Medical Physics
The International Journal of Medical Physics Research and Practice

Cloud computing in medical imaging

George C. Kagadis, Christos Kloukinas, Kevin Moore, Jim Philbin, Panagiotis Papadimitroulas, Christos Alexakos, Paul G. Nagy, Dimitris Visvikis, and William R. Hendee

Citation: Medical Physics **40**, 070901 (2013); doi: 10.1118/1.4811272
View online: http://dx.doi.org/10.1118/1.4811272
View Table of Contents: http://scitation.aip.org/content/aapm/journal/medphys/40/7?ver=pdfcov
Published by the American Association of Physicists in Medicine

Cloud computing in medical imaging

George C. Kagadis[a)]
Department of Medical Physics, School of Medicine, University of Patras, Rion GR 26504, Greece

Christos Kloukinas
Department of Computer Science, City University London, Northampton Square, London EC1V 0HB, United Kingdom

Kevin Moore
Department of Radiation Oncology, University of California, San Diego, California 92104

Jim Philbin
Department of Radiology, Johns Hopkins University, Baltimore, Maryland 21287

Panagiotis Papadimitroulas
Department of Medical Physics, School of Medicine, University of Patras, Rion GR 26504, Greece

Christos Alexakos
Pattern Recognition Laboratory, Department of Computer Engineering and Informatics, University of Patras, Rion GR 26504, Greece

Paul G. Nagy
Department of Radiology, Johns Hopkins University, Baltimore, Maryland 21287

Dimitris Visvikis
INSERM, UMR1101, LaTIM, CHRU Morvan, Universite de Bretagne Occidentale, Brest FR 29609, France

William R. Hendee
Department of Radiology, Mayo Clinic, Rochester, Minnesota 55905

(Received 25 March 2013; revised 13 May 2013; accepted for publication 3 June 2013; published 21 June 2013)

Over the past century technology has played a decisive role in defining, driving, and reinventing procedures, devices, and pharmaceuticals in healthcare. Cloud computing has been introduced only recently but is already one of the major topics of discussion in research and clinical settings. The provision of extensive, easily accessible, and reconfigurable resources such as virtual systems, platforms, and applications with low service cost has caught the attention of many researchers and clinicians. Healthcare researchers are moving their efforts to the cloud, because they need adequate resources to process, store, exchange, and use large quantities of medical data. This Vision 20/20 paper addresses major questions related to the applicability of advanced cloud computing in medical imaging. The paper also considers security and ethical issues that accompany cloud computing. © *2013 American Association of Physicists in Medicine*. [http://dx.doi.org/10.1118/1.4811272]

Key words: cloud computing, cloud services, medical imaging, PACS, security, ethics

I. INTRODUCTION

The expression "cloud computing" refers to the access of computing resources through the Internet for purposes of data storage, aggregation, synthesis, and retrieval, together with the capacity to act on the data with computational algorithms and software packages.[1] Cloud computing is available on-demand and provides flexible and scalable computing resources from remote locations. It is particularly useful in research applications involving multiple investigators at different institutions, and in large-scale data processing applications such as those in clinical medicine. It may well become a major resource in efforts to identify surrogate measures to clinical trials for evaluation of new drugs and devices in biomedicine. A number of cloud-computing resources are available for use in research applications.

Cloud computing is emerging as a solution to the challenge of delivering complex services and data interchange over the Internet. It has quickly attracted worldwide usage and is now part of our daily life, with applications such as Gmail, Google Docs, Dropbox, etc. The increasing success of cloud computing is due to the ever-decreasing cost and increasing ubiquitous presence of fast networks, which make it economically viable to access large amounts of data remotely and in real-time. It has built upon ideas and technologies initially developed for other initiatives such as grid computing, and reflect an extension of the general architecture and technologies for managing the distribution of resources on a hardware infrastructure. The main difference between grid and cloud computing stems from their different orientations. Grids aim to provide computational power similar to large distributed and parallel hyperperformance computing systems, while

clouds are challenging Internet-scale computing limitations such as application accessibility and storage space. Cloud computing focuses on the abstraction of shared resources through the use of virtualization, whereas grids are most useful where the workload distribution management and application parallelization is of primary concern. Furthermore, cloud-based applications are developed fundamentally on Internet technologies (PHP, AJAX, ASP.NET, HTML, CSS, REST, etc.), while grid applications are based mainly on parallelization and workflow management programming (MPICH-G2, Linda, CoG Karajan, etc.). From a user perspective, cloud services are easier to use because they utilize well-known, standardized components.[2] But what does cloud computing mean exactly? Is it useful in general and in healthcare, and if so, why, and in which situations? What are the benefits and risks of computational platforms distributed in the Internet cloud? In this Vision 20/20 paper we attempt to answer these basic questions, and review the most important applications and ethical challenges arising from this new approach to manipulating healthcare data through distributed computers around the world.

Several authors have tried to provide a succinct definition of what cloud computing is.[3–6] However, these definitions are still not universally applicable because each is oriented toward specific applications. This is understandable, as cloud computing is quite young and researchers and engineers from various fields have different points of view.

While recognizing that it is an evolving paradigm, the National Institute of Standards and Technology (NIST) gives a brief definition of "Cloud Computing" as "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".[1] The term derives from the cloud metaphor, which is widely used in diagrams for representing an underlying network infrastructure. The cloud is used to depict services and data that are removed from user computers, servers, and other local systems, and that provide a virtual bucket of information, functions, applications, and infrastructure that can be accessed and used by any user and any system, anywhere in the world.[4]

I.A. Cloud services

The main concept of cloud computing is the delivery of various services to end-users through the Internet. These services cover all the properties of Information and Communication Technology (ICT), from a virtual hardware infrastructure to software applications. All of the services can be remotely accessed and configured by the end user through a web browser. The exposure of user data to a world-wide accessed network raises the need for secure control and communication technologies in order to offer protected ways to collaborate. In the later part of 2007 the term cloud computing started to be used across the IT community.[5] Indeed, the period between 2006 and 2008 saw the introduction of a number of major commercial cloud providers. In 2006 Amazon released its Amazon Web Service™ (AWS) cloud service, delivering server virtualization technology. In 2008 Google published Google App Engine™ (GAE), a platform with tools for developing cloud applications. During the same year, Microsoft released Microsoft Azure™, a cloud infrastructure based on Windows Azure Hypervisor (WAH), providing .NET based tools for application development.[7] Using these commercial cloud infrastructures one can quickly prototype cloud applications. When more control over the underlying cloud infrastructure is needed, there are solutions that permit the deployment of private clouds. Eucalyptus was the first open-source platform for deploying private clouds.[8] While all clouds are based upon the Internet, a private cloud encrypts data in flight and is available only to authorized users. OpenNebula became the first open-source software for deploying private and hybrid (public-private) clouds.[9]

I.B. Cloud types: IaaS, PaaS, SaaS

Cloud platforms have been categorized into three main groups according to the type of provided services: infrastructure, platform, and service (Fig. 1).

