virtualworks

How to Tame Data Sprawl Through Content Virtualization

A light approach to a heavy problem

Table of Contents

Two of the greatest information trends today are the huge value of digital information within an organization and the near impossibility of accessing and using this information because of data sprawl.

If information is the lifeblood of any organization, then today's enterprises are bleeding badly.

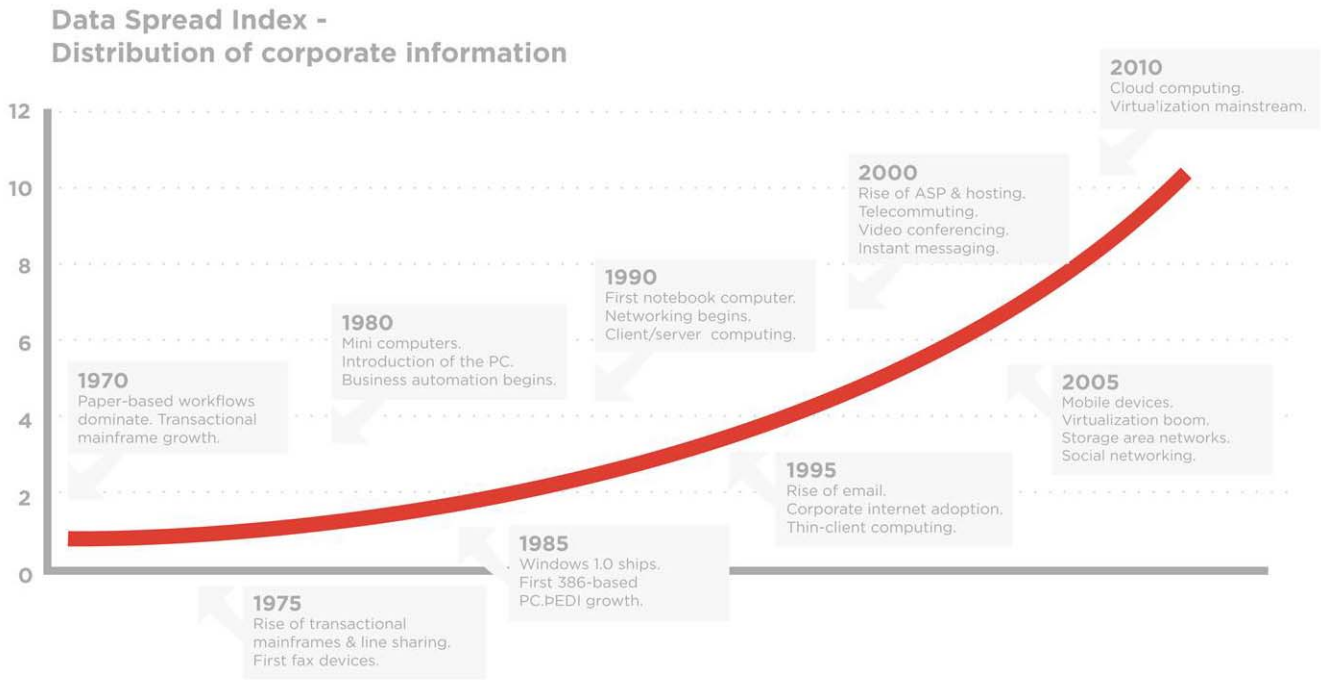
Today's enterprises have never been more fragmented. The data that used to live in structured warehouses and databases is now scattered around in emails, file servers, intranet sites and scanned files—not just behind the firewall but out in the cloud as well. And it's only getting worse as data volumes continue to double each year.

How Did We Get Here?

Thanks to the precipitous drop in storage costs—from $700 per megabyte in 1981 to $.002 per megabyte in 2010 – the need to carefully manage storage is a thing of the past. Today, if you need more storage you simply plug-in more terabytes, copy your archives and keep going, never looking back…until you actually have to find something.

We're now spending more money to maintain data than to create it. Then there's so much of it that we can never put our hands on what we need, when we need it. The simple truth is that data has grown beyond our control, and it just keeps on growing. According to IDC, there's been a ten-fold increase in data stored during the last five years (2006-2011). In the world of commoditized storage, the price you pay is now on the back-end in the form of data sprawl. And the penalties are very real.

