and the EHR system [4]. The MyPHRMachines architecture clearly separates PHR data from the software to work with this data. This paper demonstrates how this creates novel opportunities for market of PHR software services without compromising patient privacy.

Commercial PHR systems positioning themselves within the cloud computing paradigm are emerging. For example, SeeMyRadiology [6] enables patients to upload their medical images and then selectively share these with caregivers. Unfortunately, such so-called Software-as-a-Service (SaaS) systems are typically (1) specialized for one medical function and (2) specifically programmed for web browsers. The SeeMyRadiology example indeed consists of a DICOM viewer that has been programmed in HTML 5 and related technologies. MyPHRMachines is an academic prototype that is more generally applicable since it exposes to its users the so-called Infrastructure-as-a-Service (IaaS) tier of cloud architectures [7]. In a nutshell, the system provides infrastructure to (1) store and share (subsets of) patient data and (2) deploy and use specialized software in remote Virtual Machines (VMs).

MyPHRMachines allows patients to build personal health records which are robust across the *space* and *time* dimensions:

Space. Patients relocating or simply traveling across different countries during their lifetime will always be able to reproduce their original health records and the software required to analyze/visualize those. This is often currently not possible because of the high functional and architectural heterogeneity of health care information systems across different countries/states [8].

P. Van Gorp and M. Comuzzi are with the School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, e-mail: p.m.e.v.gorp@tue.nl, m.comuzzi@tue.nl .

Manuscript received XXXX 00, 2012; revised XXX 00, 2012.

Time. As technology evolves, application software typically becomes obsolete. On the server-side MyPHRMachines prevents deprecation problems by virtualizing execution environments holistically. The software to create the *idealized* environments on contemporary hard- and software is maintained by big vendors [9], regardless of the MyPHRMachines-specific extensions. On the client-side, MyPHRMachines does rely on contemporary web technologies, but only to realize a generic remote desktop client. Hence, also client software maintenance is decoupled from the number and complexity of PHR software services.

PHR systems typically offer functionality to share, visualize, and analyze PHR data [10]. MyPHRMachines also enables its users to share software to work with the health-related data, keeping data and software clearly separated in the system architecture. Having separate data and functionality also allows a finer grained delegation of access to different stakeholders. Specifically, MyPHRMachines allows patients to selectively reveal health information to other stakeholders and it guarantees that, once shared with a stakeholder, health information cannot be improperly stored. First of all, the software specialists deploying third party PHR services to MyPHRMachines never get access to patient data; secondly, even those whom have been given access to patients' remote VM sessions cannot use or store the data/software beyond the time frame that is offered by the session owner (i.e., a patient or his or her guardian). Currently available PHR systems do offer selective delegation mechanisms, but pose fundamental privacy threats in this context. Examples of typical threats characterizing currently available PHR systems are discussed later while reviewing related work.

Before discussing the implementation and application of MyPHRMachines, we introduce two use cases exemplifying the potential for innovation in health care brought about by our prototype. The first use case concerns radiology image sharing, showing how MyPHRMachines can be used to build and maintain efficiently a lifelong PHR of radiology images. The second use case concerns genomic data analysis in the context of personalized medicine, showing how the separation between PHR data and PHR functionality allows finer grained privacy-related control over PHR data access and utilization.

The paper is organized as follows. Section II introduces the two use cases in different application scenarios that we use to exemplify the functionality of MyPHRMachines. Section III presents the design of the prototype and discusses its implementation. MyPHRMachines's potential for innovation in the health care domain and the limitations of our approach are discussed in Section IV, whereas related literature is reviewed in Section V. Finally, we draw our conclusions and discuss future work in Section VI.

II. Motivating Use Cases

In this section we introduce two use cases to support the description of MyPHRMachines implementation in Section III. The first use case puts the accent on the *spatial* and *temporal* pervasiveness aspects whereas the second use case is used to highlight privacy-related aspects.

A. Radiology: Lifelong back injury condition

The first use case considers the case of non-severe scoliosis (spine curvature of less than 20 degrees) and discopathy (intervertebral disk fracture) due to physical traumas. The diagnosis and treatment of such conditions is not an easy task and physicians often tend to waive intensive and expensive treatment referring the patient to physiotherapy or even commercial fitness clubs for palliative therapy. The condition, however, may remain latent for years and reappear in the long run. The decision to start a professional, long-term revalidation program may be postponed too long especially when caregivers lack access to prior scans and analyses.

This use case concerns the medical history of a real patient of the Belgian health care system affected by the above mentioned condition. For reasons of privacy, the case has been made anonymous. The medical history of the patient can be synthesized as follows:

1) At the age of 15 the patient injures for the first time his back in a home maintenance task and receives chiropractor care to relieve acute stress between the shoulders;
2) at the age of 18, the patient experiences a wintersport accident, leading to a severe hematoma in the lower back; a RX scan is made and analyzed at the foreign holiday location, after which the patient is sedated and transferred to his home country, where he undergoes various medical scans (RX, MRI, bone scan with chemical tracer); the patient is referred to kinesitherapy for four months and is discharged with the instructions to continue performing regular sports activities, which should drain the hematoma and relieve the pain;
3) after seven years (at the age of 25), the patient is still bothered by the hematoma consequences and visits a physiotherapist, the patient undergoes a new RX and MRI scan but the physiotherapist does not find noteworthy problems; the patient is also referred to a neurologist, who orders a new bone scan (the old one being at another hospital and not retrievable); the bone scan again does not reveal bone traumas.
4) at the age of 30, the patient visits another team of specialists (an orthopedist cooperating with a neurosurgeon working outside of a hospital). The orthopedist again asks for RX and MRI scans but also inspects the previous MRI scan (which is supplied on a laptop by the patient). The specialist discovers the discopathy (intervertebral disk fracture), which may have been caused by the wintersport accident (12 years before). Since surgery only has an 80% success rate, the patient engages in an intensive lower back revalidation program. This will also reduce the pain caused by the scoliosis.

The diagnosis and treatment of our patient could be improved in several ways. First, for minimizing the treatment costs and patient stress, the patient should not have undergone more than once the same scan or, generally, examination, unless strictly required for formulating a diagnosis. Second, our patient has never been able to show his entire medical history to caregivers. Specialists, in fact, often based the diag-

nosis only on the exams that they ordered. In this regard, the diagnosis of the discopathy could have been anticipated if the patient would have been able to consistently show his complete medical history to all the specialists and institutions that he visited. Eventually, our case shows that IT infrastructures of hospitals and GPs are not sufficiently integrated yet nationally and especially internationally to provide a lifelong EHR for our patient.