I.B.1. Infrastructure as a Service (IaaS)

IaaS uses *virtualization* technology to allow several virtual systems (referred to as virtual machines) to operate on top of a single physical hardware infrastructure in an isolated manner. The key software module in *virtualization* is the hypervisor that manages and organizes the virtual resources on the physical hardware (memory, processors, storage). In this category, cloud providers can deliver on-demand virtual machines with configurable resources.[10]

Essentially, IaaS allows clients to dynamically rent (virtual) machines on which they can install their own operating system (OS) and other applications without worrying about the machines getting old and needing replacement, or staying idle and wasting resources. Furthermore, virtualization facilitates a faster recovery from hardware and system failures, as the VM's snapshots can be copied and booted up in other cloud nodes. Also, the clients can avoid the initial investment that would be needed to purchase their own hardware platform.[1,9]

I.B.2. Platform as a Service (PaaS)

PaaS includes all the features provided by IaaS, but in this case the user is able to use the provider's system platform. PaaS allows clients to develop their own system using the platform tools, without having to install and maintain these tools themselves.

In this category, users obtain access to a specific OS (e.g., a version of Windows or Linux) and associated tools (e.g., SQL Server, MySQL, Apache web server, etc.).[3,11] Users do not have to worry about keeping their OS and tools updated,

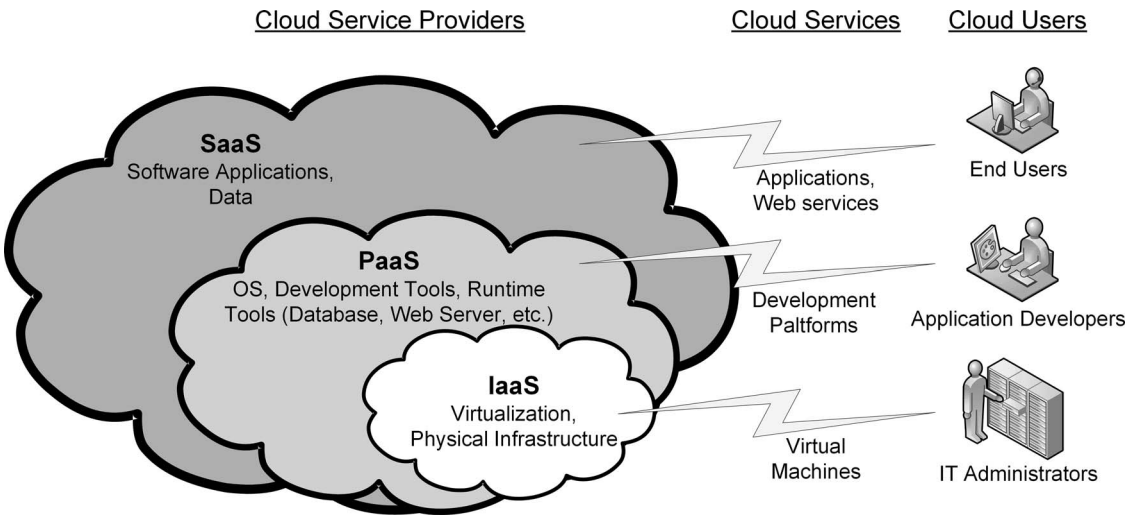

FIG. 1. Different types of cloud services.

virus-free, etc. At the same time, users may not be able to control the updates of the tools. If a user application depends on a specific version of the OS, it may suddenly stop working because an OS update has been installed by the cloud provider.

I.B.3. Software as a Service (SaaS)

This type of computer cloud eliminates the need to install and run software applications on the client's local computers. With SaaS, cloud providers install, manage, and operate the software application, and the user has neither knowledge nor control of the underlying infrastructure. With this type of cloud service, the end-user has the least flexibility but the cost is dramatically lower.[3,6,11] Examples of this category of cloud computing include GMail, Google Docs, Dropbox, etc.

I.C. Cloud computing in offices, companies, and organizations

Businesses, government agencies, organizations, and individual consumers are rapidly adopting cloud technologies. Businesses are becoming increasingly receptive to the potential for broadening their development, services, and marketing through information technologies.[12] Still, some companies have concerns about transferring to cloud providers their computing capacity, and data storage and retrieval needs. There are four major obstacles that impede companies and organizations from adopting a cloud solution.

The main concern deals with data security when data are stored in the cloud. Data protection and privacy is one of the most important issues that providers must confront. The continuous improvement of security technology allows cloud providers to offer clients services with lower risks and improved privacy protection.[11] This issue is discussed more fully in Sec. V.

The second concern pertains to reliability and performance. Companies and organizations worry about not being able to access their data because of the unavailability of cloud services when they need them. Usually there is little if any forewarning of when services may be unavailable.[13,14] Some industry leaders argue that it is still an advantage for businesses to keep services and valuable data in their own infrastructure, to ensure they can access them at any time.

As more companies adopt cloud services, the issues of interchanging data and standardization of APIs become more relevant. Common data file formats and interoperability standards are still immature in this emerging industry, which may lead to vendor lock-in.[15,16] Managing vendor lock-in is currently an important issue. Most cloud providers use proprietary APIs for both application development and data storage, which impedes the development of vendor-agnostic applications. Data or application migration from one vendor to another is needed for scalability and to take advantage of lower-cost offers when a cloud vendor stops offering a required service or goes out of business. The solution to this problem is the standardization of APIs, which would permit the development of cloud applications compatible with different public cloud vendors as well as with in-house private clouds.[4] The exchange of data between similar applications on different cloud systems and the ability to move data (and applications developed using cloud provider frameworks) from one cloud service vendor to another are still challenges with incomplete solutions. A precedent for solving this challenge is the common file format entitled DICOM that permits medical images to be transferred among imaging systems and between different institutions.

The final challenge for utilizing cloud services is cost. This is a quickly evolving field, with rapid reductions in cost. Different vendors have different pricing models, and there is encouraging information that the cloud is cost-competitive when compared with local server solutions for applications that require very large storage (hundreds of TB) and performance requirements.[17,18]

I.D. What is the need for cloud computing in healthcare?

Healthcare systems are trying to evolve and benefit from cloud services. This is because information technologies and information-rich services such as medical imaging can be greatly enhanced by use of cloud technologies. Collaboration among medical institutions and hospitals is required for sharing medical data and images. Patient data can be easily stored in virtual archives that are accessible by different healthcare providers, thus facilitating data sharing and significantly reducing local storage requirements. Privacy issues arise from use of cloud systems for confidential personal data. Nevertheless, there are significant advantages in the interpretation of difficult clinical cases when employing cloud computing services. Experts from different medical fields can consult on the diagnosis from around the world.

