Global Information Industry Center
at the School of International Relations and Pacific Studies
University of California, San Diego

A Senior IT Executive Perspective on Data Growth in Hospitals

Senior Lecturer Emeritus Jack Rockart, CISR, MIT Sloan School

September 2009

Abstract:
This paper reports on the results of interviews with eight senior IT hospital executives and two medical insurance organization CIOs on the subject of data management in the hospital industry. The executives were asked to talk about their current amount of data storage, its growth in the past several years and their expectations for data storage growth in the future. They also discussed the value they saw in current and future data and the issues they face in the management of this data. All saw very significant growth in data in the next several years caused by six major factors. These are the growth in the volume as well as the technology of imaging, expanded use of electronic medical records, increased use of "Web 2.0" software, growth in "decision support" capabilities, the development of new health information networks and the advent of genetic testing. The executives saw significant value in today's and future data both for patients and the healthcare system. Finally, they noted many issues, both current and future, in the management of data.

For additional information please contact Jack Rockart at jrockart@MIT.edu.

☐ **Executive Briefing:** a summary of one or more research projects with preliminary findings for a non-academic audience.

☑ **Research Report:** a completed report drawing on one of more research projects that presents study data, findings and management implications.

☐ **Case Study:** an in-depth description of a firm's approach to an information management issue.

☐ **Research Article:** an academic research paper with sections on hypotheses tested, methods and data, analysis, findings and references.

Contents

A Senior IT Executive Perspective on Data Growth in Hospitals

Senior Lecturer Emeritus Jack Rockart,
CISR, MIT Sloan School

September 2009

Executive Summary

The convergence of a number of factors strongly suggests that data and information storage will grow rapidly in hospitals in the United States in the next several years. Among the most important of these factors are (1) federal and business interest in promoting the use of electronic medical records and (2) major technology advances, the most important of which are in images and genomics. This expected large growth in both the kind and amount of data will, however, provide significant additional data management challenges to IT departments.

"The requirements for data storage will increase exponentially as we move deeper into the structural biology of a patient. We are not talking about the distant future but about the next five years," said a former hospital CIO, now a major consultant in the field. His comment came as part of a set of interviews of nine leading IT executives in the medical field.

The interviews were part of a set of interviews of 30 IT executives for the "How Much Information?" project[1]. The project is supported by several technology producers. While the above quote is the most striking of those made by the interviewees, its direction is shared by all. Data and information being stored by hospitals will increase significantly in the next few years.

The nine executives we talked with included three hospital CIOs; two former hospital CIOs; two IT Directors with major responsibilities for data who reported directly to the CIO; and two IT executives with medical insurance companies. All were selected because of their leading edge knowledge in the field of healthcare IT. All of the hospital participants were, or had been, involved with major research-oriented medical centers. They were selected, not as a representative sample but as executives from whom we could best learn the future of data growth in hospitals. The interviewees' hospitals, or hospital systems, range in size from 350 beds to more than 1000. Revenues range from $1 billion a year to more than $7 billion. The insurance companies are large and are leaders in the field of health insurance.

Past and Future Growth in Storage

The table below shows past growth rates, current storage numbers and expected future growth in a five-year horizon cited by the seven current or former hospital executives. As can be seen, some data elements (noted by "NA") were not provided by all. Actual numbers, rather than estimates, for the past growth were provided by most of the interviewees.

1 For more information on the HMI? project, please go to: http://hmi.ucsd.edu/howmuchinfo.php. For a copy of the project report, "Capturing Value from the Information Explosion," please contact Cynthia Beath at cbeath@mail.utexas.edu or Pepper Lane at pelane@ucsd.edu.

Table 1. Storage Growth Rates
(See Appendix: How Many Bytes is That?)

Interviewee	Past Growth Rate	Current Storage Size	Future Growth Expectation
A	30%[1]	High Terabytes	A hundred times in five years.
B	18-20%	1 Petabyte	Will double every year.
C	60%[2]	4.5 Petabytes	Will reach zettabytes in 5 years
D	40-50%	900 Terabytes	Uncertain, but some large clinical systems growing 30-40% per year
E	30%	1.8 Petabytes	N/A
F[3]	60%	about 1 Petabyte	Probable zettabytes in five years
G[4]	N/A	N/A	Probable zettabytes in five years

[1] This is the smallest of the hospitals, behind the others in medical record automation and performing less research.

