PACS, iSyntax and iSite

J. Huffman Senior Fellow, Philips Healthcare Informatics, Foster City, CA, USA.

▶ **Early PACS installations were expensive and confined to the radiology departments.**

PACS

PACS stands for "Picture Archiving and Communication System". The term is traditionally used to describe the system which handles the archiving of all medical images generated by a hospital. The primary purpose of PACS was to end the use of film as the legal mode of archiving medical images, and to provide the ability to diagnose from these images with a computer. Over the years, the interpretation of the term PACS has expanded to include the workstation, the network, the archive and in many cases parts of the medical record.

The history

The first PACS systems were developed with U.S. government funding in the 1970's. The primary digital modality was X-ray, with images scanned into the system at a resolution of approximately 2048 x 2536 (2 K x 2.5 K). In those days, CT and MR images were many millimeters thick and scans were only a few tens of slices. There were no defined standards for image communication between systems, and all diagnosis was done on film with lightboxes. The medical record was a folder with papers and film.

In those days, computer technology was still rather primitive: program entry was via paper tape and punch cards, disc capacities were measured in megabytes and cost thousands of dollars, 8 MB of RAM came in multiple modules and cost over $1000, a 9600 baud modem cost $1000, graphics workstations were commercially non-existent, and one mega-flop was a super computer. Of course, these conditions rapidly improved over the next decade, when the first commercially available PACS were developed. However, the immense resource requirements of the PACS application led to a persistent divergence of the PACS systems from the mainstream information systems which became ubiquitous in the same timeframe.

▶ **Early PACS installations were complex and required specialized support.**

During this period, the workstation, followed by the desktop workstation and then the personal computer, all came into existence. Computational capabilities were still low, disc space and RAM were still extremely expensive, but text editors and information systems to facilitate patient information and billing were commercially available and cost-effective. The PACS system, when implemented – and there were very few implemented – resided exclusively in the radiology departments, cost millions of dollars and had marginal benefits. Because the systems were only available in the radiology department, film still had to be printed for distribution to other departments and referring physicians.

Network and disc speed limitations meant that any use of images had to be scheduled, and the data had to be routed in advance to a specified workstation, due to the latencies involved in fetching. The systems that were implemented comprised exotic hierarchical storage systems that were outdated months after installation, the most expensive direct network connections to the radiology workstations, and the workstations themselves, which were the most expensive graphics stations on the market. Due to the unique nature of this combination of exotic hardware, a cadre of support personnel were required to maintain the systems. All-in-all, this was barely a proof-of-concept and certainly not a viable solution. It is not surprising that PACS took another decade to develop any market penetration.

During the second decade of PACS, hierarchical storage systems became more widely available, disc and memory became less expensive, and graphical workstations were widely available. Nevertheless, all of these components were extremely expensive and required specialized support.

The problem

The history of the development of PACS, and the divergence from the information systems which were universally deployed in this same timeframe, led to an accepted description of the PACS problem as:
There are hundreds of MB of medical data on a server, that a doctor wants to be able to read within two seconds, and then wants to be able to manipulate multiple 2K x 2.5K images in real time.

This of course leads to the requirement for the systems to be server systems with vast arrays of expensive, rapid-access discs, with direct optical connections to a graphics workstation costing hundreds of thousands of dollars in order to be able to manipulate the images in real time.

There were improvements with the adoption of complex routing rules to pre-fetch data to a particular workstation for a doctor to use. However, if a workstation went down – which was quite common in that era – or a doctor could not make it to the workstation room, film still had to be printed. Consequently, not only were there no cost savings, but an entire layer of support costs was added to the existing hospital systems.

This led to the continued non-adoption of PACS for many years. And this is where Stentor and iSyntax came into play.

The problem re-stated

The insight that led to the development of the iSyntax technology and the creation of Stentor was that the statement of the PACS problem was incorrect, as it was not directed towards solving the workflow problem of the institution. The correct formulation of the PACS problem is: *There are hundreds of MB of medical data on a server, that a doctor wants to query from a workstation in real time.*

In other words, if a doctor is sitting at a workstation with a screen size of, for example, 1280 x 1024, then if the PACS system is able to "fill" this window with the requested information interactively, that solves the problem.

An additional insight was that there were two ways to solve the increasing divergence of the PACS and Information Systems in medical institutions:
- either enable the exotic PACS systems, that were not cost-effective nor integrated into the institutional workflow, to handle information
- or have a technology that makes images just another piece of information to integrate into the ubiquitous information systems.