Note that not only are longitudinal health records useful for assessing the evolution of individual patients better, but they also open doors for "big data" clinical research, which may generate much stronger medical evidence than conventional trials (such as [11] for scoliosis).

B. Personalized Medicine: Genomic diagnostics

Ginsburg et al. describe their vision of personalized medicine as follows: *"tailored care is given for every individual based on their specific, molecular disease will become the standard of care. In the prototypical office visit of 2015, the physician will examine a patient's genetic profile (stored on CD ROMs or equivalent), lifestyle, and results from objective molecular screening and monitoring tests. Algorithms, derived from previous research efforts, will be used to compute the likelihood that a patient develops a host of chronic diseases."* [12]. In this paper, we do not focus on the algorithms that are needed to realize this vision. Instead, we show why MyPHRMachines should be used instead of CD ROMs to realize the above vision, mostly focusing on the PHR data privacy issue.

In order to benefit from personalized medicine, a patient needs to get a digital representation of his/her genetic profile. This involves a one-time analogue to digital conversion (called DNA sequencing [13]). The cost of this process is dropping at such a dramatic rate that it can soon be expected to be a free service for citizens of developed countries [14]. The major issue in this context becomes quality of software services to give personalized medical advice based on a genetic string. Among such quality requirements, privacy plays a prominent role. Clearly, a patient's genomic data is quite privacy sensitive, as it may reveal intrinsic limitations that among others can have a negative influence on someone's career, mortgage negotiations, social relations, etc. While, in fact, market competition among diagnostic software services is likely to foster the quality of personalized diagnosis, an open market of such services can be deemed safe only if patient privacy is safeguarded at the platform level.

Hence, in this use case we consider the case of a patient who has already sequenced his or her DNA, that is, who has stored the DNA sequencing string (PHR data) in the PHR system. The patient would like to run different sorts of genomic data analyses on the DNA string. However, the patient is also concerned about the improper usage that the application software provider could make of the data. Data made available by patients can be sold to other commercial institutions, after which further control becomes very complicated. If leaked for instance to employers or insurance companies, the data may unwillingly influence the patient relationship with such institutions.

C. Requirements for a PHR system

In this section we discuss a set of requirements for PHR systems directly derived from our two use cases. In the next section we demonstrate that MyPHRMachines is able to satisfy these requirements comprehensively, whereas later in the paper, while discussing related work, we show that current PHR systems available commercially or in the academic literature can satisfy the requirements only partially or only in very specific application scenarios and configuration settings.

From the first use case of radiology, we derive a set of requirements capturing the need to build PHR systems that are robust across the space and time dimensions of information sharing for the patient. Specifically, requirement RA1 focuses on the *space* dimension, whereas RA2 captures the *time* dimension.

- **RA1.** PHR systems should allow patients to reproduce their medical data to any interested care institution, irrespective of the physical location of those and/or the maturity of their IT support;
- **RA2.** PHR systems should allow patients to reproduce their lifelong medical history to any interested care institution.

The requirements derived from the second use case of personalized medicine capture important privacy-related features to protect the owner of PHR data, that is, the patient. Privacy is about the ability of the data owners to selectively reveal data to interested actors and to be assured that, once shared, their data cannot be used improperly (e.g. for commercial purpose or for extracting sensitive information about the patient).

- **RB1.** PHR systems should allow patients to selectively share their PHR data to interested care institutions;
- **RB2.** PHR systems should ensure that PHR data shared by patients will not be used improperly by the care institutions with whom such data are shared or by providers of application software.

III. Design and Implementation of MyPHRMachines

In this section we first present the technical architecture of our prototype. Then, we discuss the implementation of our two use cases.

The main idea behind MyPHRMachines is to leverage the cloud for allowing patients building their own personal health data repository and share these data with different care institutions. In the current implementation, patients have to manually upload the data they obtained from care institutions, e.g. in a DICOM CD, in the repository. In a near future, we envision that care institutions could directly push patient data to the repository. Once stored in MyPHRMachines, patients can flexibly share these data with any other care institution or interested stakeholder. Access to MyPHRMachines, in fact, requires only a Java-enabled browser, and access to a selected part of the repository can be easily granted by patients to any care institution, e.g. a GP, a hospital, or an insurance company. Moreover, MyPHRMachines also allows care institutions to make available specialist software required to view and/or analyze health-related data. In this way, caregivers need not

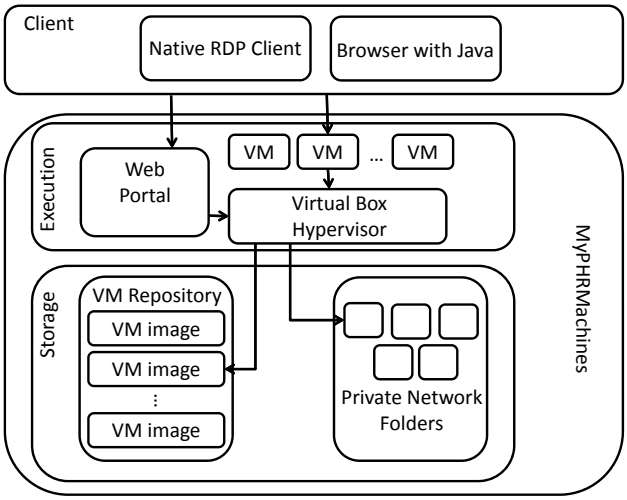

Fig. 1. Technical architecture of MyPHRMachines.

be able to run specialist software, since they can get access to this software directly from the cloud.

A. Technical Architecture

Figure 1 shows the technical architecture of MyPHRMachines, identifying, besides the components constituting MyPHRMachines, also the components of the front-end *Client*.

The prototype reuses parts of SHARE [15], a mature system for making computational research results more accessible and reproducible. The key technological components have therefore undergone various development cycles, which adds to the robustness of MyPHRMachines technical architecture. On the one hand, MyPHRMachines excludes functionality developed for the SHARE-specific use cases (e.g. generating BibTeX code for conveniently citing a VM image from a research paper). On the other hand, MyPHRMachines required the development of new functionality specific to the PHR context (e.g. access delegation to a VM session). We also redesigned the user interface of the Web portal to become simpler and coherent to facilitate access by non-expert users.