Continuing education and teaching efforts can also be facilitated by the cloud. Teaching files can be accessed by several institutions, and training courses can be co-organized to provide shared access to learning tools such as software, presentations, and medical images of clinical interest.

While the number of medical imaging studies is increasing approximately 3%–5% annually, this rate is not a significant driver of growth. However, the size of medical imaging studies, especially CT and MRI, is growing considerably faster, increasing storage requirements from 10% to 25% annually. Cloud storage prices have been dropping faster than enterprise storage prices, and this trend will likely lead to faster cloud adoption for medical image storage.[17–20]

An important driver of cloud storage is the observation that as CTs and MRI studies increase in size, longer times are required to transfer them to imaging workstations. Rendering imaging studies from the cloud to zero footprint viewing applications (considered in more detail in Sec. III.B.1) provides imaging studies anywhere they are needed. These factors are significant drivers of the move to cloud PACS, including storage.[21,22]

II. CLOUD COMPUTING AND MEDICAL IMAGING

Cloud computing is still in its infancy in the medical imaging domain, and there is currently low market penetration within the field. However, this situation may change rapidly in the near future. Among the potential driving forces for the increased use of cloud computing in medical imaging are raw data management and image processing and sharing demands, all of which require high-capacity data storage and computing.

Medical image reconstruction is a rapidly evolving domain for different imaging modalities, which over the next few years will require accurate quantitative analysis of reconstructed images. Quantification will be achieved, in part, through improved understanding of various biological processes that influence the quantitative analyses of medical images.[23] It will also be improved by considering these biological influences within the same reconstruction framework, for each of two or more imaging modalities.[24,25] The development of "quantitative imaging" has been hampered in the past by the heavy computational workload of reconstruction processes and associated correction algorithms, leading to long execution times that are incompatible with their use in clinical practice. One such example is the use of Monte Carlo modeling approaches in tomographic reconstruction to accurately account for effects associated with the physics of the detection process, considering both detector and patient interactions. One potential solution to this computational efficiency challenge is the sharing of computing facilities through the use of cloud infrastructures. In this context, advantages can be expected in the case of reconstruction algorithms and incorporated correction methodologies that can be executed in a multi-threading fashion with high potential for parallelization. One potential issue with this approach concerns the raw data transfer, which must occur over a high-speed network to exploit the speed advantage of a cloud computing infrastructure. This is a crucial factor in quantitative image reconstruction in the clinical setting, although it may be less critical in research applications.

Medical image processing is another domain that can benefit from access to cloud computing. Research in different areas of medical image processing over the past decade has led to continuous algorithmic improvements. Any new image processing algorithm should be evaluated in comparison with current algorithms to determine if it improves image quality and clinical yield. One of the challenges facing medical image processing today is the development of benchmarks[26] that allow image processing algorithms to be compared under common measures and standards. The cloud can contribute to such benchmarks by facilitating their creation as well as their widespread availability and use. Such databases combine different datasets necessary for the assessment of image processing algorithms (e.g., in segmentation, denoising, registration, fusion). An important component of these databases is realistic simulated medical image datasets,[27–30] which usually are a first step in the evaluation of any image processing algorithm. These simulations could model the different components of an acquisition process to produce highly realistic datasets, with the advantage of a known ground truth. However, such simulation datasets are often associated with long execution times and, therefore, their production is largely confined to a few specialized centers. The same concept could be developed for the previously mentioned image reconstruction challenges, with the creation of databases hosting raw datasets of different medical image modalities that would facilitate benchmarking of future reconstruction algorithms.

III. CLOUD COMPUTING AND PACS

A *Cloud-Based PACS*, or *Cloud PACS*, is a PACS system that is located in the cloud and is accessible to both users and administrators though Internet-based user interfaces. Cloud PACS offer the promise of both location and device independence. That is, a Cloud PACS can be accessed by the user from any location, with the assumption that it has sufficient network connectivity and an appropriate connected device that has sufficient display characteristics.

Until recently, PACS clients have needed powerful workstations to render imaging studies. This requirement has tied the radiologist or cardiologist to a location with a powerful workstation and a high performance network connection.

Three recent technological developments have made Cloud PACS possible: (1) remote visualization, which moves much of the graphics processing to the cloud and decreases the amount of data that has to be moved from the cloud to the end-point device, (2) increased processing power and resolution of end-point devices, and (3) maturation of HTML5, CSS3, and JavaScript to the point where zero-footprint, diagnostic-quality, and browser-based applications are feasible. These developments have eliminated the need for high-end workstations and have also reduced networking throughput requirements. Because image rendering occurs at the server side, fewer data must be transmitted to the client.

III.A. Cloud PACS

As discussed earlier, the cloud offers many advantages in terms of shared resource utilization, economies of scale, and lower maintenance and management overhead. This section focuses on the advantages of Cloud PACS including location and device independence.

Location independence means that physicians can read studies created at other hospitals or outpatient centers. They can read from home, office, or from different locations inside or outside hospitals or clinics. Further, physicians can refer patients to other physicians or request consults from physicians at different locations, with the consultations taking place through the cloud. Device independence means that an application will run on desktops, laptops, tablets, smart phones, or other end-point devices. This device independence is achieved by using various Internet technologies, especially browser based applications.

III.B. PACS three components

A PACS is composed of three main components: an image visualization application, a workflow engine, and an image archiving system. Moving each of these components to the cloud has specific advantages and drawbacks.

III.B.1. Cloud based image visualization

Image viewing is probably the aspect of Cloud PACS that provides the most immediate benefit to the end user. As the amount of data in imaging studies grows, the time required to move studies to a diagnostic workstation increases. Cloud PACS uses remote visualization, also known as remote rendering, where servers in the cloud data center are responsible for rendering the images and then sending them to the remote client or end-point device. There are several technologies that can be used on the end-point device to communicate over the Internet to the Cloud PACS: (1) thin client applications; (2) rich Internet applications (RIAs); (3) browser-based zero-footprint web applications; and (4) desktop virtualization.

Thin executable client applications have been available for some time.[31] They are typically written as a desktop application, but the image rendering is done remotely on a cloud server. Thin clients have several disadvantages: (1) they depend on the client OS and thus have to be rewritten for different end-point devices, (2) they have to be installed on the client desktop, and (3) they must be managed and maintained, which can be expensive and time consuming.

