HITACHI
Inspire the Next

WHITE PAPER

Healthcare Data's Perfect Storm

Why Healthcare Organizations Are Drowning in the Data They Are Creating and Why They Need Even More Data to Weather This Storm

By William A. Burns

August 2014

Contents

Executive Summary

Today's modern healthcare organizations, from hospitals to life sciences companies to healthcare payers, are struggling to keep up with the very data they are creating. Electronic medical record (EMR) 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 are chief information officers who are faced with an onslaught of data, the likes of which they have never seen before. Unlike the complex but tractable problem of managing structured databases from the massive to the mundane, the new menace is the management of individual files, themselves. Ranging from file servers to Microsoft® SharePoint® sites to the thousands of medical applications that populate today's healthcare landscape, this new hurdle is simply called unstructured data.

This white paper will help you understand both how and why this problem exists. It looks at why injecting even more information into this paradigm of unstructured data assets is the only way to stay ahead of the very data we are creating. It discusses the importance of restoring long-term value to this most critical class of data and delivering a cost-efficient infrastructure to manage today's healthcare information. And it explains how these efforts are key to reining in healthcare costs and maximizing the value of medical information.

About the Challenge

There is an unprecedented data surge happening in healthcare organizations around the world. Far from a “hockey stick” graph that will hit us in 5 to15 years, this is the reality today. Data growth is happening with such speed and veracity that looking at the numbers alone will leave organizations wondering how to survive this data explosion, let alone make sense of this valuable information.

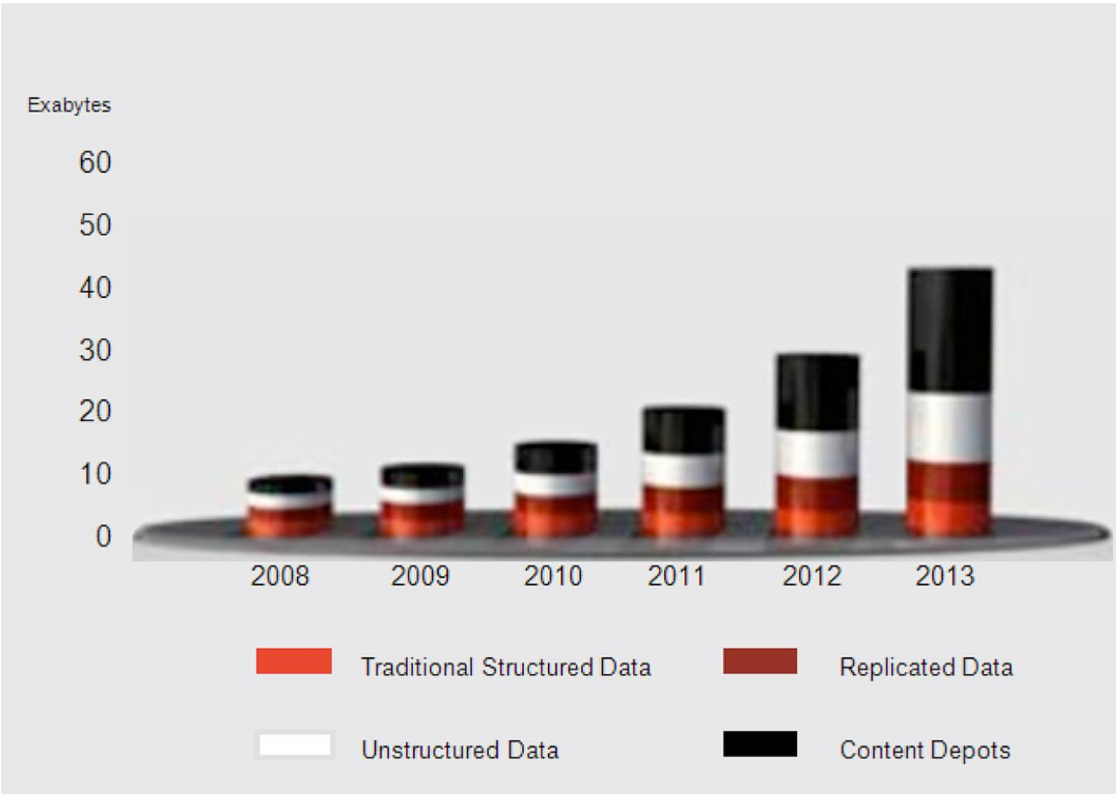
Consider the many sources of data. Current medical technology makes it possible to scan a single organ in 1 second and complete a full-body scan in roughly 60 seconds. The result is nearly 10GB of raw image data delivered to a hospital’s picture archive and communications system (PACS).