Within MyPHRMachines, we distinguish between the *Execution* and *Storage* layers. Each Virtual Machine (VM) in the execution layer represents the virtualization of specific application software (or a software bundle) serving the purpose of either viewing or analyzing patients' health data. Patients can log into MyPHRMachines and decide which VM to load in a given session using a standard Web portal. The *Hypervisor* is a generic piece of software to start, stop, clone VMs, and control their Internet access. For our prototype, we decided to use VirtualBox, an off-the-shelf hypervisor. Being heavily used in several industries, VirtualBox benefits from periodic functionality updates and security reviews. Note that, as discussed more in depth later, the VMs for specialist software are stateless and deprived of Internet access.

The storage layer includes the repository of VM images, i.e. virtual disks containing a bootable operating system and additional applications. Patient-specific VMs are simple instances of these VM images. In order to publish new VM images, software vendors go through the following procedure: first, they clone an existing VM containing the right operating system and perhaps some additional libraries of interest through the MyPHRMachines Web portal. At this stage, only the vendor can instantiate the VM image and modify the VM image (install new software, adjust configuration files, etc.). Finally, the vendor "publishes" the VM image for other users of MyPHRMachines. Users cannot change the published VM image since any personal instance of a VM image is stateless. By keeping VM instances stateless, one can deploy updates at the VM image level, which is much more scalable and secure than trying to do this at the level of patient-specific VMs.

The labor cost of requesting a VM clone via the MyPHRMachines portal is negligible. Other labor costs relate to (1) uploading application executables to MyPHRMachines, and (2) configuring them in an instance of the new VM image. The first cost is unavoidable since by definition it becomes relevant any time a software vendor wants to deploy software to a cloud-based system. About the second cost, any IaaS-based approach would provide the same level of flexibility as MyPHRMachines. However, in more general IaaS platforms (e.g. Amazon EC^2), VM images would have to be cloned explicitly for each end-user. Also, end-users would be able to change the VM images, introducing huge maintenance costs. Instead, the MyPHRMachines approach of using stateless VM sessions (i.e., sessions that do not affect the VM image involved) avoids that cost problem by design.

The PHR data are stored into network folders, which remain private folders within the MyPHRMachines domain. Put differently, the VM-based architecture ensures that all patient data can remain server-side, on a trusted infrastructure. The latter feature, combined with stateless VMs deprived of Internet access, guarantees the privacy of the patients health-related data. In particular, even if software in one of the VMs is programmed with some sort of malware, this will not be able to push PHR data outside the network domain of MyPHRMachines.

Clients can view remote VM sessions using the Remote Desktop Protocol (RDP [16]). Therefore VM sessions can be viewed in any Java-enabled Web browser without installing any additional software, by using a simple applet-based RDP viewer. For operating systems not supporting Java, e.g. iOS, a native RDP client is required. All communications between the Web portal and the hypervisor are delivered via SSH, a secure and stable communication protocol that, among others, provides measures to prevent Man-In-The-Middle (MITM) attacks. MITM attacks can also be prevented for the HTTPS traffic between a client browser and the MyPHRMachines web server [17]. Since MyPHRMachines is currently deployed only as a prototype we have not invested yet in the required certificates issued by a major certification authority. MITM attacks can also be prevented for the RDP traffic: the MyPHRMachines hypervisor supports RDP over TLS [18] so both RDP client and server identity can be protected using certificates. Again, for the prototype deployment, certificates are not yet used. Certificate configuration needs to be handled once at the level of the web server and once at the level of

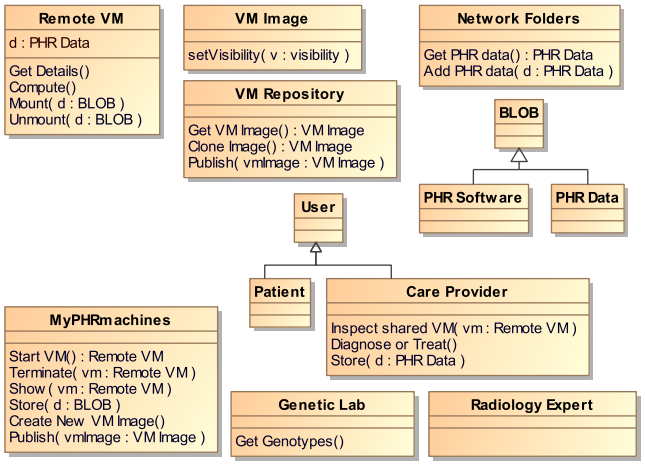

Fig. 2. Conceptual model of MyPHRMachines.

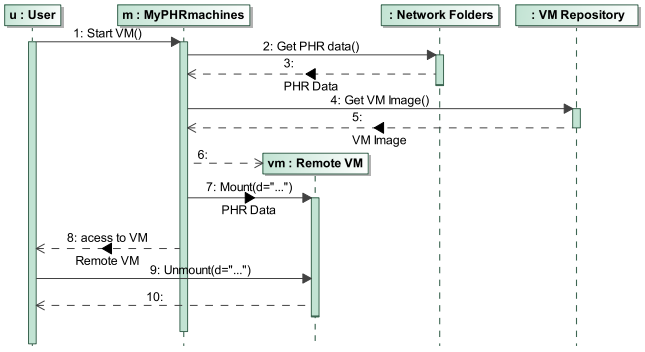

Fig. 3. Starting a new VM session in MyPHRMachines.

the hypervisor. VM-level certificates are not needed, so the scalability of the architecture is safeguarded.

The UML sequence diagrams in Figure 3 and Figure 4 show two typical interaction scenarios supported by MyPHRMachines, i.e. starting a new VM session and sharing access to a VM between a patient and a medical expert in a care institution. Figure 5 shows a less common interaction scenario involving software vendors. We include this diagram since it clarifies how MyPHRMachines differs from other cloud platforms. All sequence diagrams are based on the conceptual data model of MyPHRMachines shown in Figure 2.

When starting a new VM (see Figure 3), the platform retrieves the VM image from the VM repository and the PHR data from the private network folders. The PHR data are then mounted into the VM. By default, all PHR data of the logged in patient are mounted when a new VM is instantiated. Before

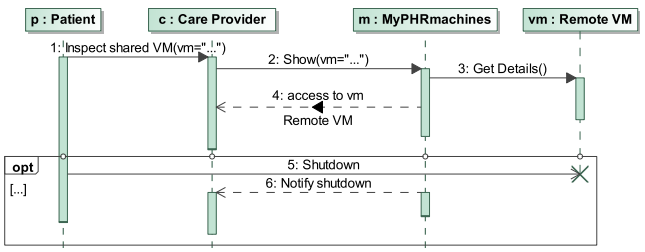

Fig. 4. Sharing a VM session in MyPHRMachines.

