feature article

創業100周年記念特集シリーズ
ヘルスケアシステム・ソリューション

Healthcare's Data Tsunami

Why Healthcare Organizations are Drowning in Data and Why They Need Even More If They are Going to Survive the Storm

William A. Burns

OVERVIEW: Today's modern healthcare organizations from Hospitals to Life Sciences Companies to Payers are drowning in data. Electronic medical record information, medical images, scanned paper reports, dictated voice files, and full motion video are just a small sampling of the massive amounts of data collected in the process of patient care, bio-medical research, and medical claims administration.

Equally affected is the Chief Information Officer who is faced with an onslaught of data the likes of which they have never seen. Unlike the complex but tractable problem of managing structured databases from the massive to the mundane, the new menace is the management of the individual files themselves. Ranging from file servers to SharePoint* sites to the thousands of medical applications that populate today's healthcare landscape, this new threat is simply called "unstructured data."

This article will help you understand both how and why this problem exists and why injecting even more information into this paradigm of unstructured data assets is the only way to escape from drowning in the very data we are creating. Restoring long-term value to this most critical class of data and delivering a cost-efficient infrastructure to manage today's healthcare information is a key way of reigning in healthcare costs and maximizing the value of medical information.

UNDERSTANDING THE EXPLOSION

There is a data explosion happening in healthcare organizations around the world. Far from a "hockey stick" graph that will hit us in to 5 - 15 years, this is the reality today. Data growth is happening with such speed and ferocity that looking at the numbers alone will leave organizations wondering how to survive this data explosion, let alone make sense of this valuable information. Consider the many sources of data. Current medical technology makes it possible to scan a single organ in one second and perform a complete full-body scan in roughly 60 seconds. The result is nearly 10 Gbyte of raw image data delivered to a hospital's PACS (picture archive and communications system). Clinical areas in their digital infancy such as pathology, proteomics, and genomics which are the key to personalized medicine can generate over 2 Tbyte of data per patient. Add to that the research and development of advanced medical compounds and devices, which generate terabytes over their lengthy development, testing and approval process. And finally, consider the impact of electronic medical records which are already mandatory in many European and Asian countries and will soon be required for every patient in the United States. These sources of data are just the tip of the iceberg when compared to the thousands of medical applications in use today that create individual's files or "unstructured data" during the patient care process and store this data on countless computers, servers and storage arrays (see **Fig. 1**). It is impossible to ignore the impact that the influx of baby boomers will have on our global healthcare systems. It is widely accepted that chronic diseases make up the majority of healthcare costs and 80% of an individual's healthcare is consumed in the last 20 years of their life. As baby boomers worldwide enter healthcare systems,

* SharePoint is a registered trademark of Microsoft Corporation in the U.S. and/or other countries.