Stentor adopted the latter approach.

Stentor

Founded in 1998 in Brisbane, California, Stentor was established as the commercial entity to leverage the technology and workstation design developed at the University of Pittsburgh Medical Center Informatics Lab, run by Dr. Paul Chang. Unlike traditional PACS solutions offered by the industry, the Stentor solutions were designed by doctors for doctors within a medical institution. This led to solutions which were far more closely coupled to healthcare workflow than the competing platforms. In addition, the iSyntax technology allowed these applications to run on off-the-shelf PC's – and have a performance comparable to that of workstations costing twenty times as much.

The novel workstation design was not the only innovation which Stentor brought to the market. The overall system design incorporated many aspects of the traditional RIS functionality to support departmental workflow without a full RIS present. This greatly increased the usefulness of the standalone system.

An additional differentiator of the iSite system is that it was designed to be deployed and monitored remotely. Field support is not typically used to install, upgrade or monitor any iSite systems. This capability is designed into the software.

But the most revolutionary aspect of Stentor's introduction to the market was the business model. Even were traditional PACS to have fulfilled the functional promises that had been made – they were still immensely expensive. Stentor eliminated the cost of entry for hospitals to have this type of solution by packaging it as a service, rather than a product.

Stentor came up with a per-study fee for storage and distribution of the data through iSite. If an institution wanted another workstation, there was no license fee – they just purchased another computer and loaded the software. Upgrades were included in the per-fee charge – no hidden costs. The institution could then pay for the system as they used it. This model uniquely aligned the interests of the customer and the supplier – producing one of the most loyal customer bases in the industry – and forcing Stentor to always provide the highest quality applications.

The previously existing PACS technology required specialized computer workstations costing as much as $120,000 each to handle the volume of data required and manipulate the images in real time without the iSyntax technology. In addition, again without the iSyntax technology, high performace local networks were required to distribute this volume of data – causing limited distribution of the solutions and expensive infrastructure upgrades. Lossy compression was used to overcome the network issues – but this just provided corrupted data to the workstation – which still had to manipulate these large images in real time. All of these issues made the traditional PACS solutions prohibitively expensive for all but the nation's biggest medical

▶ **The insight that led to iSite was that developers were addressing the wrong problem.**

▶ **iSite is packaged as a service, rather than a product.**

1

Figure 1. The iSyntax technology makes it possible to display high-quality medical images on a standard PC.

centers. And even those institutions could not afford to use the expensive workstations outside their radiology departments.

The iSyntax technology makes it possible to distribute and view medical images over ordinary networks on desktop PCs (Figure 1). Dr Chang came up with the inspiration for the iSyntax approach after visiting a factory that had done away with its parts warehouse by adopting "just-in-time" delivery of its supplies. Because doctors do not focus on an entire radiological image at one time, it is only necessary for the system to deliver the data for the region of interest at the moment that it is being viewed.

Medical images can be distributed over ordinary networks and viewed on desktop PCs.

The iSite software incorporating iSyntax technology was first deployed at UPMC Presbyterian and has subsequently been deployed at over 300 medical institutions in the United States and Europe.

The iSite system is designed to give all physicians within a healthcare system - not just the radiologists - access to full-quality diagnostic images anywhere and at any time. It ends the practice of "patients as couriers" to take X-rays from one doctor or institution to another. Patients and doctors at a community hospital in the UPMC system, for example, are able to get immediate consultations with subspecialists at another hospital viewing the same images at the same time.

iSite gives physicians access to full-quality diagnostic images anywhere, at any time.

In August 2005, Stentor merged into Philips Medical Systems' Healthcare Informatics business, and is now the Radiology Informatics Business Group for Philips, headquartered in Foster City, California. The iSite platform is now marketed worldwide by Philips Medical Systems.

iSyntax and iSite

iSyntax is the underlying representation of images residing on a server, and the communication protocol used to transmit the requested information to a client in the iSite system. The images are represented by wavelets (see the Intermezzo) allowing the maximum information in the signal to be represented by the minimum number of coefficients, without loss of information.

The typical way to optimize the solution to any complex problem is to examine each component of the problem and optimize the steps. This is what led to the impractical early PACS solutions. When there is a complex system, such as information workflow in a hospital, the solution to the whole problem, i.e. the system optimization of a solution, is rarely the same as the individual component optimization.