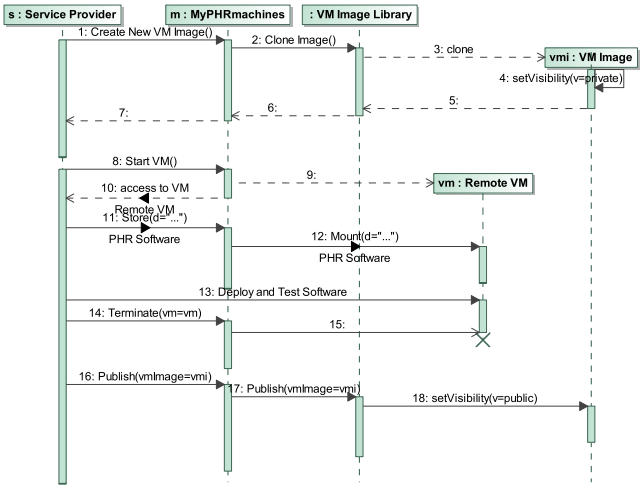

Fig. 5. Adding a new software service to MyPHRMachines.

sharing access to a VM, however, patients can selectively unmount those folders that are considered too sensitive to be shared with a given care institution.

Patients can instruct MyPHRMachines to forward by email a long, ciphered string identifier of a VM session to share with a specific care institution (see Figure 4). Using this identifier, the user at the care institution is able to access the VM with one click, even without having a system account (i.e., without the need to login). Patients may at all times decide to shutdown a VM, for instance in case they realize that the care institution to which access they granted access is misusing their PHR data.

The long string is based on applying a hash function to parameters of the VM session. The long strings can theoretically be guessed. Practically however, various system administrator techniques for blocking distributed denial-of-service attacks can be used to prevent scripted guessing, even when the attacker uses multiple machines. Moreover, even if an attacker correctly guesses such a string, that secret is valid only for the lifetime of one VM session.

One important downside of sending around URLs that provide direct VM access is that, without additional security measures, the access delegation messages could be intercepted by malicious Internet users. Fortunately, care institutions are likely to have secure messaging tools in place and therefore the access delegation message can be sent securely from the MyPHRMachines web server to the inbox of the caregiver. Therefore, we do not consider this as a major threat.

The workflow related to publishing new VM images has already been discussed. Figure 5 clarifies the details in sequence diagram syntax. Steps 1 to 7 involve setting up a new VM image. Steps 8 to 12 involve uploading application binaries. Steps 13 to 15 involve the installation and configuration of these executables. Interestingly, MyPHRMachines enables specialization among software vendors: some may specialize in setting up developer-friendly VM images with application infrastructure (e.g. a complex web and database server environment). These images can be offered to other software vendors, who want to specialize in offering end-user

oriented VM images. Steps 16 to 18 involve publishing a VM image to a library. The library concept is important both to separate the developer-oriented images from the end-user oriented ones, but also to organize end-user images in various more fine-grained categories (e.g. per medical condition or per insurance plan).

After having presented the technical architecture of our prototype, we can now go back to the requirements listed in Section II and discuss how MyPHRMachines addresses explicitly all of them.

About requirement **RA1**, PHR data can be stored by patients using the cloud storage provided by MyPHRMachines. When required, PHR data can be easily exposed to care institutions, e.g. a physician, as long as an Internet connection and a Java-enabled browser are available. As demonstrated before, access to VMs is granted by forwarding a ciphered VM identifier by e-mail to care institutions. About requirement **RA2**, PHR data will be available in principle forever within MyPHRMachines to be shared among patients and care institutions. Moreover, the application software required to view and analyze such data will also be always available. In particular, MyPHRMachines enables the virtualization of any type of operating systems and application software. These will remain available to patients and care institutions even when they become no longer in use or accepted in practice.

The requirement **RB1** is implemented as a feature of the Web portal. When launching a VM, in fact, patients can select only part of their PHR data currently available within MyPHRMachines to be shared with a given care institution. Finally, the requirement **RB2** is forced by design because, as we discussed before, VMs do not have Internet access and, therefore, the PHR data used by them cannot be pushed outside the domain of MyPHRMachines to pursue improper use. Having VMs without Internet connection may represent a limitation of our prototype. This issue is discussed more in depth in Section IV-A. It is however also an essential security strength: by cutting off internet access at the level of the hypervisor, MyPHRMachines ensures that even when end-users or applications tamper with the firewall settings of a VM, no harm can be done.

B. Use case implementation

Demo instructions for the two use cases are available on a companion Website [1]. In this section we provide a brief walkthrough of the use case implementations.

Figure 6 and Figure 7 show the interaction models of the radiology and personalized medicine use cases, respectively.

About radiology, the patient obtains PHR (radiology) data, e.g. a DICOM CD, from the radiology provider *r*. The sequence diagram in Figure 6 shows two options for loading PHR data into MyPHRMachines, i.e. by the Radiology provider (see the first *opt* block in the sequence diagram) and by the patient (see the second *opt* block in the diagram). Note again that automatic file transfers from PACS archives to MyPHRMachines network folders has not yet been implemented. Once the radiology data is in MyPHRMachines, the

[1]https://sites.google.com/site/myphrmachines/demo-phr

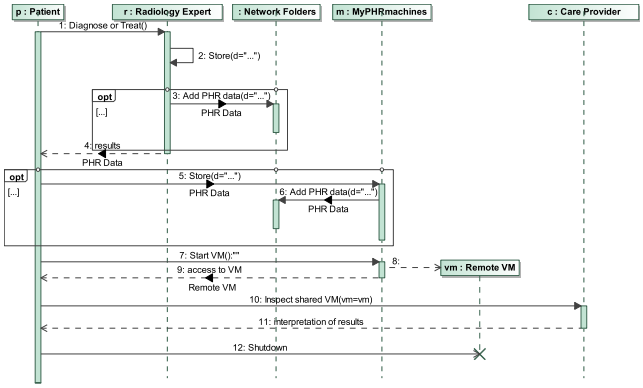

Fig. 6. Interaction model of the radiology use case.