[2] This percentage is taken from a graph of growth in their production data in the SAN.

[3 &4] F and G are the two former CIOs who responded based on their knowledge of the field in general.

To summarize the above, past growth rates were believed to be in the 18-30 percent range at the three institutions with the lowest growth rates. Past growth at the other three, all larger research institutions, ranged from 40 to 60 percent. The future growth estimates, in particular the estimates of moving to zettabytes of storage in the next five years from three of the respondents, were startling. It should be noted that the estimates were given during a one-hour interview and indicated to the interviewer that, in all but one case, the interviewees were expressing a strong belief in very significant future growth (rather than sharing the results of a formal analysis of future growth). Two of the "future zettabyte" estimates, it should be noted, came from CIOs[2] who are thought-leaders in the field. The third, however, from the CIO of a major hospital system, is in fact based on calculations made by the research department of the system, where the use of genomic data is expected to fuel this growth.

2 In this paper we refer to all current and former CIOs as "CIOs."

What is Causing Data Storage Growth?

Why these high growth rates? With regard to past growth, the IT leaders named multiple factors. Included in these were email, voice, pay for performance requirements, changes in data structures, data replication, and image technologies. The most important of these was image – which is expected to continue to stimulate data growth in the future. Five additional factors, however, were also seen to be the major reasons for future growth. They were the need for better decision support for physicians, new web 2.0 applications, implementation of electronic health records, new medical networks and innovations in genetic testing. Next we discuss the six drivers of future growth.

(1) Image Technology

Almost every interviewee cited image technology as a major contributor to both past and future growth rates. As one CIO put it, "You have to recognize that every "ology" – radiology, cardiology, pulmonology, OBGYN, etc. – all have image

generating systems. Most of those images are not under our control. They are out in the labs. And, the storage they need will continue to grow, both locally and centrally." There are many reasons for this growth. The key reasons appear to be increasing digitization of films, thinner and denser slices of organs, "backloading" of current analog films, and the need to keep all data for research and discovery reasons.

"The average size of images today is very different because of the devices that are being used to capture these images. They are so fine in the number of slices that they take and so dense in their resolution of each image of any organ that the whole area of imaging just is phenomenal in growth. You are looking at images in which a single slice can be almost a gigabyte in size."

"On the digital pathology side, we are going to see tremendous growth in part from backloading existing analog material. The bone marrow department has 5,000 glass slides that they want to digitize. They want to get rid of the cabinets and get access to the data. Each slide is anywhere from 15-25 gigs of data." (Five thousand slides times an average of 20 gigabytes per slide is 100 terabytes of data.)

"We need to keep all of our research data not only because of future research and discovery reasons. We even need to keep the data from researchers who have left us. In the event that they find a new wonder drug, we have to prove whether it was done here or somewhere else."

Image date will continue to grow – sometimes by leaps and bounds. Interestingly, the data noted in Table 1 does not, in most cases, include image storage that is kept out in the "ologies". There is, however, now a desire in some organizations to bring this data into well-managed "central" data storage (see "Data Management," below).

(2) Clinical Decision Support for Physicians

While image was the major driver of past growth in the mind of the respondents, two other factors were seen to be of some importance in the past but of growing importance in the future. They are "decision support," or instant access to medical records and other data to improve medical decision making by physicians, and new web 2.0 applications, which will create and draw upon major new data bases of "unstructured" data.

As physicians learn more about the technologies and data that are available in their institutions, the demand for decision support grows rapidly. As one CIO puts it, "The docs today expect decision support. They expect automated medical reconciliation, e-prescribing and getting to an image at the touch of a button. And they expect this anywhere in the world. The demand for having this information accessible via the web anywhere, any time is incredibly high in my organization."

Another CIO, well along with enabling physician access to data explained what his organization was doing in this way: "We have thousands and thousands of rules today which look at the data and can help a physician. The major driver of the use of data and rules is that our docs, particularly the primary care physicians, have no time. As people get chronically ill and live longer, there's a lot more going on. And physicians only have a very short time that they can spend with a patient. So we have to get smarter about saying with our decision support, 'Doc, I can think of the two most important things to pay attention to'. We need to be able to help the physician with regard to all of the things that are known about the patient. So we need to put much more focus on highlighting the data that matters and helping physicians to make the best possible decisions."

If this is a key part of the future of medical care, and our respondents believe it is, it will require data-rich electronic medical records (see "The Electronic Medical Record" below) as well as effective rules. It will also drive the need for more primary storage as well as software that can quickly retrieve data from tertiary storage.