Clinical areas in their digital infancy, such as pathology, proteomics and genomics, which are the key to personalized medicine, can generate over 2TB 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 processes.

And finally, consider the impact of EMRs, 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. Consider the thousands of medical applications in use today that create an individual’s files or unstructured data (see Figure 1)[1] during the patient care process. That information gets stored on countless computers, servers and storage systems. It is impossible to ignore the impact that retiring 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 his or her life. As retiring baby boomers worldwide make greater use of healthcare systems, the sheer volume of data will push many institutions past the breaking point.

[1] “Consumption of Disk Storage by Capacity: Forecast, Recovery, Efficiency and Digitization Shaping Customer Requirement for Storage Systems,” IDC, May 2010

Figure 1. Consumption of Disk Capacity by Data Type



Unstructured Data

Healthcare, like many other industries, is rushing to unlock the power of broad-based analytics. Healthcare seeks the abilities to compare millions of medical imaging scans for common relationships and scan laboratory results for population patterns and genetic markers. These and many 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. This includes medical imaging files, treatment reports, and both scanned and paper records.

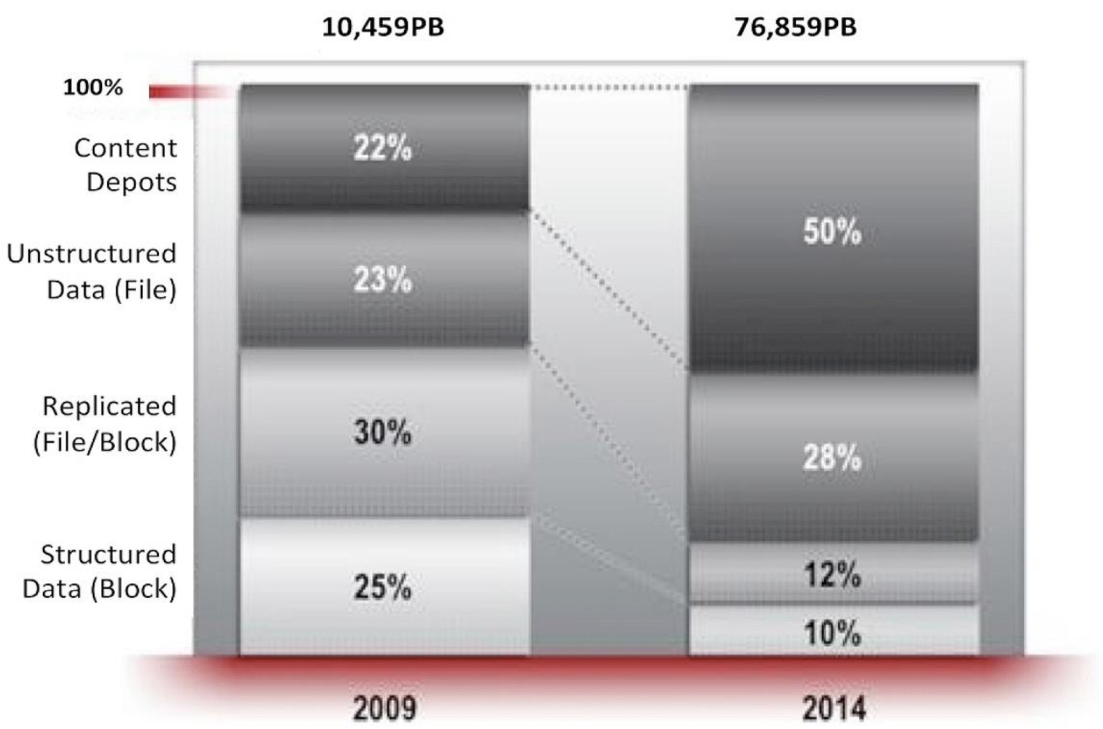
The growth of unstructured data is nearing epidemic proportions. In addition, amassing this very data into the content depots we need to execute our analytical efforts is proving next to impossible (see Figure 2)[2].