The Data Sprawl Index (DSI) is our way of calculating how far information has spread and expanded over the past 40 years since the advent of modern computing. The DSI is a measure of the average change in the distribution of enterprise information over time, representing the data usage patterns of U.S.-based organizations as it relates to major technology events and usage trends.



The Data Challenge

"Half of all organizations that have moved to the cloud are now transitioning back to on-premise because of data integration issues."

Gartner Group

The Perfect Data Storm

Since 2000, when the cost of storage hit $.01 per megabyte, three IT trends have risen to dominance, creating a perfect storm that has resulted in the deep fragmentation of enterprise data. These trends encompass virtualization, cloud and mobile computing.

Virtualization has radically transformed IT, saving millions for organizations as they consolidate servers, storage and networking. The core technology came at a perfect time, when IT shops were being asked to do more with less.

Virtualization breaks the physical and logical link in a new way, spreading the entire computing process across a wide range of physical devices, whether it's processing, end user devices, storage or networking. However, virtualization also adds a new layer of complexity for the organization.

As it turns out, physical server and storage limitations put a cap on the growth of data. Virtualize them, and you introduce a new element to manage: data sprawl. Virtualization accelerates the duplication, disintermediation and confusion of data.

A Cloudy Forecast

Virtualization has contributed to the accelerated mobility of information while spurring the growth of cloud computing, which for many organizations, has raised data sprawl to the next power. More than just the evolution of hosted applications, cloud-based computing now offers alternatives for almost every part of the modern computing stack, from raw storage and processing to finished applications and services.

Cloud vendors have successfully leveraged the public Internet backbone with advances in distributed computing, virtualization and multi-tenant architecture to create compelling technology offerings for all types and sizes of organizations. According to IDC, the number of servers deployed in cloud applications is expected to triple to 1.35 million by 2014.

While organizations of all sizes are rushing to embrace the cloud for the financial, speed and agility benefits it can bring, they are quickly discovering that it adds yet another layer of complexity to the mix by creating an unlimited ability to quickly push valuable corporate information beyond the firewall into new public and private cloud infrastructure at a reasonable price. In fact, Gartner Group reports that "half of all organizations that have moved to the cloud are now transitioning back to on-premise because of data integration issues."

Information Goes Mobile

Let's not forget the rise of mobile computing, which amplifies the problems of splintered data and different data views. Data is being generated by more than 4.3 billion cell phones worldwide, making it increasingly difficult to manage

and secure information. The consumerization of IT contributes to rogue content saved locally, which in turn, creates enterprise compliance and financial risk.

The High Cost of Data Sprawl

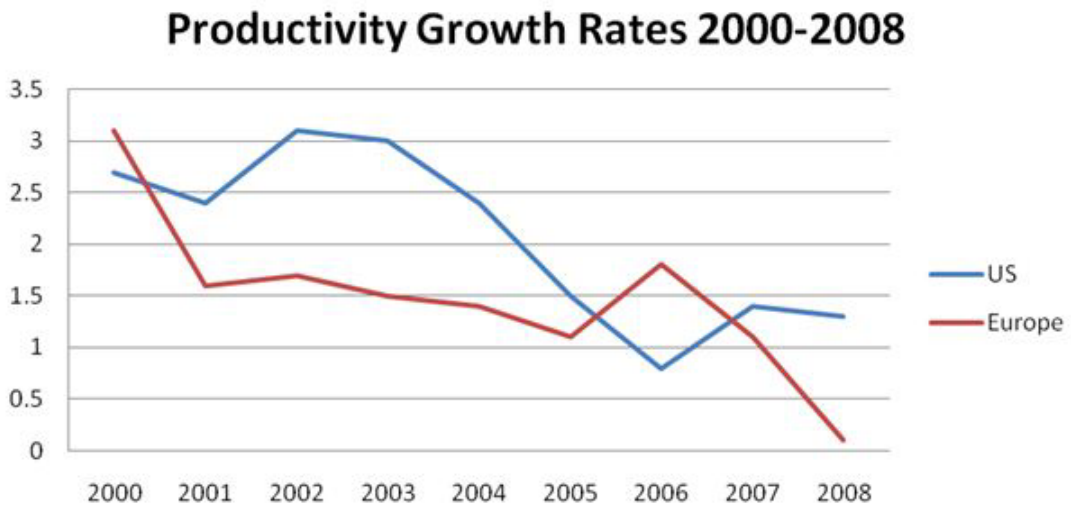
For decades, advances in information technology have significantly boosted worker productivity with everything from email to new ERP and CRM apps. In fact, global productivity has grown steadily since the 1990s. But since the millennium, we seemed to have reached a point of diminishing returns; we're simply not getting the same rate of return with new IT investments.