RIAs have become available recently and are intended primarily for clinical viewers. RIAs are applications provided over the Internet, usually through a browser, that offer functionality comparable to what one would expect with a standalone native application. A RIA relies on a client environment such as Adobe Flash,[32] JavaFX,[33] or Microsoft Silverlight[34] that must be installed in the client browser(s). The RIA can then be loaded or upgraded automatically by the browser. New versions of the RIA environment must be managed and maintained in the traditional manner, which can add significant cost, although probably less than that associated with thin clients.

With the advent of HTML5,[35] *zero-footprint clients*, also known as *zero clients*, have become ubiquitous for clinical, i.e., nondiagnostic, viewers. Zero clients can be loaded into a browser transparently from a remote website. This action simplifies the maintenance and management of the application, as users only need to have an up-to-date browser on their device. It also allows the viewing application to be deployed on any device that has a supported browser. Typically, the four main browsers, Chrome, Firefox, Internet Explorer, and Safari, are supported. HTML5 has substantially improved the capacity for displaying images, and several HTML5 diagnostic-quality zero clients have recently been introduced. This trend is expected to continue and accelerate.

Finally, desktop virtualization[36] has shown significant promise in converting traditional thick clients into cloud-friendly applications. Desktop virtualization provides remote access to a full featured OS environment running in a remote (possibly virtual) machine. It also improves zero client performance on broadband networks. Desktop virtualization has not yet received FDA approval, but is expected to gain approval in the near future.

III.B.2. Cloud based workflow

The second major component of PACS is the workflow engine. Moving the engine to the cloud provides several opportunities for improving imaging workflow within and across healthcare organizations. The most obvious is the ability to distribute work more efficiently. Specialty and subspecialty studies can be put on the worklist of the most appropriate physician. Urgent (stat) cases can be sent to the most available physician. The work can be balanced across the physician population by whatever criteria deemed most appropriate, such as balancing the workload or rewarding more efficient physicians. Since the work can be distributed independently of where physicians are located, it can be allocated according to almost any criterion desired.

Cloud PACS also makes intrahealth and interhealth system ordering, referrals, consults, second opinions, and patient transfers much easier as well.

III.B.3. Cloud based image archive

The third major component of PACS is the image archive, and again, the cloud provides several opportunities for improvement. Most of the opportunities are afforded by cloud based vendor neutral archives or Cloud PACS. The most fruitful opportunity is the aggregation of imaging records within and across healthcare organizations that encompass several geographic locations. This aggregation provides the following advantages: (1) prior studies are available for comparison even if they were done at another location; (2) unnecessary imaging, along with unnecessary radiation dose, can be reduced since the imaging study from another location is available; (3) the use of CD/DVDs can be reduced or eliminated, which in turn reduces cost and enables more timely care; (4) image sharing for referrals, second opinions, and subspecialty consults are enabled; and (5) cross-enterprise archives provide the foundation for patient-controlled image sharing.

IV. IMPACT AND ADVANTAGES OF CLOUD COMPUTING IN RESEARCH

Cloud-based research applications make parallel computation on large datasets easier and more cost-effective for researchers.[37,38] Examples of distributed computational infrastructures span the range from the multi-institutional Large Hadron Collider's Worldwide LHC Computing Grid[39] to the use of commercial cloud-based systems to speed individual project simulations.[40] It is likely that the discipline of bioinformatics[41,42] will be particularly amenable to distributed computational analyses, because of the large size of the datasets and the parallelism inherent to bioinformatics data-mining operations.

Clinical trial investigations are, in principle, very well suited for cloud-based infrastructures because each enrollee in a clinical trial is treated as an independent datum. Analyses on the trial can be parallelized by treating each patient concurrently. Even if complex calculations are required for each patient in a specified cohort, the identical operations can be easily distributed in parallel to a cluster of processors, provided that the infrastructure is designed to facilitate the distribution. In medical imaging trials, in particular, all data are readily digitized and formatted according to the DICOM standard. A cloud-based platform for a clinical trial might function as follows:

- Trial coordinators (cooperative groups, quality assurance centers, commercial partners) establish data formatting rules, data storage, and an infrastructure for management of trial data.
- Participating institutions upload data to the trial storage cloud, pending standard quality assurance verifications.
- Standardized analysis engines are provided for queries on the trial data, facilitating both primary analyses (comparing outcomes in distinct arms) and approved secondary investigations (correlating variables ancillary to main study aims).

A schematic of such a system is depicted in Fig. 2.

The main benefit of cloud-based clinical trial research is the greatly improved accessibility of data and efficiency of analysis. Beyond data storage and accessibility, the critical component of the cloud is the analysis platform, which must support a wide spectrum of queries of the data. Straightforward queries may be sufficient to achieve many study endpoints, but analyses that depend on, for example, dose-volume information could require patient-by-patient operations on DICOM data. To provide maximum flexibility for secondary investigations, advanced operations on trial data should be possible. For example, consider a secondary analysis that requires autosegmentation of cone-beam computed tomography (CBCT) scans. Such advanced operations should be imbedded in the analysis so that the investigator is not required to download each CBCT image set, a time-consuming and costly operation. A versatile DICOM toolkit coupled with generalized programming language will make almost any conceivable investigation possible.

If postmarket surveillance (PMS) is required as part of a clinical trial, the same infrastructure used for the trial can be leveraged to make the PMS process more efficient and transparent. PMS typically involves larger patient numbers but less per-patient data input to the system. Incorporation of this phase of the trial can be designed upfront to ensure maximum concordance between trial and PMS data. Premarket analysis tools regarding toxicity can likely be repurposed for PMS, speeding the analyses and facilitating independent oversight. With the potential adoption of cloud archives that aggregate patient health records across healthcare enterprises in future clinical installations, PMS could be distributed to regional nodes very effectively.

Machine learning is an underutilized resource in the analysis of clinical trial data or large-volume retrospective clinical data. Machine learning is well-suited to cloud-based infrastructures, with Google's Prediction API being one example. Machine learning is a body of techniques that allow unknown correlations to be discovered without an *a priori* hypothesis. Insofar, as clinical trials are designed to test specific hypotheses, machine learning could potentially assist researchers in establishing the validity of primary endpoints and in discovering unforeseen correlations. For example, consider a large-scale clinical trial comparing an image-guided intensity-modulated radiation therapy treatment with and without a particular chemotherapy regimen. While the primary endpoint is the efficacy of the chemotherapy regimen, there are other inferences that might be drawn from the data. Multi-institutional trials will inevitably use a range of treatment planning systems; perhaps, there is a preferred dose distribution that is unique to one platform. Or perhaps there are systematic differences in treatment plan quality among individual institutions. These types of discovered

