CaseServer: a Multipurpose Infrastructure for Clinical Cases Collections Management

M. dos-Santos
Department of Librarianship and Documentation at School of Communications and Arts,
University of Sao Paulo
Av. Prof. Lucio Martins Rodrigues, 443, rm 257
05508-020, Sao Paulo-SP, Brazil

ABSTRACT
Medical diagnosis is an outcome got from a complex process that implies multiple inferences from sets of clinical information. A large body of research has been conducted to develop tools in order to aid analyzing the amount of clinical information. In this paper, a system - called CaseServer - is presented. It supports radiology teaching and research practices. The initial intention was to create an electronic and MIRC-compliant infrastructure to provide integrated and context-based access to all relevant patient data at the time of image interpretation. A successful implementation has integrated information collected from various sources. Furthermore, the application of role-based coupled with context-based principles to acquisition, organization, control, dissemination and use of clinical cases in teaching and research has brought out a multipurpose infrastructure which promotes and makes feasible the usage of clinical cases within academic and research communities.

General Terms
Medical images, clinical information, radiology teaching.

Keywords
Medical image database, clinical case, electronic teaching file, MIRC.

1. INTRODUCTION
Making diagnosis is the fundamental step of a health care delivery workflow. It is on the basis of medical decision-making process to settle a course of action. Accordingly, just as stated by Kummar et al. [1], "diagnostic decision-making has traditionally involved using evidence from a patient's data (image and nonimage) coupled with the physician's prior experiences of similar cases". Medical diagnosis is therefore resulted from a complex process that had implied multiple inferences from sets of clinical information. To this end, it is required from physicians to have integrated access to all relevant patient data at the time of interpretation.

Currently, thanks to the use of medical imaging devices and other diagnosis systems, health care professionals and researchers can collect complex data in ever increasing amounts, which has resulted in a wide increase of clinical information in both variety and volume. As a consequence of the wide use of clinical information in daily patient care delivery routine, health care information systems (e.g. hospital information system[HIS], radiology information system[RIS], picture archiving and communication system[PACS], electronic health records[EHR], etc) are developed to archive the large and increasing volume of clinical information generated in clinical practices.

Although current health care information systems have rendered great benefits concerning patient care, these systems are usually not designed for using in teaching and research practices. Naturally, they are designed focusing on archiving and making patient's data available to the professionals at the moment of health care delivery. On the one hand, these systems store rich collections of clinical information which could have multipurpose uses (e.g. medical educational). On the other hand, clinical practices environments are highly busy and by employing clinical information in contexts rather than health care is required protecting patient privacy and information security actions, as mandated by regulations (e.g. Health Insurance Protection Agency [HIPAA] [2]).

Remarkably, medical education and research extensively employ clinical cases. In short, this approach (also named case-based learning [3]) employs clinical information from health care routine (real patients) to illustrate pertinent teaching points and highlights parts of the physician's decision-making process. However, the identification and selection of interesting cases for teaching purposes are traditionally done in a busy clinical practice environment. With attention to teaching files, radiology learning is widely dependent on images (usually key-images) and written-up notes (case presentation and reports). In order to train radiologic diagnosis, the learners (medical students, residents, and trainees) are exposed to amounts of imaging studies (visual information) and asked to interpret images or provide a read.

Beginning the research within a radiology teaching file project, the intention was to create an electronic infrastructure to provide integrated access to all relevant patient data at the time of image interpretation so that, in radiology residence programs, residents can examine all available clinical information. This will also allow residents to know how to reach sound conclusions and communicate these conclusions to radiology information consumers. The purpose of this paper is therefore to describe a system, called CaseServer, that supports radiology teaching practice.

2. MATERIALS AND METHODS
An important step aiming to make available collections of clinical information to different purposes is the design and implementation of an electronic database infrastructure, which integrates fragments of information existing in patient history records, clinical reports, laboratory tests, and ordered procedures. In consideration of various user groups have diverse interests on clinical information, it is mandatory to adapt the information presentation to the different users’ need and contexts of use. To accomplish this, firstly it was fundamental understanding by whom, when, what, where,
why, and how clinical information is used (particularly medical images). In this work, this task has been accomplished based on the view provided by [4] where the authors investigated the medical image use and search behaviours of health care professionals.