Data sprawl is simply choking the system. As a result, the actual productivity growth rate of information workers has slowed dramatically. The reduction in productivity is positively correlated to data fragmentation and unmanaged information stores.

If information can't be found when it's needed, it's a huge waste; and wasted time looking for information is only the beginning. There's also wasted time, effort and budget resources for acquiring, creating, storing and managing the information. Not to mention the extra waste when it all gets duplicated because the next person who needs the information has no idea it already exists in the system and starts the whole process over again.

Scattered information silos also cost the enterprise in terms of risk; ad hoc workarounds can compromise security, and data silos can make it difficult, if not impossible, to conform to regulatory and compliance requirements. All this adds up to a negative impact on performance and competitive ability.

The actual productivity growth rate of information workers has slowed dramatically since 2000 — a positive correlation to data fragmentation.



Source: OECD productivity growth rates 2000-2009 www.oecd.org

Early Approaches for Taming Data Sprawl

The results so far have been over-engineered systems that are expensive to buy and maintain, create a huge drain on performance, and are very likely to be ignored or worked around by the end user.

Putting Structure Around an Unstructured World

How can you get your arms around information when it's chaos out there?

Previous attempts at solving these problems involved putting controls and structure around unstructured data either by forcing users to behave according to a set of rules or managing data across the entire enterprise. Such measures are simple for databases, but impossible for emails, charts, documents and almost any other type of unstructured data.

If lack of structure was the problem, early solutions focused on providing more structure, in the form of:

- **Data warehouses.** Massive, centralized data stores that were meant to replace the hundreds of silos.
- **Document management systems.** Extensive (and expensive) knowledge management stores with pre-structured data schema and intensive metadata rules.
- **Data governance initiatives.** New rules, policies and processes designed to enforce better data hygiene practices.
- **Traditional enterprise search.** Too centralized, too hierarchical and too insecure for the way data lives.

Unfortunately, while each of these attempts delivered some benefits, none of them solved the problem. The results so far have been over-engineered systems that are expensive to buy and maintain, create a huge drain on performance, and are very likely to be ignored or worked around by the end user. The moral of the story is that when you try to cage and control information, it just doesn't work.

One reason may be the futility of trying to get users to change their behaviors. Another might be the sheer scale of the problem: the volume of data; the number of applications silos, servers and stores; and the many different locations for these silos – in the cloud and behind the firewall. Data continues to pour out of every corporate crevice, lurking in odd nooks and crannies and defying structure.

Clearly the old ways of dealing with data just can't keep up.

The March of IT Virtualization

Few trends in the history of IT have been as all-dominant as virtualization. From its quiet beginnings as a mainframe technology that enabled better CPU usage, virtualization has swept through enterprise IT over the past decade, progressing in two big waves.

Virtualization of the Data Center

The first virtualization wave encompassed how servers and storage systems are managed in the data center, delivering dramatic improvements in resource utilization, cost, energy use, footprint and flexibility. Whereas the hand wiring of servers used to be complicated, time consuming and prone to errors, virtualization now enables IT to pop out servers like cookie cutters by simply pointing at the system; and virtual storage in the cloud allows for dynamic expansion or contraction without forklifting in new hardware.

Virtualization of the Desktop

The second virtualization wave migrated to operating systems, desktops and applications, making the provisioning of new desktops easier and cheaper and secure access to corporate resources simpler. The 'one user, one desktop, one app' model is quickly coming to an end.

The abstraction of the physical reality into a flexible and adaptable virtual form provides a more usable and efficient environment for IT professionals to logically replicate physical hardware, minimize reconfiguration, grow storage transparently, deploy desktops faster, and make their lives easier. In short, you can virtualize just about anything and move it around logically in IT's world.