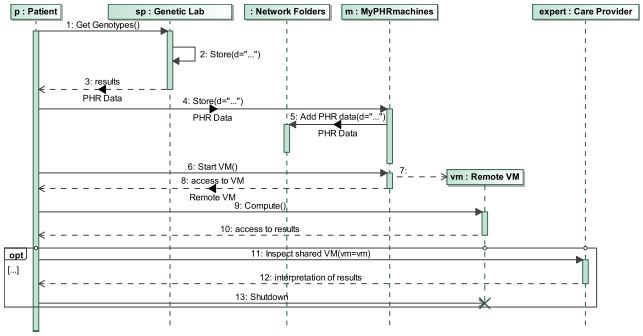

Fig. 7. Interaction model of the personalized medicine use case.

patient starts a VM and shares the access to this VM with provider *c*. Steps 10 and 11 in the diagram represent the case in which the patient delegates VM access to a care provider *c* which represents for this use case a physiotherapist from another hospital.

About personalized medicine (see Figure 7), the patient first acquires the DNA sequence from a specialized care institution (such as *baseclear* [19]) and then stores it into MyPHRMachines. In order to receive genetic counseling, the patient starts a new VM with software specialized for genome analysis and grants access to it to a medical expert. MyPHRMachines can also provide VMs with genome sequence file converters [20], but for the sake of simplicity we focus on genomic diagnostics in this paper.

Figure 8 shows the menu of the Web portal. The patient in this case has access to three VM images. Two of these images (the ones whose name starts with *Radiology*) can be used for the radiology use case while the third one is designed to support the personalized medicine use case.

For the radiology use case, one VM is designed to run the DICOM viewer provided on most DICOM CDs. The DICOM viewer is loaded by clicking on one of the virtual CD icons (see Figure 9). By loading different radiology scans into this VM, the patient will be able to reproduce his or her entire medical history (as far as radiology is concerned) to caregivers anywhere in the world. A second radiology related VM contains a specialized DICOM viewer that can also visualize DICOM data in case a viewer has not been embedded in the hospital-provided DICOM CD (or DVD).

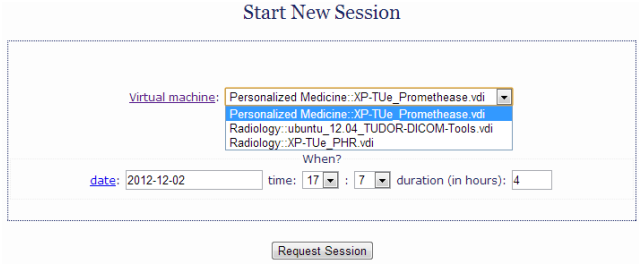

Fig. 8. The *Web Portal* component in the MyPHRMachines execution tier.

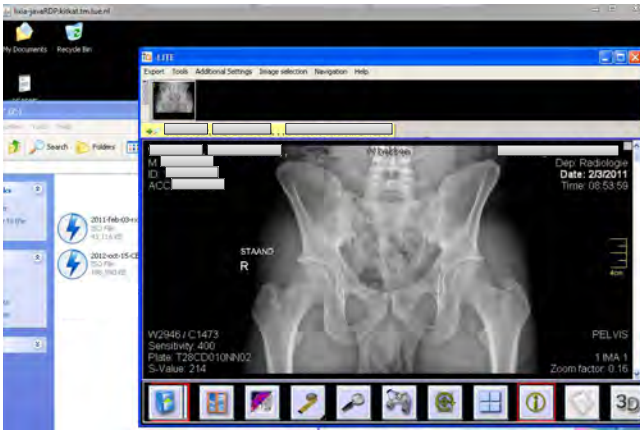

Fig. 9. Radiology scan within MyPHRMachines.

This is the case for example for CDs containing DICOM data of Cone Beam Computed Tomography (CBCT) scans.

The VM for the personalized medicine use case combines the DNA data of an anonymous patient available on the Internet with the open source Promethease software as application software, in this case to *analyze* the PHR data. Figure 10 shows an example report generated by Promethease within MyPHRMachines. Note that the information generated by VMs cannot by default be pushed out of the MyPHRMachines network domain by a possibly malicious implementation of the application software because Internet access is by default

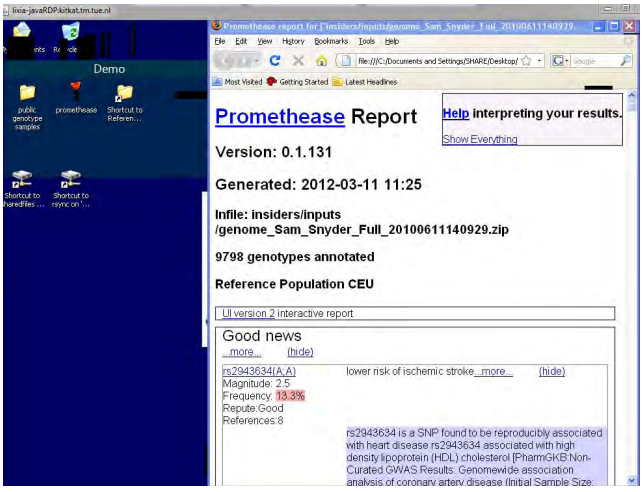

Fig. 10. Genomic data analysis within MyPHRMachines.

disabled for VMs. The Promethease software, however, happens to require an Internet connection for dynamically fetching the latest expert rules for genome interpretation. For such cases, MyPHRMachines VMs can be given Internet access through a virtual network proxy. MyPHRMachines platform administrators can define fine-grained policies, e.g. to give the trusted Promethease virtual machine access to the Internet address of the genomic expert rule repository. An alternative solution would be for the application software vendor to routinely update its provided VM image with the latest expert rules for genome interpretation.

For both use cases, access to a running VM can be delegated by the patient simply specifying the email address of the caregiver. The caregiver will receive an email with a secure URL to access the running the VM. In this way, access to patient health records can be delegated by patients to any care institutions or stakeholder requiring so. Following requirement **RB2**, the shared VMs do *not* enable one to download the patient-owned medical data. This functionality is implemented by enabling the Web portal to instruct the hypervisor to never give Internet access to shared VMs. Hence, as such, this functionality did not require any use case specific programming.

IV. Discussion and Limitations

In order for MyPHRMachines to become a viable solution to achieve health care cost reduction and quality of care improvement constantly advocated in modern societies [21], researchers will have to pay attention to several issues arising from the contextualization of MyPHRMachines in the complex health care ecosystem.