By applying experiences reported in other studies [5] - [9] our primary hypothesis was based on the development of an electronic database infrastructure as a repository for clinical information (patient's clinical history records, laboratory tests, reports of visits, image-based studies and reports, ordered procedures, treatments, follow-ups, etc) which, first of all, could be used in radiology teaching activities. Further, it was realized that this set of clinical information - when appropriately archived - could be used in other purposes related to teaching and research activities.

The infrastructure was designed as a modular and extensible medical database, integrating heterogeneous clinical information from different domains and sources. In particular, this infrastructure has been developed such as an instance of MIRC teaching file system [5] and [10]. Also, the same infrastructure offers a set of customized services to be used in teaching and research contexts.

2.1 Conceptual View
Just as Kohli and Bradshaw [11] stated “educating radiology is a complex procedure that requires hours of self-study, an enriching didactic environment, and extensive clinical experience”. This enriching didactic environment should be very similar to the real day-to-day clinical practice environment, where the learners are exposed to real clinical cases. Therefore, to teach radiology it is essential to have a set of clinical data (images and related documents) organized and readily available to use.

Based on that statement and the daily radiology department routine, the proposed infrastructure considers the multiple issues related to clinical information production and use in the workflows. The design started investigating the basic workflow of a radiology department [12] and [13] aiming to identify frequent users, their roles, information needs and contexts of use. The radiology workflow initiates when a referring physician orders a image-based examination. The order is received and images are acquired. Accordingly, it follows with image interpretation and finishes delivering results (radiology reports). In such context, documents containing information about patient's clinical history, visits, and early studies are highly desirable to be available for consultation.

To archive each clinical case, a database system was designed to store images and related information. At that moment, three important issues concerning medical knowledge were identified and considered. They are [8]: (1) parameters (imaging protocol) used during the image acquisition process (e.g. imaging modality, examined body part, functional system under investigation, and reasons which support the study/procedure order); (2) regions of interests and findings in the images; and (3) information about the findings and their properties, as well as the relationship among these findings. The datasets with reference to these issues are also archived in the database system along with radiological studies. The infrastructure, which its conceptual view is summarized in the Fig. 1, is accessed through a standard web browser, from any personal computer connected to the Internet. This allows the users to perform a set of tasks (e.g. data entry [case authoring], medical image viewing, experts can assess the contents of clinical cases, and clinical information querying).

The system is developed to offer access to a group of heterogeneous users with diverse interests on clinical information. For this reason, the granted permissions of access are associated to users' role and context of use. In this way, the set of information is presented using appropriated and customized interfaces. Since clinical cases are created and

Fig. 1 - System overview: the conceptual view and data flow.
updated by many users, it is necessary examine (quality control) all stored datasets to offer relevant and reliable clinical information. A group of reviewers inspects, comments, and flags each stored dataset.

Although this infrastructure supports archiving single radiological images (key images) and, also, in order to have minimal disruption of the radiology department's workflow, as part of a clinical case, the entire radiological study and post-processed images should be stored by using the Digital Imaging and Communication (DICOM) standard [14]. By extension and employing this standard, DICOM studies can be retrieved and processed or viewed in diagnostic workstation or any other medical image DICOM compliant software. Similarly, each text-based clinical document (e.g., radiological report) is stored as an extensible markup language (XML) document.

In order to import and retrieve studies directly from imaging scans, PACS, and viewing workstations, a DICOM client/server peer was implemented and made ready to operate. This allows the infrastructure capture and archive radiological studies using DICOM communication protocol. All imaging data (DICOM and non-DICOM), including related information (summaries of DICOM headers), stored into the infrastructure can be viewed by using up a web-based medical image viewer.