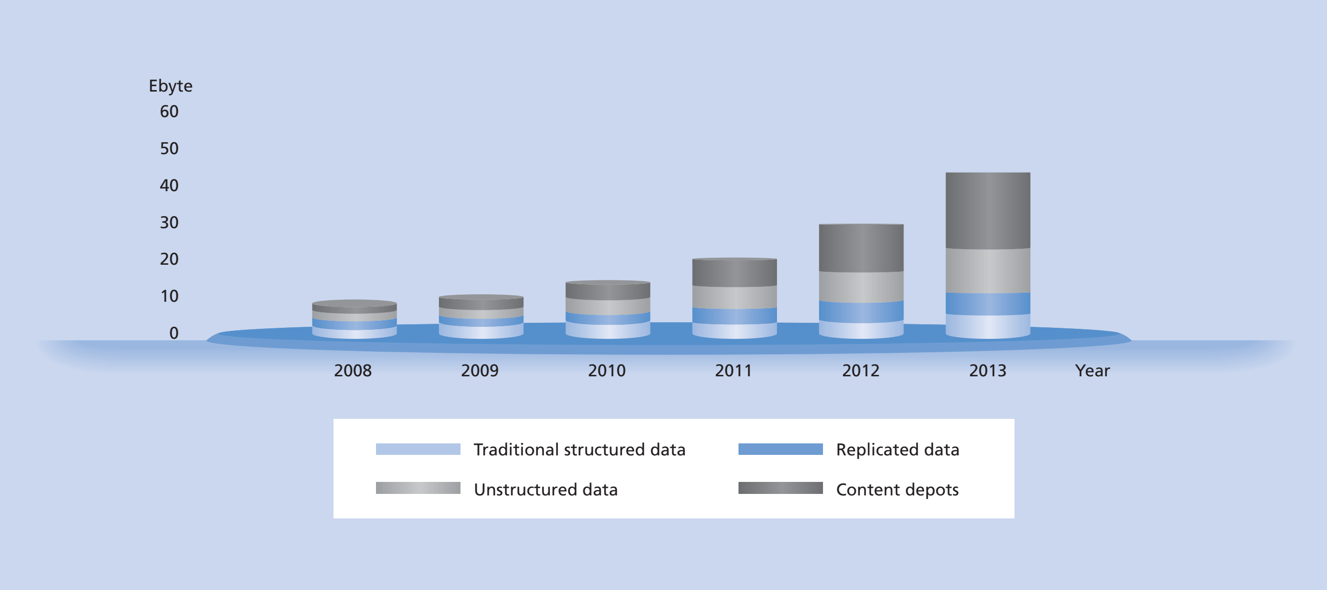

Fig. 1 | Disk Capacity Usage by Data Type[1]
The aging of the baby boomers will have a major impact on our global healthcare systems.

the sheer volume of data will push many institutions past the breaking point.

Healthcare, like many other industries is rushing to unlock the power of broad-based analytics. The ability to compare millions of medical imaging scans for common relationships, scan laboratory results for population patterns and genetic markers, and perform other forms of analytical mining are but a few of the keys to unlocking new medical compounds and delivering care in a far safer and more cost-effective manner.

However, it is the very data itself which is preventing us from unlocking these untold analytical riches. Far from the orderly, aligned, and obedient world of databases and structured data lies the fastest growing type of medical data asset we have today: unstructured data. They include medical imaging files, treatment reports, and scanned and paper records.

Not only is the growth of unstructured data nearing epidemic proportions but amassing this very data into the content depots we need in order to execute our analytical efforts is proving next to impossible (see **Fig. 2**). Unstructured data is big, hard to move, and its very definition fails to provide information about its content, value, or purposes for existence. To streamline the management of unstructured data and unlock its underlying value, storage vendors and application providers alike must aggressively move to embrace metadata and develop new paradigms for creating, managing, and utilizing it.

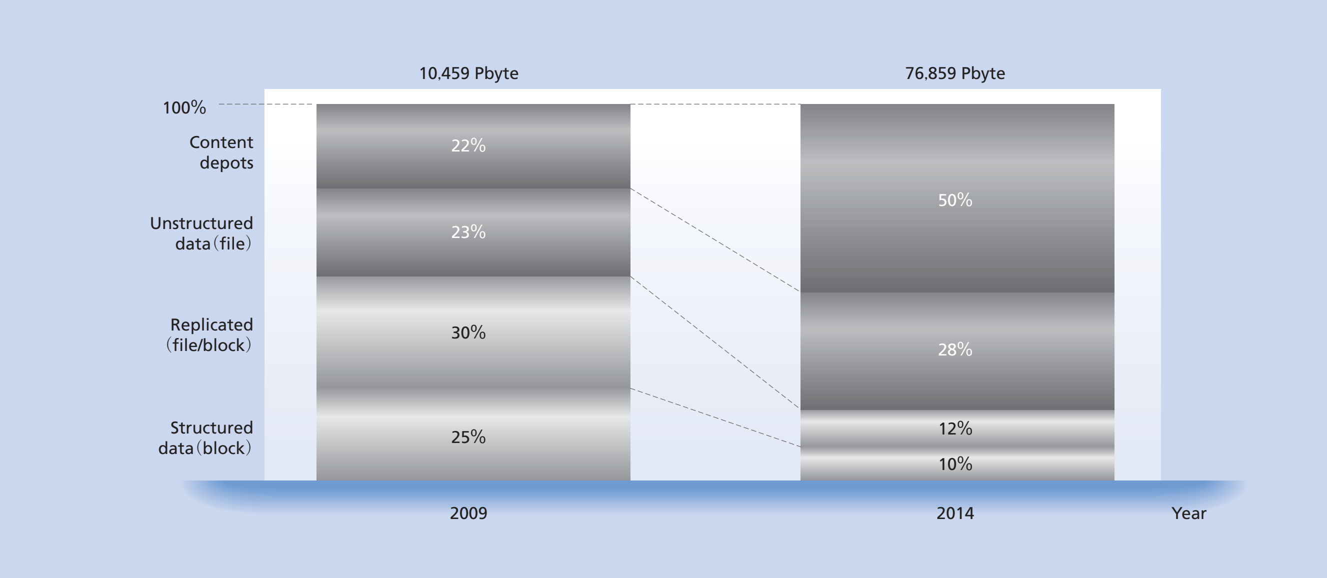

Fig. 2 | Storage Consumption by Data Type[2]
Amassing unstructured data into the content depots we need in order to execute our analytical efforts is proving next to impossible.