The iSite solution is designed to optimize the system (i.e. workflow) problem in the healthcare environment. The design of the iSite system and the iSyntax representation has a number of critical characteristics:

- The representation is scalable. Any image or sub-image can be retrieved with bandwidth and computation proportional only to the number of pixels requested – independently of the size of the original data set.
- The server access from the client is stateless. This allows load balancing and optimal use of available resources. Human interaction with computers is at a snail's pace relative to the available computational and bandwidth resources. Keeping a direct connection to a resource with long idle periods is inefficient.
- The server access to data is computation-free.

iSyntax: the essentials

Wavelets

In the same way that an x- and y-coordinate can represent a point in a plane, the branch of mathematics known as "functional analysis" uses coordinate systems called "basis elements" to represent functions. The sine and cosine functions in the Fourier transform are the best recognized example of this approach. Any unknown function can be represented by these basis elements multiplied by a series of constants. The characterization of the function is then the characterization of a known set of these basis functions.
Wavelets are the mathematically optimal way of representing any signal. "Optimal" means that the maximum information in the signal can be represented by the minimum number of coefficients. It should be noted that there are an infinite number of wavelet systems which are parameterized. A particular system of basis elements with desired characteristics can be found for a target application.

One major advantage of the wavelet representation is that geometric detail is preserved – that is, the transform coefficients preserve the relative positions of features (see Figure *I*). This allows extraction of specific coefficients to reconstitute a specific feature in an image (see Figure *II*). Additionally, this is well matched to the human visual system and the way in which images are perceived. Frequency-based methods, such as the DCT in the original JPEG format, create artifacts that are easily seen as they corrupt the geometry of the original image. ▶