(3) Web 2.0 Applications

The capabilities commonly known as "web 2.0" applications are beginning to play a part in medical care. Using this fairly new, web-based software that engenders collaboration, some hospitals are opening up new ways to improve patient care and learning throughout the organization. While current use is in its early stages, there are many major expected uses for this new technology.

One early use of the concept that has been implemented in one hospital is for patient medication list management. The CIO notes, "Every member of the care team has the ability to adjust the medication list, delete, or annotate it. It is not just one doctor maintaining an up-to-date medication list, it is a team. It's wiki-like, where there's a journal of every change that was made and who did it. Obviously, a clerk cannot prescribe a new medication but they can annotate old ones. It is truly a social network of team members working in real time on that patient's record. This has been used for the past year."

He also notes, "We have also installed software which promotes collaboration among researchers. It has been used about 12,000 times a day since we put it up six months ago."

In another hospital, residents use wikis to record what they are learning and other residents correct entries, where appropriate. The same hospital has set up what they term "e-rooms" to allow caregivers the capability to work together to develop guidelines for treating a disease.

These sample applications are first steps in using collaborative capabilities. What comes across from the CIOs, however, is the clear sense that this technology has the possibility to allow significant improvements in medical care. As one CIO put it, "Of real interest to us are sites where patients can come together to discuss diseases like multiple sclerosis or dementia. We need to figure out what to do with public sites and whether to initiate them on our own. We believe that we could pick up early signs of adverse drug reactions as well as what is working well. There is a chance that the community could turn on you, which would certainly be difficult. But we suspect that these patient 'reports' could be much more useful than the reports put out by the state."

A recent issue of the Economist (April 18, 2009, p. 18, "Medicine goes digital") speaks to this also. It notes that "…research has shown that the most active communities on social networks such as MySpace concentrate not on celebrity gossip or sport but on chronic illnesses – especially stigmatized conditions such as depression." The article goes on to point out that increasing patient involvement in their own care, made possible by collaborative capabilities, may even change the process of medical care.

Just how much, and in what ways, this newer technology will affect medical care is certainly unclear at this time. And what this means for data storage is uncertain. What is clear is that is that these applications create large streams of unstructured data, all of which is typically maintained in high-availability primary storage. In addition, storage is needed for the lengthy audit trail of changes that are made in the data. Our respondents in leading hospitals are working hard to understand exactly what impact web 2.0 applications will have. They do expect them to play an important role in the future.

(4) The Electronic Medical Record

The development and use of electronic medical records[3] (EMRs) is one of two areas in which storage needs are expected to grow dramatically over the next few years (the other is genomics, discussed below). Since the number of physician offices and hospitals where EMRs are in effective use is currently extremely small, there is the potential for a very significant increase in data storage with more widespread adoption. For this reason and because the pros and cons of EMRs are widely debated today, we spend considerable time on this growth factor.

3 A patient's medical record today is called by many na mes, such as the "electronic health record" and others. The distinctions are meaningful but, for simplicity, the term electronic medical record (EMR) will be used here.

While a medical record for everyone is probably an unobtainable goal in the next five years, it is not unreasonable to expect significant progress in moving from paper to digital records in the next few years. There are two major external drivers for the widespread development of electronic medical records today. They are the $19.2 billion federal stimulus package and increased attention to rising medical costs by major businesses and government. The stimulus provides $44,000 for every clinician in America if they have "meaningful use" of electronic medical records by 2011. If implementation is in 2012 or later, the amount is reduced. Hospitals also stand to receive significant funds from the stimulus package. One hospital in our survey estimates that they will receive about $7 million. There is a stated US goal to have 90% of US citizens with EMRs by 2014.

While not as potent a factor as the federal stimulus, there is significant stirring in the business community, toward the concept of the "medical home." There are strong beliefs in many large companies, such as IBM, that the healthcare system is too expensive and that a more patient-centered approach is necessary. A patient's primary physician will be the lynchpin of the "home," providing connections to all other caregivers. The hallmarks of a medical home, according to Dr. Paul Grundy, director of health care technology and strategic initiatives at IBM, are an ongoing relationship with a doctor; a team approach to delivering comprehensive, coordinated care that is integrated across the health care system; the use of tools such as electronic medical records, to ensure that care is delivered safely; and expanded access by patients to their physicians - including evening and weekend office hours and the use of email and telephone consultations.[4]

As one of our CIOs noted, it may take ten years for this to happen but it is the future. With the medical home, he says "You are focusing on prevention. You are focusing on helping patients be accountable for their health and thereby improving the quality of the patient's care. You are also reducing the costs of health care." The glue connecting physicians to the patient, and to other physicians, will be the EMR, driving the need for increased data storage in physician offices and hospitals. One other former CIO was also adamant in his belief that this approach had to be the future. And there are companies, like IBM, that are providing incentives to physicians to begin providing a medical home.