So how do we fix this problem? How do we survive the data tsunami and leverage this most valuable type of information –information regarding our very health and well-being? We do it by adding even more data to the pile. Not just any data but a very dynamic and living type of data called metadata.

What is metadata? Simply put it is "data about data." Metadata describes other data and, in the context of our discussion, the massive amounts of "unstructured data" files being amassed at a blinding rate. It provides information about a certain item's content. For example, a medical image may include metadata that describes how large the picture is, the bit depth, the image resolution, when the image was created, and other data about the medical procedure. A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document to include even clinical opinions or findings.

While this may seem rather academic and ethereal, many of us are exposed to metadata on almost a daily basis through the world's most popular and ubiquitous metadata driven device, the iPod*. Apple Inc. and its universally pervasive iPod is nothing more than a fixed content storage device managed through metadata. We load our MP3s (MPEG audio layer-3), MPEG (moving picture experts group) movies, audio books, and photos (fixed content) on our iPods by the thousands and this data never changes. It is static data that remains the same until we delete it or it is removed in some other fashion. Our metadata is often created for us, like Song Title, Artist and Album. However, it can also be a dynamic and living form of data such as "this song is one of my favorites, I have listened to it "x" many times, I last listened to it on "x" date and it is similar to these other songs of the same genre." It is the metadata that we create about each fixed content object on our iPod that makes it extremely user-friendly and unique in the world of MP3 players.

While the iPod example is fairly basic, this notion of self-assigned metadata on the iPod is far more advanced than what can be found within corporate and healthcare information systems today. Commonly accepted types of metadata in use today include:

(1) Basic metadata–low level data such as block-level information about where data is stored and how often it is accessed.

(2) File-level metadata–more complex data from file systems.

(3) Content-level metadata–metadata that might be found in content management systems such as file type and other more meaningful data derived from the file's contents such as "is this reporting referring to an MRI?" or "is the MRI report positive?"

Each metadata type is actionable. For example, basic metadata can be used to automate tiering, file-system data can be used to speed performance, and high-level metadata can be used to take business actions. The key challenge is how to capture, process, analyze, and manage all this metadata in an expedient manner. Static metadata (our iPod example) will only take us so far. To reach the next level, metadata must be dynamic, automatically generated, must change over time, and be associative with the world of applications we interact with. Much like our traditional models for computer hardware or systematic interaction with servers, networks, and storage devices, metadata models will need to advance beyond the basics to such models as that shown in **Fig. 3**. Associative metadata is but one of the many new and exciting paradigms for helping us deal with the massive amounts of unstructured data we are seeing in our healthcare environments today. Associative metadata is a paradigm in which unstructured data assets are indexed based on a disparate range of information, such as source, content, creation context, and relevance to a user, allowing a user to locate such files without the need to record a filename or location. Associative metadata allows a user to tailor the criteria used in a file search, and to dictate which criteria are and which are not important in the search for a certain file. However even this paradigm of associative metadata applied statically or by the applications that generate this content falls far short of where we need to be for true success. Pioneering work is underway right now on metadata robots that apply associative metadata as the content data itself is being formed and continues to enhance a content type metadata stream dynamically and far into the future. Bridging this gap will also allow us to enter the age of "Polymorphic Data Content" where our root content data exists in many forms throughout our data universe. Metadata itself is quickly becoming the barrier to enterprise data management and analytics. It has been said that our recent economic downturn has sped the adoption of cloud computing, with the promise of reduced capital expenditures, pay-as-you-go service models, and an on-demand world at your fingertips. Metadata creation, management, and utilization in business applications have the potential to stop cloud computing in its tracks. The growing

* iPod is a trademark or registered trademark of Apple Inc. in the U.S. and other countries.