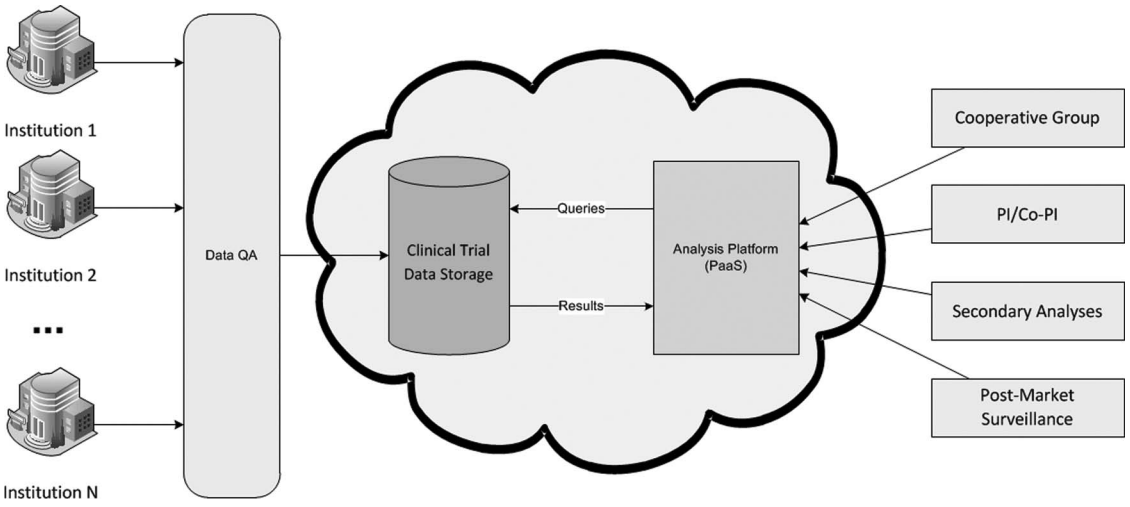

FIG. 2. Schematic diagram of a cloud-based clinical trial system. While the PaaS cloud configuration is not required, it best fits the needs of multiple investigators interacting with a regimented dataset in user-defined ways.

features could not be hypothesized at the beginning of a trial, and yet treatment plan quality variations may, in fact, subvert the aims of the trial. Applying machine learning techniques to identify such unheralded features could both improve the power of the trial's conclusions and provide comparisons between interinstitutional practices that would be virtually impossible to detect at each institution alone. Analyzing features of a large, uniform dataset is an operation best accomplished with a cloud-based platform, as all data are easily available to researchers from different institutions and the cloud can elastically provide enough resources for the purposes of the analysis that may be impractical for any single institution to provide on its own.

V. ETHICAL ISSUES AND SECURITY

Cloud computing raises several ethical issues that are less paramount when research data are managed by a single institution. These issues acquire greater significance in cloud computing[43,44] because control over research data is transferred from the institution to a third party, namely, the service provider for cloud computing. Foremost among the concerns is the need to ensure the privacy and security of patient data, and to make certain that only authorized individuals have access to the data. The service provider that is managing the computational cloud must take thorough measures to prevent security breaches, and suitable encryption should be used during uploading, transfer, and downloading of patient data. If the service provider is accessing computational resources disseminated across the globe, clients must consider whether patient data are adequately protected, especially if the resources are outside the legal jurisdiction of the country of origin. The service provider must provide written assurance that measures are in place to protect data from unauthorized use or from uses not originally intended by the researchers or clinicians. Indeed, the cloud provider may need to guarantee that data will not be transferred outside a particular jurisdiction, as, for example, is the case with the EU.[45] Finally, the service provider must be able to quickly destroy data upon instructions of the clients if a security breach occurs.[46]

Contracting with a cloud computing service provider exposes clients to the risk of dealing indirectly with several entities for data management, computation, and analysis. This risk is referred to as parameterization, and is a consequence of multiple linkages developed by the service provider.[47] In a networked management architecture offered by a cloud service provider, it may be difficult to ascribe consequences of actions to a single person or organization, as forensic investigators cannot seize equipment and analyze it offline. It is important that healthcare researchers reach an agreement that research data will not be used by the service provider or its subcontractors for purposes other than those intended in the research study. Researchers should also ensure that they retain ownership of the research data, through a written clause in the contract for cloud services.

One of the risks of cloud computing is the challenge of migrating research data from one cloud-computing platform to another, or even returning the data to an institutional platform, if such actions are deemed desirable in the future. This challenge reflects the proprietary nature of the platform of a service provider in the case of SaaS, and the unique structural requirements for data entered into it. This challenge should be addressed before services for cloud computing are contracted for, and not later when the need for data transfer arises.[48]

Cloud computing is a relatively recent addition to the world of research, and all of its potential ethical implications cannot be foreseen. For this reason, the precautionary principle should be applied in all negotiations for cloud services.[49] The precautionary principle attempts to prevent harm from unknown consequences, without hampering progress and innovation altogether. It states that one should refrain from actions in the face of scientific uncertainty about serious or irreversible harm. Furthermore, the burden of proof for assuring the safety of an action falls on those who propose it.

Security in ICT can be defined with the classic CIA model, named after the three properties it considers: confidentiality,

integrity, and availability. Confidentiality relates to the prevention of information disclosure to unauthorized entities (systems or individuals). Integrity ensures that the data cannot be modified without detection. Availability refers to the provision of the services or data when they are requested by authorized entities. Mechanisms for providing CIA include a combination of encryption schemes, access control, data backup, data replication, and data storage safety measures.[50] For medical records, security is complicated by various ethical and legal requirements that must be fulfilled to ensure protection of sensitive personal data. Furthermore, national regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health Act (HITECH) in the United States, demand additional layers of security for medical data storage, exchange, and use.[51] The location of the cloud infrastructure and the country where the cloud provider is incorporated can be issues as well. For example, a country could introduce a law by which the cloud provider must make available to the authorities sensitive data about patients of another country.[52]

Medical applications store, exchange, and use sensitive personal data, and protection of the data from unauthorized access is an ongoing requirement. User authentication and authorization, as well as information ownership, are the main security issues that must be taken into account when considering cloud-based medical services.[53] In SaaS and PaaS service models, the software application is responsible for managing user authentication and information access rights, and the cloud provider has direct access to these data at any time. Lightweight Directory Access Protocol (LDAP) based servers are commonly used to store user security credentials and permissions. LDAP servers can be installed in an organization's premises, thus increasing the protection of sensitive employee and patient information. Data access policy is specified in terms of user roles (Role-based Access Control – RBAC) in combination with access permissions to the data objects (Attribute-based Access Control – ABAC) at both application and OS levels. In the IaaS model, where the OS and hardware are delivered as virtual systems, the security of the hardware and hypervisor software is a responsibility of the cloud provider that manages the security controls for hardware access, environmental security, and hypervisor security.[54] Hypervisor security refers to restrictions on access to the content of the virtualized resources (e.g., hard disks and processors).