In a database infrastructure, information retrieval is a fundamental functionality. On the back end, this infrastructure sets up an application programming interface (API) designed to create tools (plug-ins) to process and analyze text-based clinical documents. Profiting from this API, developers may create plug-ins that operate on XML documents and add such plug-ins as new feature to the infrastructure. An indexing mechanism – which automatically extracts some experts and relevant concepts (terms) from radiological reports – was created to organize reports and contribute to process query requests. The found and extracted terms are expressed in an acceptable terminology employing controlled vocabularies (e.g., Radlex [15]). Conversely, a retrieval mechanism was also created to retrieve clinical information by means of customized query interfaces. Another API, based on ImageJ's API [16] and [17], has provided means to implement medical image processing algorithms (as plug-ins) which straightly operate on image data structure of the database.

2.2 System Requirements

Wong and Tjandra [18] presented a set of issues which requires attention and are relevant in medical domains. Some of these issues are: fragmented sets of clinical information (clinical information is traditionally stored in heterogeneous and not always integrated systems), create interoperable interfaces with various clinical applications, archive heterogeneous datasets, patient privacy, clinical dataset security and conformity, use of open standards, and compliance with clinical systems. Such issues address the essential step towards integrating patient data acquired from different sources.

Ideally, a suitable infrastructure for clinical cases archiving and management should have a plenty of features, like [8]: the datasets should be uniquely identified; offer customized query and retrieve tools; carry out datasets (complete clinical cases) on demand; provide tools for text and image-based information viewing, and clinical case authoring; provide context-sensitive and role-base user access control [19]; offer tools for medical images and health care documents de-identification; retrieve radiologic studies and related information from health care information systems (e.g., HIS, RIS, and PACS); offer means to be integrated into clinical workflow; provide vendor-neutral tools for case input; the clinical cases should be reviewed and assessed by experts (quality control); rapid query processing to summarize contents found in various repositories; and content-based tools for image and text-based information retrieval.

Specifically for teaching files, Scarsbook et al. [20] present a number of characteristics (versatility, accessibility, ease of use, compatibility, and flexibility), which has been added to this infrastructure in order to maximize its usefulness. Another important system requirement is related to distance learning regulatory aspects.

2.3 System Architecture and Implementation

In order to create an extensible and multipurpose clinical database infrastructure, the system was designed and implemented in nine major modules: (1) clinical case database; (2) learning management system; (3) query, retrieve, and viewing tools; (4) learning management system; (5) usability framework; (6) MIRC broker - MIRC broker is an application designed to query, retrieve and share cases by using MIRC schemas [21]; (7) DICOM client and server peer; (8) role and contextual-based access controls whose basic notion of this application is that granted permissions are associated with roles and contexts of use (where, why, what, by whom, when, and how clinical information is used), by using the directory service application named Lightweight Directory Access Protocol (LDAP) [25]; and (9) medical image and text-based documents processing tools.

Moreover, in this extensible infrastructure, it was created four major user and system interfaces to provide means and facilities to use the services available in this infrastructure. These interfaces are: 1) context-based GUI - that offers web-based user interfaces and is context-sensitive; 2) developer interface - new services and features are added to the infrastructure by means of plug-ins, which are developed using APIs that provide standard methods to manipulate images and text-based documents; 3) DICOM - it is an interface designed to communicate with DICOM compliant applications; and 4) MIRC - it is an interface designed to set up the MIRC broker module, setting schemas and adding MIRC peers.

![Fig. 2 - CaseServer architecture layers](image-url)

The system was implemented as an Internet portal, which is accessible through a standard web browser, allowing the users to perform many tasks. In terms of architecture, the system
organized in a three-tiered application model (see Fig. 2). This architecture - whose tiers are: data (it is responsible for data storage and persistence tasks); and it manages access to the clinical cases database and information about users stored in the LDAP system; application (the set of applications and rules to operate on clinical datasets); and presentation (it is what the users see and interact with) - features the main stages of the datasets flow. The architecture is scalable and designed to handle large amounts of diverse clinical data.