These two external drivers are complemented by two "internal" drivers for both hospitals and physicians. One is the desire to use EMR technology in order to be more efficient. As one CIO noted, "Our younger physicians are ready for this. They put in the time to learn it and like it." He noted that of the house staff in his hospital, 100 per cent were eager to use EMR and other technology as it becomes available. This is not surprising. These physicians are mostly young and members of "generation Y," a generation sometimes called "digital natives." They are now 25 percent of the workforce and will be 45 percent in five years. This suggests that the rate of adoption of technologies like electronic medical records will quickly increase. A second internal driver is the desire to provide better patient care. As we note below, our CIOs all agreed that an EMR enabled improved care of the patient. One noted that even some of the physicians who were not going to adapt the use of the EMR in the near future saw this benefit.

How widely are electronic medical records in use today? The most recent study, published in the New England Journal of Medicine[5] ("Use of Electronic Records in U.S. Hospitals, A.K. Jha and others, 360:16, pp 1628-1638, April 16, 2009) and co-

4 See 'Medical home' concept embraced by IBM, other employers," at financialweek.com, March 12, 2008, at http://www.financialweek.com/apps/pbcs.dll/article?AID=/20080312/REG/446511355/1014/NEWS (accessed March 22, 2009).

5 A.K. Jha, C.M. DesRoches, E.G. Campbell, and others, "Use of Electronic Records in U.S. Hospitals," New England Journal of Medicine, v360:n16, pp. 1628-1638, April 16, 2009. The article was viewed online at http://content.nejm.org/cgi/content/full/360/16/1628 on Aug 15, 2009.

authored by Dr. David Blumenthal, who was recently named to lead the Obama push to increase the use of information technology in healthcare, noted very little use. Utilizing 24 factors in areas such as clinical documentation, test and imaging results, computerized provider order entry, and decision support related to the medical record, the study found only 1.5% of 2952 hospitals surveyed had a "comprehensive" electronic records system that included these factors. Some 7.6% had a "basic" system, usually without computerized order entry and decision support, present in at least one unit of the hospital. These findings are in line with other recent studies.

A survey by the same group in 2008 found just 17% of US physicians using electronic medical records.[6] Only 4 percent had fully functional systems including order entry. As one of our CIOs noted, that leaves some 664,000 out of 800,000 physicians in the US without some type of electronic medical record and even more without a fully functional one. There is clearly a great deal of room for an increase in EMR use – and for the concomitant growth in data storage. With most medical records in paper form today, the movement of this data to electronic storage will greatly expand the amount of storage needed.

For the leadership hospitals, having a comprehensive medical record is nothing new. As one CIO said, "I have a thousand physicians that I am directly responsible for. They all have electronic health records and have had them for years." The issue for these hospitals is system improvement and ensuring effective use. As one CIO says, "We are relentlessly forever knocking out paper and replacing it with computer functions. We are at the point now where, if you are a physician and not using the electronic medical record by the end of next year, you're out of here."

The CIO from the smallest hospital, with a medical record system installed in only one department, is clear that his organization needs to and will catch up. He talked at length about how the organization realizes the importance of the EMR for effective patient care. For this hospital, the stimulus package is a major driver leading toward implementation.

Value. Why should the government and medical organizations spend the billions of dollars necessary to implement medical records? The answers that emerged from all of our respondents were cost (e.g., cutting down on unnecessary duplication of tests), patient safety (e.g., catching medications that might be harmful to a patient before he takes it) and quality (e.g., the ability to follow appropriate guidelines for each patient's situation). While these responses were universal, four other interesting value opportunities were mentioned: (1) reducing the capacity gap, (2) reducing length of stay (LOS) of patients in hospitals, (3) eliminating a significant portion of defensive medicine, and (4) enabling preventive care.