Unstructured data is big and hard to move. Its very definition fails to provide information about its content, values or purposes for existence. To streamline the management of unstructured data and unlock its underlying value, storage vendors and application providers alike must quickly move to embrace metadata; they must develop new paradigms for creating, managing and utilizing metadata.

---

[2] "Storage Consumption by Data Type: Forecast, Recovery, Efficiency and Digitization Shaping Customer Requirement for Storage Systems," IDC, May 2010\

Figure 2. Storage Consumption by Data Type



Metadata

So how do we fix this problem? How do we avoid the data storm 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.

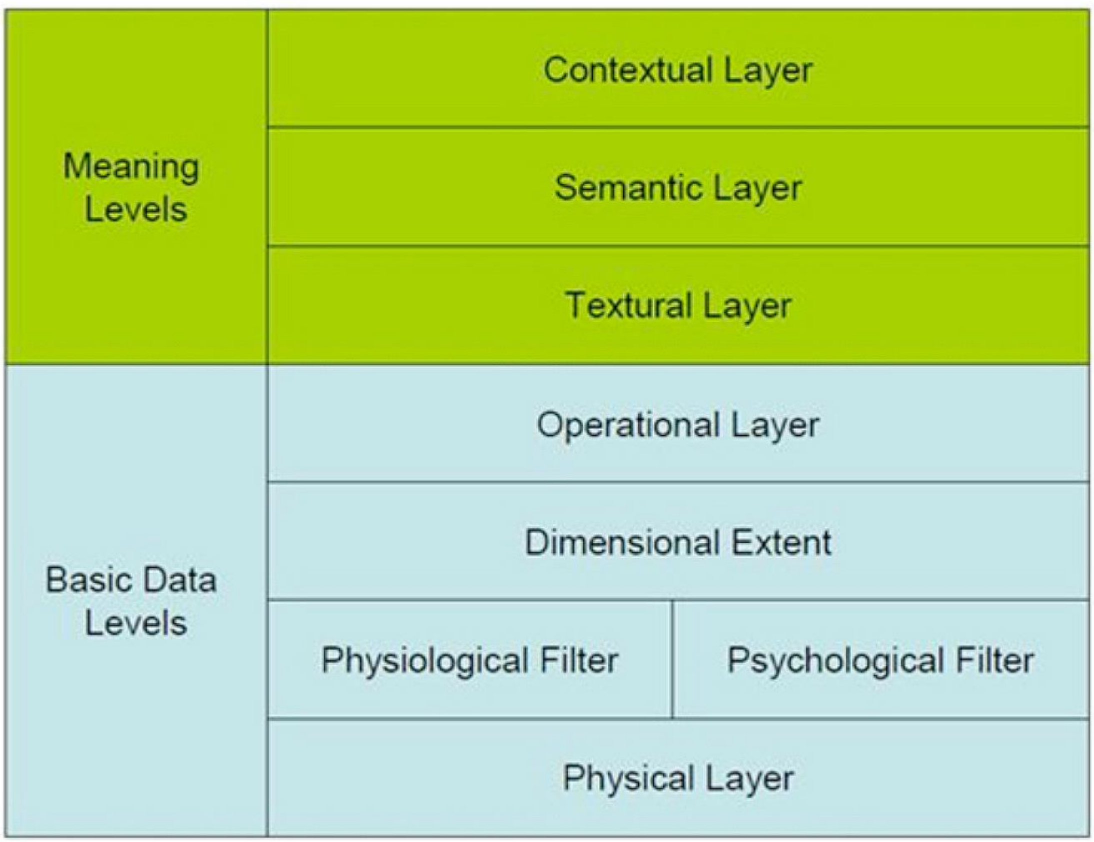
What is metadata? Simply put, it is "data about data." Metadata describes other data and, in the context of our discussion, the massive amounts of unstructured data files being amassed at a blinding rate. It provides information about a certain item's content. For example, a medical image may include 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 Computer's universally pervasive iPod is nothing more than a fixed-content storage device managed through metadata. We load our MP3s, MPEG 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. For example, it may report: "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 in these areas today include:

- *Basic metadata.* This is low-level data, such as block-level information about where data is stored and how often it is accessed.
- *File-level metadata.* This is more complex data from file systems.
- *Content-level metadata.* This is metadata that might be found in content management systems, such as file type. It also may include other more meaningful data derived from its very contents, such as: "Is this report referring to an MRI?" or "Is the MRI report positive?"

Figure 3. Associative Metadata Model



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, able to change over time and associative with the world of applications in which we interact. 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. They must move to models such as those shown in Figure 3[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 disparate types of information. This data may range from source, content and creation context, to its relevance to a user, and allow a user to locate such files without the need to record a filename or location.

Associative Metadata

Associative metadata allows a user to tailor the criteria used in a file search. The user can 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 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. The cloud has promised reduced capital expenditures, pay-as-you-go service models and an on-demand world at your fingertips. Yet, metadata creation, management and utilization in business applications have the potential to stop cloud computing in its tracks. Growing, 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 enable 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.

Conclusion

Weathering the data storm gets harder every day, and few healthcare organizations are effectively dealing with it. The challenge of gaining 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. Without paradigm shifts in metadata management, such as to 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 Data Systems is enabling industries to take their first steps toward a more productive data future. Hitachi Content Platform 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

[3] “Angels in our Midst: Associative Metadata in Cloud Storage,” Tom Coughlin and Mike Alvarado, downloaded 11/3/2011, Coughlin Associates, www.tomcoughlin.com/Techpapers/Angels+in+our+Midst,%20102710.pdf

yields that is trailing today. Metadata is becoming the key to winning today's healthcare data challenge. Our ability to embrace it is our only limitation.

Appendix A: References

- "A Novel Taxonomy for Consumer Metadata," T. M. Coughlin and S. L. Linfoot, 2010 ICCE Conference in January 2010.
- "Angels in our Midst: Associative Metadata in Cloud Storage," Dr. Tom Coughlin and Dr. Mike Alvarado, September 2010, http://www.tomcoughlin.com/Techpapers/Angels+in+our+Midst,%20102710.pdf.
- "Distributed Computing Economics," J. Gray, Queue, vol. 6, no. 3, pp. 63-68, 2008.
- "A Case Study: Polymorphic Query Languages and Their Impact on Data Structures," A.E. Lindite, University of Utah, April 2009.
- "Efficient Metadata Management for Cloud Computing," Verma, Abhishek; Venkataraman, Shivaram; January 19, 2010, http://hdl.handle.net/2142/14820.

Appendix B: About the Author

William A. Burns

Vice President, Global Health and Life Sciences, Hitachi Data Systems

William Burns joined Hitachi in 2008 and leads the Global Health and Life Sciences Team at Hitachi Data Systems. With extensive experience in the digital healthcare arena, Burns has worked with 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.

Burns has served on the healthcare advisory boards of Microsoft, 3M, McKesson Corporation, The American Red Cross and Gillette Children's Hospitals. He is a lifelong member of 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.

WHITE PAPER

Hitachi Data Systems

Corporate Headquarters
2845 Lafayette Street
Santa Clara, CA 96050-2639 USA
www.HDS.com community.HDS.com

Regional Contact Information
Americas: +1 408 970 1000 or info@hds.com
Europe, Middle East and Africa: +44 (0) 1753 618000 or info.emea@hds.com
Asia Pacific: +852 3189 7900 or hds.marketing.apac@hds.com

WP-419-C W.Burns August 2014