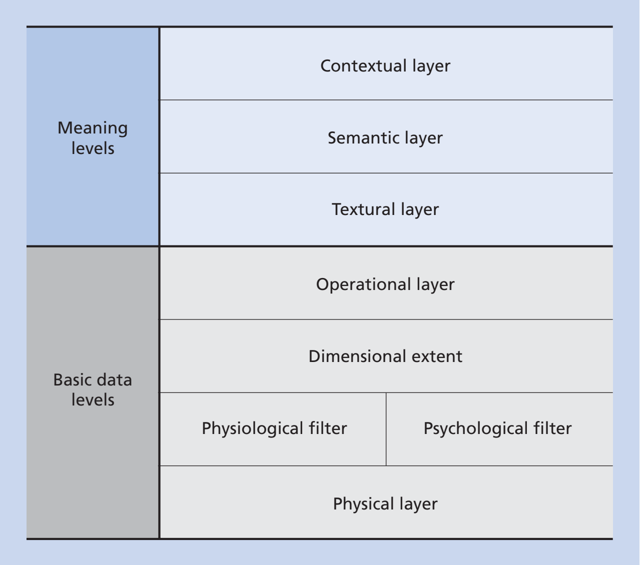

Fig. 3 Associative Metadata Layer Model[3]
Much like our traditional models, metadata models will need to advance beyond basic models.

cloud-based content depots and storage pools will quickly become black holes where we dump our data never to be seen, used, or understood again. Unlocking metadata however holds great promise and paradigm shifts for how we deal with our data. Rather than shoving the data into a big data repository, concepts like associative metadata allow us to distribute the metadata and allow parallel processing concepts to operate in tandem. By allowing the metadata to remain distributed, massive volumes of data can be managed and analyzed in real or near-real time, thereby providing a step function in metadata exploitation.

CONCLUSIONS

The data tsunami in healthcare is washing ashore today and few healthcare organizations are effectively dealing with it. Better understanding of how to create, harvest, manage, and exploit metadata is a very near-term problem to be addressed by today's information management professionals. Distributed data storage has been identified as one of the challenges in our paths towards cloud computing and without paradigm shifts in metadata management such as associative metadata our cloud computing initiatives risk quickly becoming "black holes" of lost and low relevance data.

As a worldwide leader in data management, Hitachi is taking the industry's first steps toward a more productive data future. The Hitachi Content Platform (HCP) and its systematically defined and embedded metadata stream capabilities have revolutionized application deployment and management.

The challenges of providing greater quality of care in a more efficient and cost-effective model are common themes across all healthcare delivery and research organizations around the globe. Our ability to generate information about our health and welfare has never been more advanced; it is our ability to understand this information and harness the power it yields that is trailing today. Metadata is becoming the key to managing today's healthcare data tsunami and our ability to embrace it is our only limitation.

REFERENCES

1) "Consumption of Disk Storage by Capacity: Forecast, Recovery, Efficiency and Digitization Shaping Customer Requirement for Storage Systems," ICD, IDC #223234, vol. 1, Tab: Markets Storage Systems: Market Analysis (May 2010)
2) "Storage Consumption by Data Type: Forecast, Recovery, Efficiency and Digitization Shaping Customer Requirement for Storage Systems," ICD, IDC #223234, vol. 1, Tab: Markets Storage Systems: Market Analysis (May 2010)
3) T. Coughlin and M. Alvarado, "Metadata Layer Model: Associative Metadata in Cloud Storage" (Sep. 2010)
4) T. M. Coughlin and S. L. Linfoot, "A Novel Taxonomy for Consumer Metadata," 2010 ICCE Conference (Jan. 2010)
5) T. Coughlin and M. Alvarado, "Angels in our Midst: Associative Metadata in Cloud Storage" (Sep. 2010)
6) J. Gray, "Distributed Computing Economics," Queue, vol. 6, no. 3, pp. 63–68 (2008)
7) A.E. Lindite, "A Case Study: Polymorphic Query Languages and Their Impact on Data Structures," University of Utah (Apr. 2009)
8) A. Verma and S. Venkataraman, "Efficient Metadata Management for Cloud Computing," University of Illinois for Digital Leadership (2010)

ABOUT THE AUTHOR

William A. Burns
Vice President, Global Health & Life Sciences, Hitachi Data Systems

Mr. Burns joined Hitachi in 2008 and leads the Global Health & Life Sciences Team at Hitachi Data Systems. Mr. Burns has extensive experience in the digital healthcare arena, including point-of-care disease management systems, diagnostic imaging, ambulatory patient monitoring, clinical research platforms, and regulatory compliance. He has proven instrumental in setting the leadership foundation in the development of technology-enhanced clinical and business initiatives for Hitachi Data Systems clients.
Mr. Burns has served on the Healthcare advisory boards of Microsoft Corporation, 3M Company, McKesson Corporation, the American Red Cross and Gillette Children's Hospitals. He is a lifelong member of the Institute of Electrical and Electronics Engineers and the American Management Association. He is also a featured speaker at several national and international forums on the topics of healthcare strategy and digital transformation.