As it turns out, physical server and storage limitations used to put a cap on the growth of data. Virtualize them, and you introduce a new element to manage: data sprawl.

The Virtualization Divide

"For every 5-10 employees, there is one application, generating copies of the same files everywhere, which in turn, leads to massive amounts of duplicate data strewn all across underutilized storage systems."

Michael Vizard, "The Real Big IT Problem," ITBusinessEdge.com, March 2011

Spinning Out End-user Complexity

While data center virtualization has lowered costs and made IT management simpler, end users suffer more complexity in trying to locate, access and manage content across myriad devices and locations.

The past two waves of virtualization have spawned multitudinous file and data locations all over the place, putting pressure on end users through physical resource multiplication. The issue is no longer about simply finding a piece of information when it's needed. The issue now centers on finding the plethora of copies housed in multiple locations—across virtual servers, off-boarded storage, attached to emails hosted in the cloud, or sitting in virtual and physical desktops.

For example, if you need to find a document to edit, you must first know where the many different versions of that document reside, locate them and then deal with them individually.

The problem is you may have one version of the file housed in one or more virtual servers, another version stored in the Amazon cloud, and a third version attached to an email you sent to your colleague for review.

Where do you even begin to look? Is the file stored in your virtual desktop 'My Documents,' or your physical 'My Documents'? Is it in the virtual desktop you use to connect to your home computer? Or did you download it to your mobile device?

This Time it's About the End-users

Delivery of information is no longer about loading up and spinning out applications anymore; it's about slicing out the specific pieces of information users need to get the job done. The enterprise is ripe for a new information paradigm where everyone can find every piece of information they need, when they need it, with a light and agile interface that actually super-charges productivity instead of creating a drain on it.

The virtualization concept has now advanced to embrace the content that lies at the heart of every organization, promising to liberate it from the storage recesses and application silos that have curtailed its use and value.

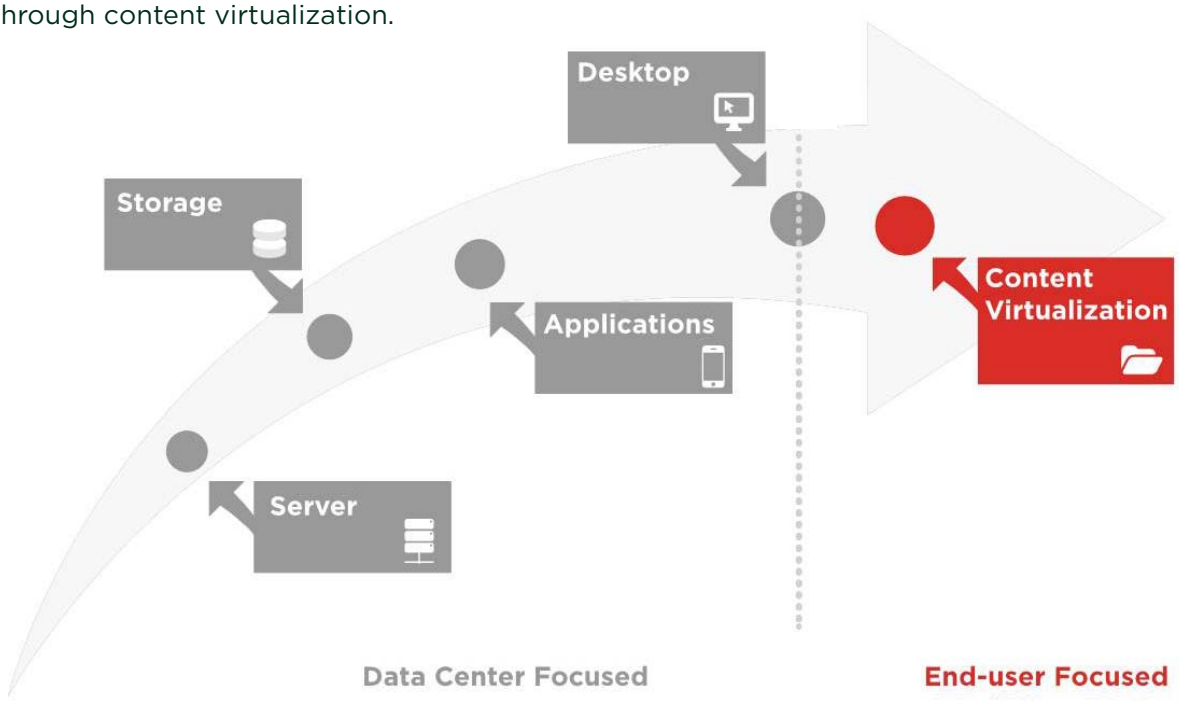
It now stands ready to transform the very core of the enterprise itself through content virtualization.