The proposed architecture is based on and implemented using open medical standards (e.g. DICOM [14] and Health Level Seven [HL7] [26]), local and international regulations for patient privacy and data protecting [2], [12] and [24]. Thus, in the implementation, open source development tools (Java, Java Server Pages, PostgreSQL, XML, etc) were used in a Linux environment. The dcm4che [27] (version 2) DICOM toolkit has been used to handle DICOM objects. The Unified Modeling Language (UML) has served as the set of symbols (language) used in the application modeling process. Also, the user-centered development lifecycle proposed by Kunert [28] has been used to design and assess the graphic user interfaces (GUI).

3. RESULTS
A version the proposed infrastructure, with all working modules and services, was implemented. The range of offered services in this infrastructure has been developed so that these services are available on an Internet Portal. Now, access to and use of this system are restricted to few users, who have assessed and provided feedbacks about the system operation.

Fig. 3 shows three screenshots of the user interface. Fig. 3(a) is the main interface, which is available to the users when they just logged in. In this interface, it is presented a list of recently added cases where the user can query, retrieve, and view clinical cases. The system provides means to create and adopt policies for clinical case retrieve and viewing. Depending on this policies and rules for use, not all clinical cases will be available for viewing by any user. Similarly, tools (e.g. clinical case editing) are available only to users with permission to create and edit clinical cases. All permissions are granted by the role and context-based access control module. Such permission are based on user role and context of use.

When a clinical case is retrieved and opened, the set of information related to it is presented and the studies (e.g. laboratory tests, image-based studies, and ordered procedures) are presented in the same GUI, as shown in the Fig. 3 (b). Clinical cases, which are retrieved from MIRC database peers, have its contents mapped to the presentation model used in this infrastructure and are identified as “MIRC Case”. Fig. 3(c) shows a report retrieved from a MIRC database peer using MIRC schemas.

For each image-based exam, in a specific tab, the image series of this study are displayed as thumbnails (see Fig. 4(a)). Each thumbnail represents a series of images and some of them (e.g. echocardiography series) are displayed as a video. This approach (using static and dynamic thumbnails) has offered a preview of the entire study and its series.

![Figure 4](image-url)

The medical image viewer application (Fig. 4(b)) is launched when a thumbnail (Fig. 4(a)) is clicked. This viewer provides a set of tools designed to view and process medical images. By using DICOM files, the images are viewed in a near diagnostic quality fashion. Therefore, the image datasets generated in an imaging scanner are displayed in the viewer without file format conversions.

Many radiologists and residents are acclimated to using Osirix [29] image viewer or another DICOM viewer (e.g. Merge Efilm®) to view and/or process medical images. They may continue using these software, and via query and retrieve DICOM services retrieve studies from this infrastructure.

Currently, fifteen frequent users (physicians, radiology residents, mastered radiologists, medical image processing researchers, and software developers) have used the CaseServer. There are approximately 146 complete clinical cases assessed and certified. Most of such cases were acquired from public repositories, personal collections, and migrated from other applications [30]. Of a total of 303 image-based studies, 142 (46.87%) are computed tomography (CT) studies; 32 (10.56%) from magnetic resonance imaging (MRI); 47 (15.51%) from ultrasound (US); 6 (1.98%) from x-ray angiography (XA); 15 (4.95%) from nuclear medicine modalities (NM); and 61 (20.13%) from r-rays (RX, including mammographic images). Most of cases have more than one study and, in some cases, studies from different imaging...
modalities. The majority of studies archived are from CT, RX, and US. This current version has proven useful in distinct purposes. The major results and use of this system are summarized as it follows:

a) **radiology learning tools**: the system has offered easy and integrated access to all relevant clinical data at the time of image interpretation. The radiology learners can interact with experts via chat and forums provided by the learning management system (Moodle [31]).

b) **data mining in clinical datasets**: correlating image, body part, pathology, and information contained in an image is a difficult task to offer it in a practical way.

c) **improving communication in radiology**: the current fifteen users comprise a multidisciplinary team that has also contributed to rethink the processes of communication in clinical environments.

d) **users’ feelings**: although this is a ongoing project, the users have acknowledged advantages over image-oriented teaching files and medical image databases. Conversely, they think gathering information to create a complete clinical case is time consuming, since the information are scattered through different systems. Developers have identified in this infrastructure great opportunities to create tools for clinical information management.