I

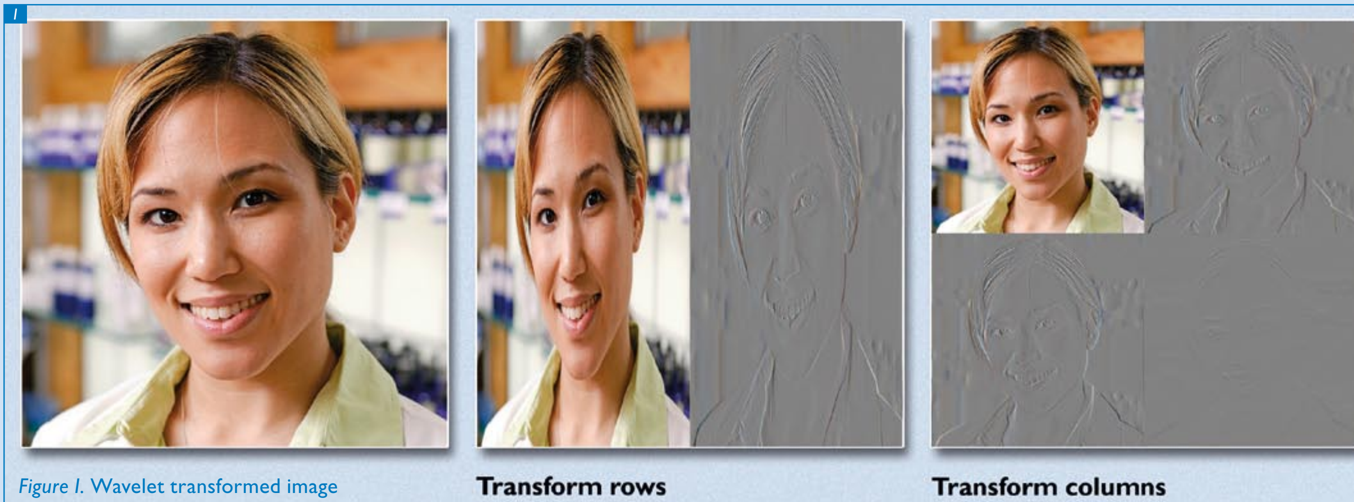

Figure I. Wavelet transformed image

II

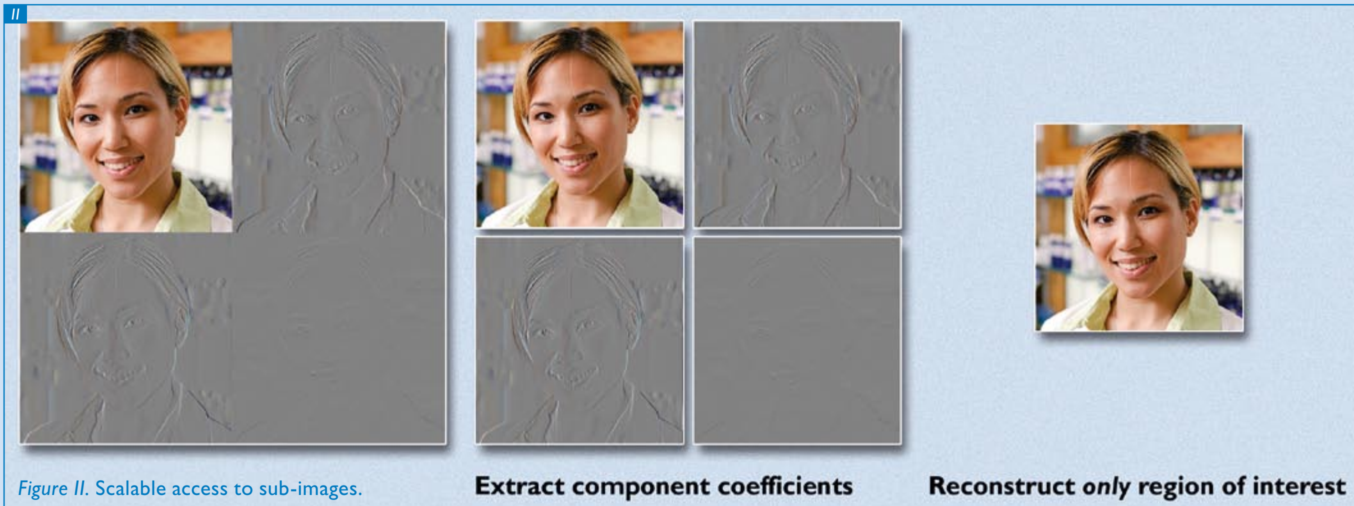

Figure II. Scalable access to sub-images.

Compression

Image compression essentials

An image is represented in a computer by a matrix of pixel values, either monochrome or multi-spectral, for example in color. Since most images of interest are created by reflected or absorbed light or other spectra such as X-rays, sampled values in an immediate geometric vicinity tend to be related, or correlated. This means that pixels that are close geometrically tend to be either similar in value, or completely different. This correlation implies redundancy in the representation of the image. The process of image compression tries to remove the redundancy in the description of the image required to reconstitute the original sampled values, in order to reduce the amount of memory or bandwidth required for storage and transmission.

There are two distinct classes of image compression methods – lossless, and lossy.

Lossless compression essentials

Lossless compression retains sufficient information to reconstitute the image exactly. Lossless image compression has absolute limits that no method can exceed. The branch of mathematics that defines this limit – information theory – is extremely mature. There is an "information content" in an image which can be measured in bits and is the absolute upper limit for how much a signal can be compressed – as long as the data is not directly derived for some easily predicted model. This "information content" is defined as:

$$\Sigma -p_i \lg (p_i)$$

I.e., the sum of the probabilities of the individual value, times the lg (log base 2) of the probability. For example, an image of any single value has a probability of 1 for that value, and an information content of 0 bits – there is no information in the image. If an image were comprised of 50% zeros, and 50% ones, then the information content would be:

$$-0.5 * \lg (0.5) + -0.5 * \lg (0.5) = 1 \text{ bit of information}$$

This formula quantitatively describes the amount of independent information in the sampled data per sample.
If there is a predictive model of the data, then the probabilities used in the equation are the conditional probabilities of the independent symbols. I.e., if a 1 always follows a 0, then the probability of the 1 after the zero is 1, and the information content is 0.

The "art" of image compression is two-part – either in the definition of a model that improves the predictive probability of the "next" pixel, or in the method used to reorganize the data into a representation more amenable to compression. In either approach, the information content of the image as a fundamental limit of compressibility cannot be exceeded.

As it turns out, a wavelet representation of the image is close to the optimal "re-organization" of the image information for input to a lossless (or lossy) compression system. The wavelet representation optimally decorrelates the data so that the input values are independent – there is no redundant encoding of data. In effect, the wavelet transform is acting as a high-efficiency predictor – but retains all of the geometric information of the image. Consequently, this representation provides an optimal representation for both image compression, and human visual perception.