The protection of individual medical records from access by unauthorized third-party entities is a major ethical and legal issue. Data encryption of objects that are exchanged or stored is the most suitable measure of data protection. For data exchange over the Internet, common network transmission encryption techniques such as Secure Sockets Layer (SSL) and Transport Layer Security (TLS) provide strong data protection. A potential vulnerability of these techniques is the insecure SSL trust configuration and the interception of the cryptographic keys by malicious entities.[54] Network security can be strengthened by establishing secure virtual private networks (VPNs) between the organization and the cloud.[55] For the encryption of stored medical data, the most prominent practices[53,56] involve data encryption before their transmission to the cloud. Following this practice, the organization holds and manages the data encryption/decryption keys without the need to share them with any third-party entity. This approach requires high and reliable network connection bandwidth with the cloud.[56]

Another major concern in information system security is service and data availability. Although cloud architecture and infrastructure can ensure a high level of availability of cloud services and data, the cloud provider faces many security challenges such as backup policies and fast recovery of failed virtual objects. Further, the protection of data from accidental or intentional modification is essential in cloud computing. Data integrity must be assured at all levels of the cloud computing infrastructure and services. Cloud providers protect data storage using redundant and fail-safe technologies such as RAID or RAIN (redundant array of inexpensive disks/nodes).[52] However, cloud storage clients need to implement their own measures for ensuring high integrity of the data, as cloud storage could lose or corrupt the data. One way of achieving this is by keeping a log of the state of the data (e.g., MD5 digital signatures) that can be used to check for corruption. These signatures are short descriptions of a file, computed based on its contents. For example, a simple (and bad) signature would be the number of vowels and consonants present in a piece of text. More elaborate signature functions like MD5 have the good property that it is extremely difficult to change a file in such a way that its signature will remain the same. Of course detection of data corruption is not enough—a remedial action is required for recovering from that state. For this reason clients should also keep backups (possibly through a separate cloud storage provider), thus ensuring the integrity of the data even if the cloud provider has a failure. In the SaaS model, where the provider often offers access to the data through Internet services, the clients should implement standard security protocols as well as introduce integrity controls in data transactions.[54]

Security vulnerabilities and system failures are part of the natural lifecycle of a computer system or application, and system logging is a crucial aspect of system management that facilitates identification of the causes of system vulnerabilities. System logging has three major attributes: application events, user logs, and resource monitors.[52] These logs help administrators isolate the causes of system failures, identify user activities during runtime, and monitor real-time information about the application status and virtualized and physical resources in use. Logging can be applied at all levels of the cloud infrastructure. The main focus of logging is the provision of logged information to healthcare administrators in case of a security or failure event. With regard to national legal issues, another important concern is who has access to what pieces of recorded information. In the United States under specific conditions, for example, the cloud provider is obliged to allow governmental access to recorded information. In most European countries the requirements are different, making it difficult for a European healthcare organization to trust a US cloud provider.[52]

In cloud computing, most of the responsibility of applying security measures is allocated to the service provider.[57] When medical data are stored outside the organization (or even the country of origin in some occasions), security requirements are becoming increasingly complicated and demanding. Cloud providers must follow specific measures for securing patient data, including the security of the physical data storage location, data availability and integrity, data access control, and audit.[58] From the opposite perspective, healthcare providers and research organizations must trust cloud providers when they decide to move to the cloud. The only business tool that can currently help establish a formal and legally binding business relationship of trust between healthcare organizations and cloud providers is a service-level agreement (SLA).[59] A SLA clearly describes all the measures, tools, and procedures to be followed by the cloud provider to provide secure data exchange and storage, the availability of services, access to logs, and the physical safety of the stored data. A SLA must include additional agreement points that clearly define how the cloud provider can access the stored data, what logging information will be provided to the customer, how data are securely isolated in the cloud from other customers, the geographic location of the data, how the customer can evaluate the security measures, and how the cloud provider is compliant with national and international regulations for secure exchange and storage of patient data. Starting with the SLA, a healthcare organization can evaluate offered services and security measures and decide whether to trust a cloud provider and what type of cloud service (IaaS, SaaS, or PaaS) best meets its requirements.

To gain a better understanding of security issues it is helpful to contrast cloud computing with other infrastructures, for example, the infrastructure for electricity production. Hospitals use public/private electricity providers but also have their own emergency generators in case of power failure. The hospital emergency generators are rarely more secure or reliable than the public infrastructure, but are independent of the electricity providers and hence are likely to be functional in the event of power failure. Similarly, a hospital IT system will never be as reliable and secure as a cloud operated by Google, Amazon, Microsoft, etc.—hospitals simply do not have the know-how and personnel resources to achieve that level of service. However, by investing so many valuable resources (patient data, etc.) in a single provider, the potential consequences of a successful attack (even a simple denial of service) are substantially greater. It is the same with electricity generation—attacking a power station or grid may be far more costly than simply cutting the cables that transfer electricity to a specific hospital but it can cause a lot more damage and, therefore, has a much higher payoff for an attacker. Therefore, use of cloud computing represents a trade-off between the cloud provider's better defenses against attack and the greater consequences of a successful attack. In fact, an organization might suffer collateral damage when a cloud provider is attacked just as, collateral damage to a hospital could result from an attack on a power station intended to disable the security system of a bank. The decision to move to the cloud, and how much if any service should be maintained locally, should be carefully considered. An added problem with cloud computing, as mentioned earlier, is that unlike electricity producers, cloud providers have direct access to the data and services of a healthcare institution. They could break their confidentiality promises either for business reasons or because the nation where they are located requires them to do so. A potential solution to this dilemma may become available through homomorphic encryption,[60,61] whereby the data stored on the cloud are always encrypted, even when they are undergoing processing. In this manner, the cloud provider never has access to the unencrypted data. Homomorphic encryption is not yet fully available and may never become a viable practical solution because of cost. For this reason other schemes are being considered,[62,63] to increase the data privacy guarantees without imposing a prohibitively high overhead. Another solution would be for a country or a group of healthcare institutions (e.g., all European Union hospitals) to form their own cloud infrastructure, as, for example, is currently possible through OpenNebula.[9] Such so-called "community clouds" have been proposed for users with special security, legal, or performance requirements.[1,64]

VI. CONCLUSIONS

Cloud computing is transforming ICT services by turning them into a virtual public service. It is a sign of industry maturation, just like electrical power generation moved from small private units to larger providers. From that perspective alone, use of cloud computing in the healthcare sector is inevitable, due to the increased functionality and economies of scale that can be achieved. At the same time, cloud computing introduces major advantages for health provision and research that are impossible to ignore, and these will undoubtedly accelerate its adoption.