A useful frame of reference in this context is the one of institutional theory, which has often been used to address the shortcomings of technological and business innovation in health care [22], [23]. It predicates that organizations are often influenced by normative pressure, arising internally, e.g. relative power of physicians and administrative managers, or at the industry level, e.g. imposed reimbursement schemes, leading them to choose legitimated elements that have often the effect of directing attention away from task performance and social welfare. According to institutional theory, the relationship among *processes*, *people*, *business models* and our proposed solution, in particular, needs further investigation.

Regarding *processes*, we need to investigate how MyPHRMachines will impact administrative and clinical processes currently in place in health care institutions. For instance, administrative processes are usually driven by data available in local EMRs, which may be inconsistent with the data possessed by the patient. Another factor influencing the success of our solution can be the management of the coexistence of patients adopting and non-adopting personally-owned health care records, since we cannot assume complete penetration of such a technology without government sponsoring, at least in the initial transitory period.

Regarding *people*, MyPHRMachines represents a technological innovation that may disrupt current medical practice and patient behavior. As such, we need to investigate its

acceptance and possible adoption by different types of users, such as patients, physicians, or administrative personnel.

Eventually, regarding *business models*, research is required to understand how to make our solution economically profitable in the health care ecosystem. While, in fact, adopting our solution may reduce the cost of data exchange and exam retake, the costs related to the implementation and maintenance of patient-owned records has to be taken into account. Moreover, we argue that MyPHRMachines can become a success only by exploiting its complementarity to existing PHR and EMR systems. At least in the initial diffusion period, the use of the system should be suggested to citizens for whom the requirements addressed in Section II are particularly critical, such as business travelers or citizens in need of specific advanced analysis that could reveal privacy sensitive information.

A. Limitations

We distinguish the limitations of our work into the ones relating to the functionality of MyPHRMachines as currently implemented and the ones relating to the research method adopted for its evaluation.

About the functionality, MyPHRMachines is likely to lead to numerous personal application islands, in which each patient collects heterogenous PHR data and application software. This can lead to a very chaotic repository of health information and related functionality that can be very hard to maintain for the average patient. The issue can be overcome by a careful design of the interface of MyPHRMachines used by patients to upload, share, and, generally, organize their PHR data, which should be intuitive and hide technical details.

Another limitation previously identified is the lack of Internet access for the VMs. In principle, this prevents a VM to call external (Web) services and, therefore, to combine together such services, e.g. pipelining genomic diagnostics services available on the Internet. We argue, however, that the same services can be deployed within the trusted domain of MyPHRMachines and be available to all patients to be used. Moreover, we clarified that, for trusted VMs, controlled access to specific Internet addresses can be configured by means of a web proxy. Users should be properly informed of the kind of VM session they are running: a session without Internet access can be trusted blindly while a session with controlled Internet access is only as trustworthy as the Internet sites for which the proxy allows access.

Another consequence of the lack of internet access in *end-user* VM sessions is that the software inside such VM sessions cannot automatically update itself. We argue that this is an acceptable limitation, too. First of all, most automated internet updates are security-related and, therefore, irrelevant for VMs without internet access. Secondly, MyPHRMachines is designed to allow frequent updates at the level of VM images. End-users are expected to have short-living, stateless, VM sessions and therefore if VM updates are provided frequently, then end-users benefit from the functional software updates soon enough.

A straightforward extension deriving directly from the analysis of the radiology use case is the integration with existing EMR systems. This is required to free the patient and caregivers from the burden of transferring to the PHR system all health information and, consequently, is likely to foster adoption. In our opinion as developers of MyPHRMachines, from the technical implementation standpoint, this extension does not represent a substantial obstacle.

About the research method, MyPHRMachines is currently fully implemented using real PHR data and real medical application software. The system, however, has not yet been experimented in clinical settings by real patients. Thus, the above discussion remains at a qualitative level, based on the analysis of the literature and qualitative interviews with key health care stakeholders. Experimentation with actual patients will allow us to evaluate the *people* institutional factor related to MyPHRMachines adoption. This is important since review results have already pointed out that the positive attitude of patients towards PHRs does not translate automatically into their effective adoption [1], [24].

V. Related Work

We can first classify current PHR solutions into free-standing (3rd party), provider-tethered, and integrated PHR systems [25]. Free-standing PHR systems are stand-alone software applications that help patients maintaining their personal health information. Provider-tethered solutions are implemented and made available by a single care institution. In terms of number of users, the most successful PHR solutions belong to the latter category, with examples such as the EPIC MyChart system [26], tethered from hospitals using the EPIC EHR, and MyHealtheVet [24], promoted by the US Department of Veterans Affairs. Besides increasing efficiency, by reducing the need for patient data collection or duplicate clinical exams, provider-tethered PHRs promote a more *stickier* relationship between the provider and the patient. At the same time, however, this type of PHRs do not address the *space* dimension in the continuity of care envisioned for PHRs. An interoperability problem remains, in fact, when the patient seeks care from a caregiver outside of the network of the provider of the PHR. Kaelber et al. have demonstrated theoretically that the large-scale deployment of such PHR systems would have significant economic drawbacks [27].

MyPHRMachines can be classified as an integrated PHR solution [25]. Integrated PHRs are free-standing solutions that collect information from a variety of information sources, such as EMRs, insurance claims, pharmacy data, or data entered directly by patients. Integrated solutions, such as Indivo X [28] or Microsoft HealthVault [29] are less successful in terms of adoption when compared to provider-tethered solutions [26], [29]. Patients, in fact, are required to proactively experiment with the technology without being pushed in doing so by a given provider. Moreover, the interoperability of the PHR with other proprietary systems and, more generally, the provider willingness to trust and use the PHR, are not guaranteed.

The MyPHRMachines solution overcomes that second limitation of integrated PHRs as follows. First, it makes the PHR information trustworthy by delivering original PHR data and related application software directly to care institutions

instead of providing patient-entered information [30]. Second, the barrier to accessing a MyPHRMachines session is minimal, since only one hyperlink needs to be clicked for accessing the trusted health data and its corresponding software.