Lossy compression methods

Lossy image compression uses various methods to approximate the original image in such a way as to improve compressibility. By lossy, it is meant that the input pixel values are changed upon inversion. As the vast majority of interesting images are not easily predictable, this approximation step usually comprises quantization of the images into a reduced dynamic range. For example, if an image were comprised of pixel values from 0 to 255 such as an 8-bit grayscale image, to improve compressibility, the data could be quantized by dividing each value by 2, yielding a new dynamic range of 0 to 127. These quantized values can then be inverted by multiplying by the quantization factor and the degradation in the image relative to the original is caused by the deviation of the approximated values from the original values.
In this case, all odd values would be approximated by the next lower even value – but the image has half the "information content".

After this quantization, or approximation step, similar lossless coding methods are used as in the lossless case above. These methods are generally called "entropy coding" methods as "information content" is a measure of the entropy of a system. The more ordered a system, the less information is present. If every value were predictable – there would be no information. If every value is random, then entropy is maximized and the information content is maximized.

This allows optimal, scalable use of the server resources. If any interaction has a computational load on the server, the server scalability is compromised. Many competitors have copied the iSyntax behavior at a single workstation – but fail when multiple users are querying the same server. The iSite system scales to over a thousand users on a single server.

- The server-client communication is in discrete requests. Streaming overloads resources. The context of a doctor's accesses can allow a predictable way to pre-fetch data – for example, the next image in a stack in the direction of query. With the use of localizers and scouts, the doctor can go directly to an important image or feature without incurring the overhead of waiting for the streaming load of the entire dataset. Workflow is optimized[1].

These capabilities combine to provide the technology to enable the incorporation of images into the ubiquitous information systems as just another piece of information.

Compression

Additional benefits, and the relationship of iSyntax to lossless and lossy compression methods are described in detail in the Intermezzo. In brief, there are two distinct classes of image compression methods: lossless and lossy. As the names imply, lossless compression methods retain all of the original information, while in lossy methods some of the information is lost.

Why lossy methods are sub-optimal for medical images

Lossy image compression approximates the original image by an image more amenable to prediction or entropy coding. Every lossy image compression method uses a target error metric such as mean-squared error or absolute maximum error as a measurement of how good a compression is. This is often indicated by the signal-to-noise ratio.

Quantization methods approximate similar values by a constant value – for example, in an area of an image with texture, the detail in the texture can be sacrificed without incurring a large "error". In an ordinary photograph, background objects such as trees or the sky can be heavily degraded without perceived loss of detail in the image. These are the methods used by every commercial image compression system, along with JPEG.
The difficulty here is in the application of these methods to medical images. In virtually all cases, the diagnostic information in a medical image is a subtle variation in low-contrast, high-frequency detail – i.e. local texture. A hard edge such as a fracture is easily detected in even highly compressed images. A subtle malignancy does not have a hard edge – but is rather identified by a variation in localized texture. The human eye is very good at detecting this type of local variation. However, all commercial compression systems preferentially discard this information first. These subtle details – such as a pneumothorax – are routinely missed in uncompressed images. The idea that compression can be "as good as" the original image neglects these considerations.

The common argument that diagnostically important areas are not compressed highly but "there is a lot of black space in medical images that can be compressed" is answered more effectively with the iSyntax approach: "just don't send it".

iSyntax image representation

The wavelet representation used in iSyntax is not compression – it is just a representation. In other words, what goes in is what comes out. What the wavelet representation brings to the table is an optimal way to represent the information needed to reconstruct any part of an image, at any resolution, with the minimum amount of data, while maximizing the data integrity relative to the original image. This is, in fact, a complicated way of saying that it is the best way to represent the data.

The lossless compression step is performed on the wavelet-transformed data. Both the iSyntax system and JPEG use the identical underlying representation.

Given that there is a theoretical limit to the maximum compression available for any given signal, how is it attained and what are the trade-offs?

There are several "degrees of freedom" in how compression can be applied to a signal:
1. maximum compression
2. rapid encoding
3. rapid decoding
4. scalable access of encoded data

All of these "degrees of freedom" have an associated "cost". For example, if one chooses to attain the highest degree of compression possible, then there is no flexibility in encoding or decoding times, nor scalable access.

▶ Commercial compression systems can discard critical diagnostic information.

▶ Wavelet representation retains all of the essential image data.

[1]Clearly, when video formats are used, streaming is the preferred delivery mechanism

An example of rapid encoding is a delta coder, where a signal is encoded by the differences in adjacent signals. This is very low computation, but has a low expected compression rate.