One former CIO focused on what he termed the "capacity gap." "I think our ability to be able to supply care at a reasonable price as the baby boomers age is in doubt. We will not have enough medical personnel. There's going to be a gap between supply and demand and that is going to cause large, large challenges for the industry. It already is happening in some regions." The answer to this, he says, is giving physicians access to all information about the patient in a timely manner so that the physician can be more productive. Effective electronic medical records will make this possible, he believes.

The same CIO also notes that reducing length of stay can save countless dollars. "If we can reduce LOS for patients, we can avoid building beds. And adding beds is not just the cost of the beds, it is the cost of the people that you need to staff those beds as well as the cost of the buildings and equipment to put them in."

With regard to defensive medicine, another former CIO states, "I think defensive medicine is one of the most expensive components of the healthcare delivery process today. A lot of it is practiced because neither the patients nor the physicians have

6 C.M. DesRoches, E.G., Campbell, S.R. Rao, and others, "Electronic health records in ambulatory care -- a national survey of physicians." New England Journal of Medicine, 2008, v359, pp. 50-60.

access to the full historical content of the patient's care. If a physician needs to order tests, and the tests have already been given yet their results are not available, any physician will go ahead and order all the tests that he believes he needs. This may expose the patient to additional radiation or other issues. And it certainly provides additional cost to the healthcare system."

Both of the CIOs noted just above believe that the "medical home" – based around a medical record – can facilitate a long hoped for goal of medicine: preventive care.

What If Everyone Had an EMR? What would happen to data storage if all 300 million people in the United States had electronic medical records? By far the major part of an EMR is the storage of images. The rest of the EMR is relatively insignificant in terms of storage. Here is a computation of this "what if" based heavily on an analysis provided by the head of medical imaging at Partners Healthcare in Boston.

Table 2. Storing EMRs for Everyone in the US (See Appendix: How Many Bytes is That?)

Assumptions & Facts	Computation
• Average image exam size = 100MB	• Average imaging exams per person per year = 750M/300M = 2.5
• Number of imaging exams performed in the US = 750M per year	• Image data per person per year = 100MB x 2.5 = 250MB
• Population of US = 300M	• Lifetime patient image data = 250MB x 40 years = 10GB per person
• Life expectancy in the US = nearly 80 years	• 10GB x 300 million people = 3 exabytes of storage
• Median age of in the US = about 40 years	

Of course these figures are based on an outer limit of the number of patients whose data might be stored. But, one can plug in whatever reach percent of the population one expects to have EMRs and draw one's own conclusions.

But Uptake of EMRs Faces Many Issues. There is a lively discussion in the press today about the ability to effectively implement EMRs. Those on the negative side zero in heavily on cost with emphasis not only on initial hardware and software implementation, but also on training and on-going maintenance and support costs. One CIO in our study estimates these costs at $40-6O, 000 per physician. (This could be lower – Wal-Mart's Sams Club is offering an EMR kit priced at around $25,000 for the first physician and $10,000 for each additional one for implementation of hardware and software. As of this writing, it has yet to be tested in medical organizations, and the costs of training and operation are not included.) Other, far from minor, roadblocks issues include security, privacy, a lack of standards if data is to be shared, and significant usability issues.

The primary deterrent, however, is the reaction of physicians to the technology. It is here where several of our interviewees cited problems. One summarized much of what others said as he spoke at length about the issue. He noted the well-known factor of resistance to change, and the fact that each specialty has a distinct need for different information, which demands that any "standard" EMR be somewhat tailored to specialists needs. More important, he said, was the fact that many, if not most, current software offerings in the market are not a good fit with perceived physician needs. Yet, as he puts it, "the

primary issue for a time-stressed physician is the ability to make the technology work for him rather than against him. It is a lot quicker for a doctor to pick up a piece of paper on a bed board or on a clip and look at that data than to go to a system and log in for one particular piece of data."

Sounding the same note, another CIO says that there is a real need for an EMR and related software applications that will be a "more complete and easily learnable solution." "You can't ask physicians to leave every day a half an hour later or to do an extra hour of work a day because the system forces them to do the data entry versus someone else. Maybe ten minutes or a half hour extra a day for two months until they get up the learning curve is acceptable but, if that loss continues then we will have to look harder at how we reimburse physicians, especially primary care docs."

On the one hand, we have today several effective electronic medical records in a few hospitals. Most were built by the organization itself with strong leadership and implemented using well-designed change management processes. On the other hand, those without EMRs see some clear challenges.