4. DISCUSSION

Vannier, Staab and Clarke [32] highlight the need for creation, expansion, integration, and sharing contents of medical image archives. According to their statement, these archives “have become central to advancement in many fields” (e.g. radiology, pathology, and image processing). The presented infrastructure is based on health care delivery and radiology workflows to archive clinical information. With attention to diagnosis decision-making this infrastructure was also designed to archive the patients’ data (laboratory tests, ordered procedures, entire radiological studies [images, key-images, and reports], etc.). Its design focus on archiving datasets from heterogeneous medical databases, taking into account the diversity and amount of clinical information. Also, in order to complement the set of features traditionally available in medical databases, the cases’ follow up and quality control interface (applying a peer review model) were implemented in this infrastructure. Coupled with, by using context-based user interfaces, the datasets can be used in multiple purposes, such as radiology teaching and assessing medical image processing methods.

In fact, medical archives are a rich source of information and their contents may particularly be harnessed for teaching and research purposes. It can be noted, however, that it is important to devote efforts in order to make the users aware with useful resources. The main focus of this work was on offering means to increase accuracy of diagnosis. In terms of image management workflow, the work started considering the major aspects related to this task (ordering an image-based examination, interpretation, reporting, and delivery of results).

Contextual factors affect clinical information needs and search behaviours. Accordingly, as suggested by Wilson [33] “information-seeking behaviour results from the recognition of some need, perceived by the user” (in clinical environments, the user may be physician, surgeon, radiologist, student, nurse, or research, among others). Such needs, either in scope of information or level of access, depends on specific context of workflow and particular requirements of health care team collaboration. This have to do, traditionally, with the relationship among organizational objectives, management process (workflow) and information needs [34].

In teaching and research environments, it is essential integrating fragments of information captured in clinical reports, laboratory tests, and ordered procedures into a same user interface [35]. While searching vast clinical information collections, users can often be frustrated with the discrepancy between their needs and search results [36]. Müller et al. [4] suggested “only when information needs are understood, real tasks can be created to evaluate retrieval systems”. In this work, why, where, and when were also added. For this reason, by using the customized set of service, regarding the user’s context, the infrastructure has provided interactive and integrated access to all available clinical information about the clinical case. Thus, this set of feature can offer the opportunity to radiology learners to train how to reach well-informed conclusions.

As a source of clinical information, this infrastructure has harnessed for other possibilities of use such as support evidence-based research activities, provide clinical datasets for assess medical image processing algorithms, content-based image retrieval field of test, and understanding the medical image consumers (users) needs and their contexts and then apply these information into indexing practices when designing medical image retrieval systems. Some features has offered means to reuse ImageJ’s plug-ins and assess medical image processing algorithms using real and large amount of clinical cases. In this way, the infrastructure is implemented as an educational resource which can be used in medical education practices, dissemination of medical knowledge, and other purposes as well.

5. CONCLUSION

CaseServer is an infrastructure which enables to archive complete clinical cases and use them for teaching and research. Successful implementation has integrated information collected from various sources. This infrastructure has also provided means to assess medical image processing methods by using diverse and comprehensive dataset. The plug-in interfaces enable to add new features and services to CaseServer. The application of role coupled with context-based principles for the acquisition, organization, control, dissemination and use of clinical cases in teaching and research has brought out a multipurpose infrastructure which promotes and makes feasible the usage of clinical cases within academic and research communities. An initial evaluation through comparison with existing similar medical image databases systems has shown this infrastructure can be effectively used to archive and manage clinical information to be used in multipurpose tasks. Further research is required to incrementally develop additional features related to the clinical information users’ needs and usage contexts, helping them to construct meaning from the patients’ data (image and non image). To sum up, CaseServer can be used to support radiology teaching practices and researches, thus offering context-based interfaces and access to a MIRC compliant database which archives clinical information in an optimized and organized fashion, all aligned with the clinical, radiology, and research workflows.

6. ACKNOWLEDGMENTS

This work was supported by a grant from São Paulo Research Foundation (FAPESP), Brazil.
7. REFERENCES