As its name implies, content virtualization is based on the ideas that have made other virtualization technologies so popular: resources on demand, distributed architectures and agile infrastructures.

By creating a virtualized representation of content or data from across the entire organization, an abstracted platform can be provided to users whereby all relevant content is quickly and securely served up, on-demand.

In an era governed by data sprawl, content virtualization is a breakthrough technology on par, or exceeding, other virtualization applications in the productivity benefits it produces.

Content virtualization is based on the ideas that have made other kinds of virtualization so popular: resources on demand, distributed architectures and agile infrastructures.



Introducing Content Virtualization

"Virtualization continues to demonstrate additional tangible benefits the more it's used, broadening its value to the enterprise at each step."

CIO.com

What is Content Virtualization?

Content virtualization breaks down the walls between application silos to ease data sprawl and deliver a more productive end-user experience.

It abstracts content from disparate data silos and captures it in a secure common index, giving users immediate access to a single, logical view of their data regardless of its location. The index is then federated across the enterprise, so users can securely find and access everything from anywhere without leaving their application or device.

Content virtualization is the lightest approach to dealing with the heavy problem of data sprawl – requiring no heavy lifting, custom integration or changes in how end-users or organizations work. It makes all fragmented applications, and their related data silos, work together as one cohesive content system, empowering users to draw upon the collective intelligence and information of the entire organization.

Defining Characteristics

Content virtualization leverages a deep understanding of applications and stored metadata to render an image of all enterprise content across diverse applications and devices, network and cloud architectures, data storage systems and formats. The defining technical attributes of content virtualization include:

Common index. The most critical component of a content virtualization platform is the core index itself, which acts as a unifier for all enterprise information and a single go-to point for finding anything within the enterprise's data stores. Content virtualization platforms must organize and index data using a common index structure and make the data accessible to any authorized application or user device, regardless of format or structure.

Non-disruptive. Unlike complex systems that try to impose structure on unstructured data or new processes on users, content virtualization platforms leave data where it lives while making it available to everyone who needs it and is authorized to see it.

The system must not interfere with neither the functioning of applications nor the structure and integrity of data in performing its role. Rather, it must sit on top of other platforms and existing security systems in a lightweight model in order to build the common index dynamically and continuously.

Any application. Content virtualization platforms possess the ability to understand the metadata of the various applications in use by an organization and provide the means for their interrogation. The system must support all file formats and data types, structured and unstructured, as well as legacy data.

Any location. Location, whether physical or virtual, must not be an impediment to the system. Content virtualization spans all types of local, SAN and NAS storage; server configuration, whether local or remote; and cloud service, whether public or private.

Any device. Content virtualization must federate disparate content into a single view regardless of end-point and support all client device models: physical, virtual, distributed and mobile.

How it Works

The 'big four' principles of content virtualization include:

1. Find all information

Discover data wherever it lives, across hundreds of applications, servers, intranet sites, email and file stores – whether hosted in the cloud or on premise in the data center.

2. Index everything

Identify and list all data in the enterprise, making it easy to index everything in a central or virtualized cluster of indexes.

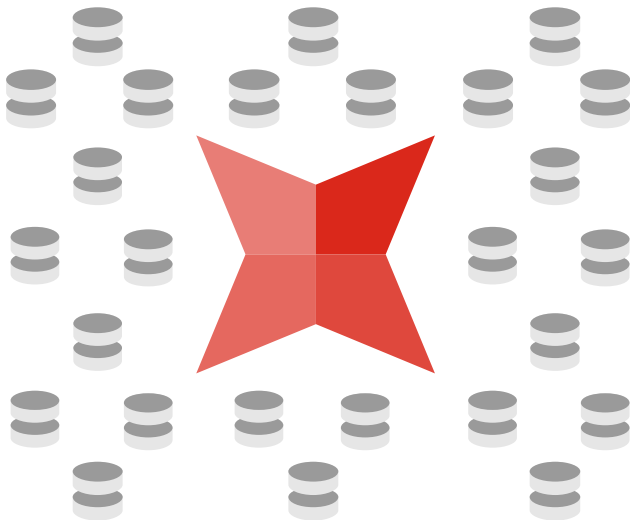
3. Secure access

Use existing security schema (such as Active Directory) to associate each piece of content with its access privileges, so only authorized users can see it.