(5) Health Networks

There is increasing activity with regard to health networks throughout the country. Health networks are community initiatives providing the technology to link physicians, hospitals, community health centers, laboratories, patients and other health organizations electronically. While their primary purpose is the exchange of information, some provide patient based EMRs – yet another approach to the design and building of EMR capability.

The eHealth Initiative's website reports that their 2008 survey showed 42 Health Information Networks (HIEs) in 2008, an increase of ten over 2007. Many new HIEs are under development. While these differ from site to site, and usually involve multiple medical organizations, they tend to collect information from multiple medical sites, process it and provide data to physicians from a patient's record.

The hospitals in our survey have either put in place or have set about building their own health networks, to provide patient medical records not only at the hospital but also at the sites of their affiliated physicians. Information about patients is transferred when appropriate.

Two recent factors are driving this move. One is the relaxation of the Stark Law which now allows hospitals to pay up to 85% of the cost of the software and the necessary equipment at the local physician's site. The second, as noted earlier, is the $19.2 billion of federal stimulus funding. Three CIOs whom we interviewed talked at length about this.

One hospital, the smallest in the group, has just started building its health network. It will have its own EMR, assist affiliated physicians in building their EMRs and provide data exchange capability. The aim is to provide access to updated status of the patient's data everywhere within the system. The benefits of an in-house patient health record, discussed in the prior section, will thus be extended to the entire network. All patients, not only those who are in-patients at the hospital, will have their data in an EMR and, moreover, have the benefit of updated information available to each provider.

In a second hospital, the CIO notes that, "We believe that all the docs in the community need to have electronic health records." He is working to provide this capability to "anyone who is a member of our physician organization – although they may be members of multiple organizations. We are 75% complete. To meet federal stimulus payment dates, we are requiring meaningful use by our clinicians as a condition of continued membership in our physician organization." A third hospital, much further along, is engaged in bringing medical record interaction capabilities and, equally important, a series of clinical systems to all its providers.

(6) And Now, Genetic Testing

According to our interviewees, genetic testing will soon be routine for high-risk patients. As one CIO puts it, "We believe that we are within three to five years of where genetic testing is just part of a patient's testing. We know, for example, that there are 76 different forms of leukemia each of which has a different genetic underpinning to it. With a genetic profile, we can pick the right drug for the particular form of leukemia. There are now 2500 diseases for which there is a genetic test."

"With the advent of genetic testing, I actually think your grandkids and mine will have care that's very different from what you see today. They will live long and cancer won't kill."

Another CIO, who has had the 30 billion base pairs of his genome sequenced, is also a strong believer in the future of genomic testing. "I have five or six conditions that one might be concerned about. I would have had a completely different experience with the healthcare system if, at the age of 20, not at the age of near 50, I actually understood all my risks. And so if it costs $1000 when you are 18 years old to have a map of many of the disease states that you could mitigate through lifestyle changes, I think it is highly likely that patients and doctors will take advantage of that."

The same CIO goes on to note that the US is preparing for this revolution. "Last year we developed a set of standards for the recording and transmission of the genome and family history. We are technologically ready and policy ready."

Both CIOs note that this revolution raises many issues. Among the most significant are problematic ethical problems, cost issues, training of personnel to effectively deal with this new technology and legal questions about retention.

Three of our respondents believe that the genomic testing, together with other medical data, will increase the amount of data storage in a teaching hospital to the zettabyte range in the next five years. Another CIO is less certain about the growth trajectory. "Will I store the 30 billion base pairs? Will I store just the important variations, or the results of the analysis? If it is the former, for clinical or research purposes, it will be a huge amount of data. "

What is clear is that the revolution has started. Massachusetts General Hospital recently announced[7] that it plans within a year to read the genetic fingerprints of nearly all patient tumors. They have already performed genetic screening on patients with, at least in one case described, very favorable results. Memorial Sloan-Kettering Cancer Center is now ready to have most patients with lung cancer screened. "Eventually, said Dr. Marc Ladanyi, chief of Sloan-Kettering's Molecular Diagnostics Service, such screening during a visit to the oncologist will be as commonplace as tests performed during an annual physical. 'You can think of it a little bit like when you get your blood drawn, he said."[8]

Data Management – A Tough Problem

Just how to manage all the data that will be available in the future is unclear to our CIOs – and worrisome. Many are concerned about how to manage the data for which they are currently responsible. The current amount of storage noted in Table 1 represents only a part of the data actually managed by IT in medical centers. To this centrally-managed clinical and administrative data must be added the image data that is now in "local" image systems. And there may be also a very large amount of research data – not typically part of the centrally managed data. Add to this the genomic data on the very near horizon. We discuss the management of images and research data first and then look at some specific issues our CIOs noted with regard to data storage.