As far as architecture is concerned, PHR systems rely on a client-server, Web-based architecture [31]. Although Web-based access provides easy access by patients and caregivers, traditional PHR systems remain passive repositories of health-related data, which still require external application software for data visualization or analysis. Software-as-a-Service (SaaS) can be used to integrate application software within Web-based PHRs. Application software will then have to be reprogrammed against the libraries and interfaces provided by the PHR platform, e.g. the Java and .NET libraries of HealthVault. MyPHRMachines does not pose that barrier.

On the one hand, MyPHRMachines preserves the benefit of a Web-based client, i.e., patient and caregivers only need a browser to access data, but, on the other hand, MyPHRMachines extends the scope of traditional PHR systems by allowing to run the original application software to visualize and analyze data through virtualization. Caregivers and software providers will not have to reprogram their application software against a SaaS specification, e.g. Web services over SOAP, but can simply deploy their existing software in a VM image.

As far as PHR data security and privacy are concerned, Web-based PHR systems usually allow patients to collect and store digitized health information, but they usually implement only very simple selective access delegation policies [32]. About commercial systems, PeopleChart [2], for instance, allows separating private and public health information and defining specific roles (e.g. provider or caregiver) to access the information classified as public. MyPHRMachines allows a finer grained sharing approach, where patients can delegate access to subsets of their PHR data to individual caregivers. Such functionality may of course be extended with a role-based access control similar to PeopleChart's, e.g. to share PHR data to all GPs known by a patient, but this extension will still build on the fine grained sharing already implemented in the current version of our prototype.

Unlike MyPHRMachines, the existing PHR platforms provide no technical measures for preventing data abuse by the plug-ins that are contributed by third party software vendors. Instead, they confront patients with take-it-or-leave-it terms of use agreements for each individual third party plug-in. Typically, in such agreements the third party vendors *promise* not to abuse the data. Consequently, upon ad-hoc end-user permission, their software service gets download access to the patient data and it is up to external audits to verify that the terms of use are adhered to. While this architecture may be adequate for sharing information to providers whose reputation is at stake (e.g. an established hospital), it seems much less adequate for a genomic analysis service provided by a niche player from the rapidly evolving bio-informatics industry.

The cloud is by nature opaque [7] and therefore may pose additional data security threats. The encryption of health-related data is a particularly relevant topic in the design of electronic health records [33]. Given that the number and type of care institutions with which health data will be shared is not likely to be known a priori, the literature suggests to use attribute base encryption (ABE) as the main encryption primitive for sharing EHR data [34]. In ABE access policies are expressed based on sets of attributes of users, rather than on the unique identity of users. This allows patients to selectively share their PHR data in a secure way to a set of users without the need to know their complete identity. A characterization of ABE encryption in the context of PHRs has been proposed in [34], whereas the implementation of ABE encryption in the PHR Indivo X is proposed by [35]. We consider ABE encryption a solution that should complement the current implementation of MyPHRMachines. In this paper, in fact, we are brief about such generic security techniques to enable a deeper discussion of the unique privacy protection mechanisms that are offered by MyPHRMachines.

[2]http://www.peoplechart.com

VI. Conclusions

In this paper we presented MyPHRMachines, a novel PHR system. Leveraging virtualization techniques, MyPHRMachines allows patients to build lifelong personal health records. The records can be shared by the patient with any stakeholder interested in those. MyPHRMachines allows also the controlled sharing of application software that is required to view and/or analyze health records. Patients seeking care by caregivers in different geographical areas will be able to reproduce their original health records, no matter the limitations imposed by the heterogeneity of local health care information systems. Moreover, as technology evolves, patients will always be able to use original software to view and analyzed data, even when that software becomes obsolete and possibly no longer supported by the stakeholder that produced the data.

Besides a clinical experimentation, to fairly assess patients' propensity in using such an innovative PHR system as MyPHRMachines, we are currently working on extending our prototype in several ways. One of the major extensions regards creating an open App market for application software, through which medical software providers could compete to provide the best suited functionality required by patients. We are currently studying the issue of how various security techniques can be employed to protect data in MyPHRMachines at various levels, such as encryption techniques at the level of VM instance logs, private key transfers between RDP clients and remote VMs, and encryption at the level of mounted network folders. Furthermore, we are surveying practitioners to understand more broadly and deeply the specific use cases for which MyPHRMachines forms a unique enabler. Finally, we will deploy data translation services to MyPHRMachines. Such services will enable a smooth transition from the already provided functional interoperability to deeper system interoperability. The private network folders will be used as the *blackboard* for exchanging data between different VMs.

References

[1] D. C. Kaelber, A. K. Jha, D. Johnston, B. Middleton, and D. W. Bates, "Viewpoint paper: A research agenda for personal health records (PHRs)," *JAMIA*, vol. 15, no. 6, pp. 729–736, 2008.