Cloud services can deliver to end users a large spectrum of computer resources through the Internet, without the surcharge of purchasing and maintaining additional equipment. The effectiveness of these services is particularly useful in storing, processing, and sharing large databases of medical images. By using the cloud, researchers can access the resources needed for executing large-scale clinical trials involving multiple institutions. The emerging technologies of cloud computing have already attracted several researchers, clinical administrators, and software developers to move medical image archives such as PACS onto the cloud, in order to improve manageability, accessibility, and storage availability. Important impediments to this migration are the security, privacy, and ethical issues that arise through the use of cloud computing in which the management and storage of medical data are moved from a local organization to a world-accessible cloud. ICT encryption technologies and safety policies can partially ensure data security and privacy. However, cloud providers must additionally ensure that their services are compliant with national and international regulations, of both their country and the country of the client. They must also take appropriate measures to safeguard the highly private nature of

sensitive patient data that may be stored on and processed by their systems.

ACKNOWLEDGMENTS

This research has been cofinanced by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) – Research Funding Program: Thales, investing in knowledge society through the European Social Fund.

[a]Author to whom correspondence should be addressed. Electronic addresses: gkagad@gmail.com and George.Kagadis@med.upatras.gr; Telephone: +30 2610 969146; Fax: +30 2610 969166.

[1]P. Mell and T. Grence, "The NIST definition of cloud computing," Special Publication **800–145** (2011).
[2]I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud computing and grid computing 360-degree compared," in *Grid Computing Environments Workshop, GCE '08* (IEEE, Austin, TX, 2008), pp. 1–10.
[3]L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: Towards a cloud definition," ACM SIGCOMM Comput. Commun. Rev. **39**, 50–55 (2009).
[4]M. Armbrust *et al.*, "A view of cloud computing," Commun. ACM **53**, 50–58 (2010).
[5]K. P. Andriole and R. Khorasani, "Cloud computing: What is it and could it be useful?," J. Am. Coll. Radiol. **7**, 252–254 (2010).
[6]L. Wang, G. von Laszewski, A. Younge, and X. He, "Cloud computing: A perspective study," New Gener. Comput. **28**, 137–146 (2010).
[7]Q. Ling, Z. Luo, Y. Du, and L. Guo, "Cloud computing: An overview," in *Cloud Computing* (Springer, Berlin-Heidelberg, 2009), pp. 626–631.
[8]EUCALYPTUS, "EUCALYPTUS open source software license agreement," http://www.eucalyptus.com/licenses/eucalyptus-software-license-agreement.
[9]Nebula, "About the OpenNebula.org Project," http://opennebula.org/about:about.
[10]G. C. Kagadis, C. Alexakos, S. G. Langer, and T. French, "Using an open-source PACS virtual machine for a digital angiography unit: Methods and initial impressions," J. Digit. Imaging **25**, 81–90 (2012).
[11]A. Monaco, "A view inside the cloud," IEEE, http://theinstitute.ieee.org/technology-focus/technology-topic/a-view-inside-the-cloud.
[12]R. Allan, "Cloud and Web 2.0 resources for supporting research," http://tyne.dl.ac.uk/NWGrid/Clouds/.
[13]CNN, "Amazon EC2 outage downs Reddit, Quora," http://money.cnn.com/2011/04/21/technology/amazon_server_outage/index.htm.
[14]H. Tsukayama, "Amazon Web Services crashes after storm," http://www.washingtonpost.com/blogs/post-tech/post/amazon-web-services-crash-after-storm-thecircuit/2012/07/02/gJQAj0VPIW_blog.html.
[15]J. McKendrick, "Cloud computing's vendor lock-in problem: Why the industry is taking a step backward," http://www.forbes.com/sites/joemckendrick/2011/11/20/cloud-computings-vendor-lock-in-problem-why-the-industry-is-taking-a-step-backwards/.
[16]B. Darrow, "Fear of lock-in dampers cloud adoption," http://gigaom.com/2013/02/26/fear-of-lock-in-dampens-cloud-adoption/.
[17]Forrester, A. Reichman, R. Whiteley III, and E. Chi, "File storage costs less in the cloud than in-house," http://www.forrester.com/File+Storage+Costs+Less+In+The+Cloud+Than+InHouse/fulltext/-/E-RES57696?objectid=RES57696.
[18]R. Vanover, "How to determine if cloud storage is a cost savings," TechRepublic, http://www.techrepublic.com/blog/networking/how-to-determine-if-cloud-storage-is-a-cost-savings/6408.
[19]P. K. Kijewski, "Radiology IT: Applications integration vs. consolidation," J. Digit. Imaging **24**, 814–822 (2011).
[20]L. A. Silva, C. Costa, and J. L. Oliveira, "A PACS archive architecture supported on cloud services," Int. J. Comput. Assist. Radiol. Surg. **7**, 349–358 (2012).
[21]P. G. Nagy, "The future of PACS," Med. Phys. **34**, 2676–2682 (2007).
[22]*Informatics in Medical Imaging*, edited by G. C. Kagadis and S. G. Langer, 1st ed. (CRC Press, Boca Raton, Fl, 2012).
[23]P. Aguiar *et al.*, "Geometrical and Monte Carlo projectors in 3D PET reconstruction," Med. Phys. **37**, 5691–5702 (2010).
[24]D. Wallach, F. Lamare, G. Kontaxakis, and D. Visvikis, "Super-resolution in respiratory synchronized positron emission tomography," IEEE. Trans. Med. Imaging **31**, 438–448 (2012).
[25]J. Yan, B. Planeta-Wilson, and R. E. Carson, "Direct 4-D PET list mode parametric reconstruction with a novel EM algorithm," IEEE. Trans. Med. Imaging **31**, 2213–2223 (2012).
[26]T. Shepherd *et al.*, "Comparative study with new accuracy metrics for target volume contouring in PET image guided radiation therapy," IEEE. Trans. Med. Imaging **31**, 2006–2024 (2012).
[27]A. Le Maitre *et al.*, "Incorporating patient specific variability in the simulation of realistic whole body 18F-FDG distributions for oncology applications," Proc. IEEE **97**, 2026–2038 (2009).
[28]P. Papadimitroulas, *et al.*, "Realistic pathological simulations of the NCAT and Zubal anthropomorphic models, based on clinical PET/CT data," paper presented at the *54th annual meeting of the American Association of Physicists in Medicine, Charlotte, NC, 29 July–2 August, 2012*; Med. Phys. **39**, 3645 (2012).
[29]P. Papadimitroulas, G. Loudos, G. C. Nikiforidis, and G. C. Kagadis, "A dose point kernel database using GATE Monte Carlo simulation toolkit for nuclear medicine applications: Comparison with other Monte Carlo codes," Med. Phys. **39**, 5238–5247 (2012).
[30]D. Visvikis, M. Hatt, F. Tixier, and C. Cheze Le Rest, "The age of reason for FDG PET image-derived indices," Eur. J. Nucl. Med. Mol. Imaging **39**, 1670–1672 (2012).
[31]C. Toland, C. Meenan, M. Toland, N. Safdar, P. Vandermeer, and P. Nagy, "A suggested classification guide for PACS client applications: The five degrees of thickness," J. Digit. Imaging **19**(1), 78–83 (2006).
[32]Adobe Systems Inc., "Flash Player," http://www.adobe.com/uk/products/flashplayer.html.
[33]Oracle Corp., "JavaFX," http://www.oracle.com/technetwork/java/javafx/overview/.
[34]Microsoft Corp., "Microsoft Silverlight," http://www.microsoft.com/silverlight/.
[35]W3C, "HTML 5.1 Nightly," Editor's draft, http://www.w3.org/html/wg/drafts/html/master/.
[36]T. Richardson, Q. Stafford-Fraser, K. R. Wood, and A. Hopper, "Virtual network computing," IEEE. Internet Comput. **2**, 33–38 (1998).
[37]A. Fox, "Computer science. Cloud computing–What's in it for me as a scientist?," Science **331**, 406–407 (2011).
[38]C. Vecchiola, S. Pandey, and R. Buyya, "High-Performance Cloud Computing: A View of Scientific Applications," in *10th International Symposium on Pervasive Systems, Algorithms, and Networks (ISPAN)*, (IEEE, Kaohsiung, 2009), pp. 4–16.
[39]G. Brumfiel, "High-energy physics: Down the petabyte highway," Nature (London) **469**, 282–283 (2011).
[40]K. Jorissen, W. Johnson, F. D. Vila, and J. J. Rehr, "High-performance computing without commitment: SC2IT: A cloud computing interface that makes computational science available to non-specialists," in *8th International Conference on E-Science (e-Science)* (IEEE, Chicago, IL, 2012).
[41]L. Dai, X. Gao, Y. Guo, J. Xiao, and Z. Zhang, "Bioinformatics clouds for big data manipulation," Biol. Direct. **7**, 1–7 (2012).
[42]A. Shanker, "Genome research in the cloud," OMICS **16**, 422–428 (2012).
[43]R. Chow *et al.*, "Controlling data in the cloud: Outsourcing computation without outsourcing control," in *ACM Workshop on Cloud Computing Security (CCSW '09)* (ACM, 2009), pp. 85–90.
[44]D. J. Abadi, "Data management in the cloud: Limitations and opportunities," IEEE. Data Eng. Bull. **32**, 3–12 (2009).
[45]EU, "European Union Directive on data protection," in *Official Journal of the European Communities* (EU, Brussels, Belgium, 1995), Vol. L.281, pp. 31–50.
[46]Stanford, "Internet research ethics, in Stanford Encyclopedia of Philosophy," http://plato.stanford.edu/entries/ethics-internet-research.
[47]J. Timmermans, B. Stahl, V. Ikonen, and E. Bozdag, "The ethics of cloud computing," http://www.academia.edu/646838/.
[48]J. Grimes, P. Jaeger, and J. Lin, "Weathering the storm: the policy implications of cloud computing," http://nora.lis.uiuc.edu/images/iConferences/CloudAbstract13109FINAL.
[49]W. Pieters and A. van Cleeff, "The precautionary principle in a world of digital dependencies," IEEE. Comput. **42**, 50–56 (2009).