Images. Image data, as noted earlier, is typically stored all over the institution. "Healthcare is the land of 10,000 data bases throughout the organization," says one CIO. One example, from another CIO, is "the guy who runs the echo machine who is capturing his echo data onto, let's say, a digital tape and then keeps a cabinet of 100 digital tapes." This CIO says there is an amount of image data in various places in the organization equal to the current amount of data he is now handling. His plans for the coming year are to establish an "enterprise image archive" to ensure that the data has all the hallmarks of well-managed data (secure, protected, available for disaster-recovery, etc.)

Research data. Research data is typically not managed by the central IT organization at the current time. With the National Institutes of Health being better funded by the Obama administration, there is every reason to believe that research, and the data it develops, will continue to grow in institutions with research missions. In the words of one CIO, "single researchers often have a data base in the terabyte range with some moving toward petabytes."

7 Stephen Smith, "MGH to use genetics to personalize cancer care," Boston Globe, March 3, 2009, page A1. See http://www.boston.com/news/local/massachusetts/articles/2009/03/03/mgh_to_use_genetics_to_personalize_cancer_care/

8 Stephen Smith, op. cit.

Research data, therefore, "swamps other data." Another CIO provided an example of one researcher with a multi-petabyte data base. Moreover, each of these data bases is an individual "silo." Each is supported by grants from NIH and other sources. They are unlikely to be brought together in the near future – if at any time. It is not clear how well these data bases are managed.

Some key issues in data management today. All in all, as one CIO says, data management is "extremely concerning." It is beginning to get the attention of senior management of some hospitals. In one institution, the head of the hospital drew attention to this issue. "He really does recognize the fact that we have huge amounts of data. And if someone asks 'what are you doing with all of this stuff,' there really isn't a good answer. No one has a great answer for it," says one interviewee.

One CIO sees the data problem as involving a number of specific issues. Among these are the difficulty of backup, replication of data, how long to keep data, vendor dictates and whether to go to an internal "cloud".

Difficulty of Backup. It is the increasing size of data bases, the CIO notes, which make them difficult or impossible to backup. As data increases to multiple tens of terabytes in one application, backup "just isn't going to happen." The solution is replication – which adds to data storage needs.

Growth of Replication. But replication is also a problem… and it is happening by default in many areas. As the CIO noted, "They come in and take your blood. They do the blood workup. That information is stored in the blood bank system. The results are then sent back to the order management system. At the same time, that information is sent to the medical records system. And then a piece of that information, with comments from the floor, also goes back to medical records. It actually resides in four places before we are through." "Do we need to have it in four places?" he asks.

Which data should be kept for how long? Similarly, he asks, "Do we really need to know what my blood pressure was 10-20 years ago?" The question is not his alone. Data retention is a particular challenge for a medical institution, since the clinical value of some items of data vary significantly from one department to another. Also, as noted above, litigation and research are also reasons why data is maintained – especially in academic medical centers.

Vendor dictates. In an era in which software is purchased, not built, vendors have an influence which strongly affects how data is stored and accessed. "They dictate to us the server hardware and how much storage needs to be assigned to it. When we try to retrofit things to provide high availability, they put up a red flag and say, 'we do not support that'. It is a problem."

Managing data in the "cloud." This CIO, like others, is now faced with the issue of whether to store data in "the cloud," whether it is by means of an external supplier or internal virtual servers. The new technology poses problems of many sorts, not the least of which is internal resistance to putting "my data" in virtual storage. "They want their own server."

The CIOs we interviewed were deeply involved in one or another aspect, sometimes several, of this storage problem. Many were working with outside vendors. One CIO puts it this way: "If I want to store hundreds of terabytes of research and medical records data, what do I do? I have been working with several vendors to think about this. Given that I have a finite budget, how do I acquire SAN, NAS, Fibre Channel and various types of RAID, not RAID-erased systems, because I have to judge reliability, security, and cost and then balance these with many other factors. Storage, in general, is a nasty, nasty problem."