4. Leverage

Make all content available to authorized users and to a new generation of index-enabled applications, on-demand.

Content virtualization is the lightest approach to dealing with the heavy problem of data sprawl – requiring no heavy lifting, custom integration or changes in how end-users or organizations work.



Content Virtualization Benefits

"Virtualization is no longer just about server consolidation... In virtualized environments, it's easier to move things around, to encapsulate, to archive and to optimize."

CIO.com

Business Advantages

Virtualization has already transformed enterprise IT. Now content virtualization is set to transform the way that users inside enterprises can harness the content they need to make better decisions, in real time, based on the right information, regardless of where it lives.

It enhances the end-user experience by providing a simple, comprehensive content experience so they can:

- Make faster, more informed decisions with the right information at the right time.
- Boost productivity by eliminating wasted time searching for or recreating existing data.
- Accelerate time-to-insight by uniting content across all enterprise data stores, structured and unstructured, behind the firewall or out in the cloud.
- Simplify the user experience by providing a single view of all data (and data relationships) regardless of the app it is created or stored in.
- Get more value from every application by unlocking the data stored inside.

It also helps IT departments do their jobs better because it delivers data control and insight without massive management overhead, making it easier to:

- Secure all content, so users see only what they are authorized to see.
- Lighten the data management burden by leaving data where it lives, but discovering and indexing it for use.
- Unite all content stores whether in the cloud, the data center or any server or storage system in-between.

The Power of Content Virtualization

By leveraging the power of content virtualization, organizations of all sizes, from SMBs to global enterprises, can solve a critical pain point for frustrated workers trying to find the information they need in a timely manner across a variety of applications, locations and devices. Content virtualization allows companies to unleash the power and value of information, dramatically improving the way organizations make decisions, exploit opportunities and respond to change.

Here are just a few examples of the many ways organizations are using content virtualization to attack some of their most pressing information challenges:

Manage unstructured files and email servers. Instantly and securely access millions of files, emails and attachments – even in the cloud.

Go paperless. Index all scanned files (with powerful OCR), and make it instantly visible to the people who need it.

Empower SharePoint and Notes. Index all the information locked inside multiple SharePoint or Notes systems to make it accessible to all authorized users.

The Indexed Intranet. Index the intranet and users will instantly find what they need and come back for more.

Real-world document management. Make document management systems user-proof — or skip the DMS entirely and index all content instead.

Let the Data Sprawl

Content virtualization dramatically mitigates the downside of data sprawl. For the first time, enterprises can let data stay wherever users want it to be without losing its value to the entire organization.

As applications continue to proliferate, more data moves into the cloud and knowledge workers mobilize their content assets, the need for content virtualization has never been greater.

About VirtualWorks

VirtualWorks is pioneering a new software category called content virtualization that breaks down the walls between application silos to ease data sprawl and deliver a more productive end-user computing experience. Headquartered in Boca Raton, FL, with offices in Norway and Sweden, the company is led by Edward Iacobucci, a prominent high-tech entrepreneur and co-founder of Citrix Systems.

The company's technology is currently used by more 300 private and public sector organizations in Norway, Sweden and the United States to unleash the power of information no matter where it resides within the organization. For more information, please visit www.virtualworks.com.

VirtualWorks is pioneering a new software category called content virtualization that breaks down the walls between application silos to ease data sprawl and deliver a more productive end-user computing experience.

USA
5301 N. Federal Highway
Suite 190
Boca Raton, FL 33487
561.327.4900

Norway
Lysaker torg 45
1366 Lysaker
(+47) 81 536 863

Sweden
Kanalgatan 51
931 32 SKELLEFTEA
(+46) 70 543 7700

www.virtualworks.com