An intermediate compression method is a Huffman coder (variable length prefix coder). This requires pre-computing a set of code tables but achieves fairly high compression – at the cost of latency (computation) in encode and decode and the overhead of storing and communicating the tables to the decoder.

▶ **Images are integrated into the enterprise information system like any other data.**

JPEG and iSyntax have the same image representation – a wavelet transform. In fact, for the lossless case, the wavelet system chosen is the same. The methods differ in the choice of lossless compression used. JPEG uses a method that achieves close to the theoretical maximum compression – at the cost of extremely high computation and no scalable access to the data. What that means is that the entire image can be retrieved with the least amount of data – but no sub-part or sub-resolution of the image can be retrieved independently[2]. That is, the entire compressed image partition must be accessed to retrieve any part of the image.

The iSyntax method uses a lossless compression method that optimizes scalable access, at the expense of 10-15% coding efficiency. This translates into the ability to extract any region at full resolution or any sub-resolution of the image with no computational overhead. This allows for the remote access capabilities of the iSyntax system, and the maximum scalability of server and client computing resources.

▶ **Visual cues achieve an enormous gain in efficiency and system utilization.**

This choice was made to optimize the characteristics of the image storage with medical workflow. Data storage is becoming progressively more inexpensive, while clinical and diagnostic interpretation of images is becoming more expensive and more complex. The flexible access to data by doctors at any location is far more critical than maximum compression. The true cost component of the overall system is doctor bandwidth. By choosing a flexible representation, iSyntax allows the integration of images into the information systems of an enterprise as just another component of data. This has allowed an entire re-design of the user interfaces to incorporate visual cues for doctors, rather than textual descriptions to identify pertinent medical information. This is an enormous gain in efficiency and system utilization, as several major medical sites using iSite have found[3].

Optimization

There is no magic to iSyntax. There are extremely solid limits to how much a signal can be compressed and iSyntax gets close to these limits. It is unlikely that any future method will ever beat iSyntax in compression rate by any significant amount.

Following the theme of workflow or system-wide optimization, the use of iSyntax also enables other optimizations.

It is not data bandwidth but latency that is important. With the iSyntax technology enabling the integration of images into the information system, the user interface for dealing with clinical data was redesigned for use in the iSite system. The doctor no longer needs to read a description of the contents of a study, but is rather visually cued by the actual contents of the study. This is far more efficient.

With an intelligent design of the interface, the problem becomes one of how to “hide” latency? This is a complicated topic with many nuances, but a simple example will show why iSyntax with no lossy compression is more effective and faster than a system using lossy JPEG compressed images.

Assume a user would like to display an X-ray image that is approximately 2048 x 2536 (2 K x 2.5 K) on their computer screen in a small window approximately 500 x 300 pixels. Although, the original image size is approximately 10 MB, the data required for the “first” image to be displayed using iSyntax is around 40 KB, whereas the lossy compressed image (assuming 20 : 1 lossy compression) requires 500 KB for any display. The user then examines the available images and makes decisions as to what to do next. During this phase, the iSyntax system is downloading increasing resolution – “hiding” load time. By the time the user decides to increase the size of the window or zoom into the image in order to investigate some structures more closely or perform quantitative measurements, all data

[2]JPEG does provide the ability to encode a pre-computed region of interest for preferential extraction. However, this is a meaningless case in medical imaging where the pertinent area of an image cannot be known beforehand.

[3]MD Anderson in Houston, TX found that the utilization of their physician portal went from 5-6% to over 90% after the inclusion of the iSyntax imaging capabilities.

will have been downloaded in the background. iSyntax gives the appearance of immediate and instantaneous image load time with little sensitivity to data size or network limitation. In contrast, the lossy system is less responsive, has higher user latencies and introduces image degradation.

Conclusion

By concentrating on optimizing the workflow solution of a combined image and information system, using iSyntax as a enabling technology, Stentor completely re-solved the PACS problem. Combined with a redesign of the data interaction model of the user interface to better support rapid interpretation, Stentor was able to integrate images into iSite: a ubiquitous information system which enabled the first practical unified electronic patient records to be developed.

It must be stressed that this is only one component of the iSite success. In combination, a unique service delivery and business model was developed which removed the financial barrier of entry to hospitals and clinics to have this kind of enterprise support. This aspect of the iSite strategy will be dealt with in another article ■

▶ **iSite is a ubiquitous system optimizing the workflow of images and information.**