[2] AHIMA e-HIM Personal Health Record Work Group, "Defining the personal health record," *Journal of AHIMA*, vol. 76, no. 6, pp. 24–25, Jun. 2005.
[3] P. C. Tang, J. S. Ash, D. W. Bates, J. M. Overhage, and D. Z. Sands, "White paper: Personal health records: Definitions, benefits, and strategies for overcoming barriers to adoption." *JAMIA*, vol. 13, no. 2, pp. 121–126, 2006.
[4] International Standards Organization, "ISO/TR 20514:2005 – health informatics – electronic health record – definition, scope and context," Jan. 2005.
[5] A. Rosenthal, P. Mork, J. Li, M.H. adn Stanford, D. Koester, and P. Reynolds, "Cloud computing: a new business paradigm for biomedical information sharing," *Journal of Biomedical Informatics*, vol. 43, pp. 342–353, 2010.
[6] Accelarad, "Seemyradiology – medical image sharing," http://www.seemyradiology.com/, Jul. 2012.
[7] S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, and A. Ghalsasi, "Cloud computing - the business perspective," *Decision Support Systems*, vol. 51, pp. 176–189, 2011.
[8] M. Beyer, K. A. Kuhn, C. Meiler, S. Jablonski, and R. Lenz, "Towards a flexible, process-oriented IT architecture for an integrated healthcare network," in *Proceedings of the 2004 ACM Symposium on Applied computing*, 2004, pp. 264–271.
[9] T. J. Bittman, G. J. Weiss, M. A. Margevicius, and P. Dawson, "Magic Quadrant for x86 server virtualization infrastructure," Gartner RAS Core Research Note G00205369, Jun. 2011.
[10] D. T. Mon, J. Ritter, C. Spears, and P. Van Dyke, "PHR System functional model," HL7 PHR Standard, May 2008.
[11] S. Negrini, S. Atanasio, F. Zaina, and M. Romano, "Rehabilitation of adolescent idiopathic scoliosis: results of exercises and bracing from a series of clinical studies. Europa Medicophysica-SIMFER 2007 Award Winner." *European journal of physical and rehabilitation medicine*, vol. 44, no. 2, pp. 169–176, Jun. 2008.
[12] G. S. Ginsburg and J. J. McCarthy, "Personalized medicine: revolutionizing drug discovery and patient care," *Trends in Biotechnology*, vol. 19, no. 12, pp. 491 – 496, 2001.
[13] M. L. Metzker, "Sequencing technologies - the next generation," *Nature reviews. Genetics*, vol. 11, no. 1, pp. 31–46, Jan. 2010.
[14] K. Wetterstrand, "DNA sequencing costs - data from the NHGRI large-scale genome sequencing program," Jan. 2012. [Online]. Available: http://www.genome.gov/sequencingcosts/
[15] P. Van Gorp and P. Grefen, "Supporting the internet-based evaluation of research software with cloud infrastructure," *Software and Systems Modeling*, vol. 11, no. 1, pp. 11–28, 2012.
[16] Microsoft Terminal Services Team, "Top 10 RDP protocol misconceptions," http://blogs.msdn.com/b/rds/archive/2009/03/03/top-10-rdp-protocol-misconceptions-part-1.aspx, Mar. 2009.
[17] E. Rescorla, "HTTP over TLS," IETF RFC 2818, United States, 2000.
[18] T. Dierks and E. Rescorla, "The transport layer security (TLS) protocol," IETF RFC 4346, 2006.
[19] Nucleics, "Reviews of dna sequencing service companies & facilities," http://www.nucleics.com/DNA_sequencing_support/sequencing-service-reviews.html, 2012.
[20] MyBio community, "Sequence format conversion," http://mybio.wikia.com/wiki/Sequence_format_conversion, 2012.
[21] R. Agarwal, G. Gao, C. DesRoches, and A. K. Jha, "The role of information systems in healthcare: Current research and road ahead," *Information Systems Research*, vol. 22, pp. 419–428, 2011.
[22] W. Currie and M. Guah, "Conflicting institutional logics: a national programme for IT in the organizational field of healthcare," *Journal of Information Technology*, vol. 22, pp. 235–247, 2007.
[23] A. Flood and M. Fennell, "The role of organizational theory and research in conceptualizing and examining our health care system," *Journal of Health and Social Behavior*, vol. 35, pp. 154–169, 1995.
[24] K. Nazi, "Veteran's voices: use of the american customer satisfaction index survey to identify MyHealtheVet personal health record users' characteristics, needs, and preferences," *J Am Med Inform Assoc*, vol. 17, pp. 203–211, 2010.
[25] D. Detmer, M. Bloomrosen, B. Raymond, and P. Tang, "Integrated Personal Health Records: Transformative tools for consumer-centric care," *BMC Medical Informatics and Decision Making*, vol. 8, 2008.
[26] J. Halamka, K. Mandl, and P. C. Tang, "Early experiences with personal health records," *Journal of the American Medical Informatics Association*, vol. 15, no. 1, pp. 1–7, 2008.
[27] D. C. Kaelber, S. Shah, A. Vincent, E. Pan, J. M. Hook, D. Johnston, D. W. Bates, and B. Middleton, *The Value of Personal Health Records*. Healthcare Information & Management Systems Society, 2008.
[28] B. Adida, A. Sanyal, S. Zabak, I. S. Kohane, and K. D. Mandl, "Indivo X: Developing a fully substitutable personally controlled health record platform," in *AMIA 2010 Symposium*, Nov. 2010.
[29] A. Sunyaev, D. Chornyi, C. Mauro, and H. Kremar, "Evaluation framework for personal health records: Microsoft healthvault vs. google health," in *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, jan. 2010, pp. 1 –10.
[30] A. S. McAlearney, D. J. Chisolm, S. Schweikhart, M. A. Medow, and K. Kelleher, "The story behind the story: Physician skepticism about relying on clinical information technologies to reduce medical errors," *International Journal of Medical Informatics*, vol. 76, no. 11-12, pp. 836 – 842, 2007.
[31] N. Archer, U. Fevrier-Thomas, C. Lokker, K. A. McKibbon, and S. E. Straus, "Personal health records: a scoping review," *JAMIA*, vol. 18, pp. 515–522, Jul. 2011.
[32] I. Carrion, J. Fernandez Aleman, and A. Toval, "Personal health records: New means to safely handle our health data?" *Computer*, vol. PP, no. 99, pp. 1–1, 2012.
[33] R. Zhang and L. Liu, "Security models and requirements for healthcare application clouds," in *IEEE 3rd Int. Conf. on Cloud Computing*, 2010, pp. 268–275.
[34] M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, pp. 131–143, 2013.
[35] C. Wang, X. Liu, and W. Li, "Implementing a personal health record cloud platform ciphertext-policy attribute-based encryption," in *4th IEEE Int. Conf. on Intelligent Networking and Collaborative Systems*, 2012.

Pieter Van Gorp is investigating and extending the applicability of transformation technology to software modeling and health information system challenges since 2002. Since 2008, he is an assistant professor in the School of Industrial Engineering at Eindhoven University of Technology. Previously he held a postdoc position at the University of Antwerp, where he also obtained his Ph.D. degree in Software Engineering. He has published various conference and journal papers. Van Gorp has developed various software prototypes, including SHARE and MyPHRMachines. He also teaches various courses, covering topics such as model-driven engineering, business process simulation and health informatics. Van Gorp has participated in the organization of national and international research events. He is the PC chair of the Foundations track of ECMFA 2013. Recently, he has visited the Clinical Informatics Research and Development group from Partners HealthCare to study the potential role of MyPHRMachines in the US health care system.

Marco Comuzzi is an assistant professor at the Eindhoven University of Technology. He received his Ph.D. in Information Technology from Politecnico di Milano in 2007. He has been visiting researcher at the McCombs School of Business, University of Texas at Austin, and a post-doctoral research fellow at City University London. More recently, he has been a visiting lecturer at UNIST, Ulsan, South Korea. His research interests concern the design and management of process-oriented service-based systems. In particular, he has focused on architectures for the automated negotiation of services quality and on the monitoring of service contracts and SLAs. He has published several papers in international journals and conference proceedings and he has been involved in several EU and national research projects.