[50]L. M. Kaufman, "Data security in the world of cloud computing," IEEE Secur. Privacy **7**, 61–64 (2009).
[51]F. Prior, M. L. Ingeholm, B. A. Levine, and L. Tarbox, "Potential impact of HITECH security regulations on medical imaging," in *31st Annual International Conference of the IEEE EMBS, Minneapolis, MN, 2009*, (IEEE, 2009), pp. 2157–2160.
[52]M. Zhou, R. Zhang, W. Xie, W. Qian, and A. Zhou, "Security and privacy in cloud computing: A survey," in *6th International Conference on Semantics Knowledge and Grid (SKG)* (IEEE, Beijing, China, 2010), pp. 105–112.
[53]R. Zhang and L. Liu, "Security models and requirements for healthcare application clouds," in *IEEE 3rd International Conference on Cloud Computing (CLOUD)* (IEEE, 2010), pp. 268–275.
[54]S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing," J. Netw. Comput. Appl. **34**, 1–11 (2011).
[55]S. G. Shini, T. Thomas, and K. Chithraranjan, "Cloud based medical image exchange-security challenges," Proc. Eng. **38**, 3454–3461 (2012).
[56]S. G. Langer, "Challenges for data storage in medical imaging research," J. Digit. Imaging **24**, 203–207 (2011).
[57]J. Philbin, F. Prior, and P. Nagy, "Will the next generation of PACS be sitting on a cloud?," J. Digit. Imaging **24**, 179–183 (2011).
[58]E. J. Schweitzer, "Reconciliation of the cloud computing model with US federal electronic health record regulations," J. Am. Med. Inform. Assoc. **19**, 161–165 (2012).
[59]B. R. Kandukuri, V. R. Paturi, and A. Rakshit, "Cloud security issues," in *International Conference on Services Computing – SCC '09* (IEEE, Bangalore, India, 2009), pp. 517–520.
[60]C. Gentry, "Computing arbitrary functions of encrypted data," Commun. ACM **53**, 97–105 (2010).
[61]E. Naone, "Homomorphic encryption," http://www2.technologyreview.com/article/423683/homomorphic-encryption/.
[62]C. Curino *et al.*, "Relational cloud: A database-as-a-service for the cloud," paper presented at the *5th Biennial Conference on Innovative Data Systems Research, Pacific Grove, CA*, 2011, pp. 235–241 (www.cidrdb.org).
[63]R. A. Popa, M. S. Redfield, N. Zeldovich, and H. Balakrishnan, "CryptDB: Protecting confidentiality with encrypted query processing," paper presented at the *SOSP '11 Proceedings of the twenty-third ACM Symposium on operating systems principles, Cascais, Portugal* (ACM, New York, NY, 2011), pp. 85–100.
[64]D. Kovachev, D. Renzel, R. Klamma, and Y. Cao, "Mobile community cloud computing: Emerges and evolves," in *11th International Conference on Mobile Data Management (MDM), 2010* (Aachen, Germany, 2010), pp. 393–395.