Final Thoughts

What are these CIOs telling us? It is important to restate that they are not describing data storage problems at small community hospitals. Most are executives who run – or did run – IT organizations in major medical centers. They were selected for their knowledge of the field. What they are saying, in summary, is that the amount of data storage is going to increase substantially, even exponentially, in the next several years. They believe that the federal incentives for adoption of electronic medical records, advances in imaging capability (particularly the capture of thinner and denser image slices), and the coming of the $1000 genetic test are all leading to significantly improved medical care and a concomitant increase in stored data.

It is clear that these advances will hugely affect large research-oriented medical centers. But what do they mean for smaller hospitals that make up the bulk of the 8,000 hospitals in the US and, more broadly, for physicians and their practices? The optimistic view, of course, is that the federal stimulus, with both its positive and negative incentives, paired with constantly improving EMR software, will lead to widespread adoption of electronic medical records – and that the "digital natives" of generation Y will further this move. The primary negative position is that inadequate technology and significant physician resistance will keep adoption below what is hoped for. While the impact of ERMs will only become evident in the next several years, it is the opinion of our CIOs that the value for patients and the entire health system far outweighs the negatives.

And, how about that $1000 gene test? Will the technology underlying gene testing become still better and cheaper and patient outcomes so clearly improved that gene testing, and the increasing data storage coming from it, will result in more widespread use? Will the cost of data storage decline enough to enable exabyte or zettabyte storage in hospitals? The answers are certainly not clear at present. But, the potential for improved patient care makes this an area to be closely scrutinized in the near future.

For IT, the management of data growth will certainly pose problems. These problems will certainly add to the many problems evident today in managing existing data storage growth. And data management costs will grow – not significantly in storage hardware where costs are declining rapidly but in software and people, whose costs are rising. But, once again, value, if the dollars are available for the necessary investments, should trump costs in the hospital sector. There are significant inefficiencies in today's processes which will yield to better outcomes in both costs and the quality of care for patients.

All in all, assuming the economy and the government cooperate, the new technology now available to medical organizations is going to make a difference in patient care. Our CIOs agree on that.

Appendix: How many bytes is that?

1000 bytes = 1 thousand bytes = 1 kilobyte (1k)

1000^2 bytes = 1 million bytes = 1 megabyte (1M)

1000^3 bytes = 1 billion bytes = 1 gigabyte (1G)

1000^4 bytes = 1 trillion bytes = 1 terabyte (1T)

1000^5 bytes = 1 quadrillion bytes = 1 petabyte (1P)

1000^6 bytes = 1 quintillion bytes = 1 exabyte (1E)

1000^7 bytes = 1 sextillion bytes = 1 zettabyte (1Z)

About the HMI? Program

The How Much Information? (HMI?) research program was formed in 2008 to answer the question, how much information is generated annually in the world, and in the United States? The program is sponsored by AT&T, Cisco, IBM, Intel, LSI, Oracle, and Seagate. The Principal Investigator is Prof. Roger Bohn and the Research Director is Dr. James Short, at UC San Diego's Global Information Industry Center (http://giic.ucsd.edu). Founded in 1960, the University of California, San Diego is one of the nation's most accomplished research universities, widely acknowledged for its local impact, national influence and global reach.

Acknowledgements

This research report is the product of industry and university collaboration in applied research. We are grateful for the support of our industry partners, sponsor liaisons, university research partners, and administrative staff at the University of California, San Diego.

The author would like to express his thanks to the following people: Eric Colliflower, Manager, Storage Management Group, Johns Hopkins University Medical Center; John Glaser, CIO, Partners Healthcare System Inc.; Leo Gilmore, former CIO, Lancaster General Hospital; Alok Gupta, CIO, Carefirst, Inc.; John Halamka, CIO, Beth Israel Deaconess Medical Center and CIO, Harvard Medical School; Don Lyons, Former Deputy CIO, MD Anderson Cancer Center and founder of InterOPERANT; Bob Romero, Senior Director IT, Johns Hopkins University Medical Center and Bill Schikolovich, CIO, Tufts Medical Center. Two other interviewees, for business reasons, asked that their names not be included.

Questions about this research may be addressed to the HMI? research program at the Global Information Industry Center at the School of International Relations and Pacific Studies, UC San Diego:

Roger Bohn, Principal Investigator, rbohn@ucsd.edu

Jim Short, Research Director, jshort@ucsd.edu

Pepper Lane, Program Coordinator, pelane@ucsd.edu