

SPONSORED BY:

MOBILE



Transform Your Business With Data

ABOUT US REGISTER LOGIN

HOME BLOGGERS MESSAGES POLLS WEBINARS RESOURCES VIDEOS | STRATEGY ANALYTICS TECHNOLOGY FB TW LI G+ RSS

FINANCIAL SERVICES HEALTHCARE GOVERNMENT EDUCATION RETAIL

BIG DATA STORAGE

Big Email Data: Junk or Treasure?



Robert Plant, Associate Professor, School of Business Administration, University of Miami
5/22/2013
Comment
10 comments

Login
50% 50%

Most of us think we have a big data set related to our email storage; I personally have about six gigabytes of email. The business school where I reside has about 3 Terabytes of live email storage containing emails going back to around 1995, when we moved to Windows NT. However, this is just a drop in the proverbial email ocean when compared to some institutions.

FAO Mr President

This was recently highlighted by the curators of the George W. Bush Presidential Center (library) who have an 80 terabyte database containing some 200 million emails from the 43rd President's time in office. These emails were collected under the auspice of the Presidential Records Act, passed in 1978 to create a collection of all presidential records during a presidential term, and the Hatch Act of 1939, which is intended to prevent "Pernicious Political Activities" but interestingly does not pertain to the president, vice-president, and other executive branch officials.

When these two pieces of legislation were drafted it was inconceivable that the data generated would take -- at a rate of 800,000 messages a year, the current maximum capacity -- 250 years to analyze. The overhead comes from the fact that each message needs to be reviewed individually and if necessary undergo censoring.

Recreating the trail

Additionally, the President's office has to work within the Hatch legislation, the aim of which is for the President's office to be able to recreate the data trail, should inconsistent or potentially illegal activities take place. Under such legal regulation, the Bush years were surrounded in controversy as allegations and claims of lost emails, up to 22 million, were made including that servers outside the existing government channels were being used for official email communications. This compounds the issue of accountable deletion of data.

The need for storage and archiving is an interesting conundrum, raising the question: should we store everything? Or just the data associated with seemingly "important" things? The onus of responsibility for emails at the White House falls on the shoulders of the President, who according to the act has "responsibility for the custody and management of incumbent Presidential records," but also gets to play Chief Information Officer, as the act allows "the incumbent President to dispose of records that no longer have administrative, historical, informational, or evidentiary value, once he has obtained the views of the Archivist of the United States on the proposed disposal." Of course, the question is: What should he, just like CIOs and individuals, actually dispose of, if anything?

This is a difficult question to answer as it depends on what the data could data be ultimately used for.

Making decisions

One use is to improved decision making, to use email data to look at connectivity within intra- and inter-organizational communication channels. For example, to identify who speaks to whom, why members of the organization were involved in a project, and how long projects took to complete. This data could be then layered with the process Meta data is passed back and forth between constituents and ultimately combined with outcome data to improve the processes.

Spotting trends

A second use is to identify trends. Product data in email sets may be very useful to flag early warnings on product issues. Just as in the way that Google uses search to identify flu outbreaks two weeks before the CDC, emails may be useful trend indicators to identify trending product issues and specific product component issues before a critical mass of formal customer complaints is logged.

Trends in the business space

A third approach is the use of email data to identify trends and cycles in the business's environment. For example, in healthcare, the patient emails to clinicians could, for example, provide data sets for the onset of symptoms during different times of the year and how they were resolved. In manufacturing industries, emails can be combined with data from marketing campaigns, POS data, and social media to build profiles of potential new products.

So, while President Obama wrestles with whether to press delete on an email or archive it, business leaders would be wise to refrain from deleting en masse an email database because firstly, they need to ensure accountability to regulators, and secondly, when combined with other big data sources it may provide a fertile and ready-made big data set for development of new ideas, innovative products, and process improvement.

MORE BLOGS FROM ROBERT PLANT

A Review: Big Data in 2013

39 comments
Data scientists, NSA, C-suite buy in... Big data changed a lot in the last 12 months.

What Would Seymour Cray Think of Big Data?

13 comments
How would the tech legend respond to the 4 Vs?

The 360 Degree View of a Customer

56 comments
Big data is helping companies like Netflix see what you really want (and even when you go to sleep).

Lessons for Big Data From President Obama's Healthcare Implementation

24 comments
Three big data lessons for CEOs from HealthCare.gov's bumpy start.

Is Big Data Ruining Sports?

29 comments
From sailing to the NBA -- big data could increase the gap between sporting haves and have-nots (while ruining the magic in the process).

More from Robert Plant



FLASH POLL

All Polls

DIGITAL AUDIO



Big Data Storage Radio

Latest Archived Broadcast
Locking Up Your Cloud: Security in the Ether
Data security is precious enough when it's sitting in your office, so what's the best way to maintain security standards once you migrate to the cloud?
Listen

Full schedule Archived shows Download to calendar

BDR IN YOUR INBOX

Enter email SUBMIT

FEATURED VIDEO

Big Data Explained: What is a Data Mart? 18



Use this video to explain quickly to someone the difference between a data warehouse and a data mart.
Watch This Video

Related posts:

- [Big Data Is a Big Mystery to Local Governments](#)
- [Big Data Explained: What Is Volume?](#)

Email This Print Comment

COMMENTS

Newest First | Oldest First | Threaded View



AlphaEdge, User Rank: Exabyte Executive  
5/28/2013 | 11:47:10 AM

Re: It Can Be Junk For Someone And Treasure For Some One

ENRON email data set provides enormous research for computer science community. It can be very useful insights on topic/key words reserach, social network analysis, as well as event detection.



50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board



a.saji, User Rank: Megabyte Messenger  
5/23/2013 | 10:37:19 AM

Re: It Can Be Junk For Someone And Treasure For Some One

@netcrawl: Yes it is a headache but we cannot do anything other than blocking but for that also we should know the sender address.



50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board



James M. Connolly, User Rank: Blogger  
5/23/2013 | 8:53:04 AM

Re: Junk

@netcrawl. Right, Email remains important from a forensic point of view, as you note in the Lehman case. However, just in day to day corporate activity it has revolutionized how we do business. You can get a message/query to anywhere from one person to the whole company in an instant. Most important is the fact that you don't have to coordinate the time for the conversation. Think how often you may have walked across a building or up and down stairs in the office to have even a brief conversation with someone, only to find that they aren't in yet, they are at lunch, or they're on the phone. For probably 90 percent of messages email is just fine. For the remaining 10 percent you need phone or face to face.



50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board



netcrawl, User Rank: Petabyte Pathfinder  
5/23/2013 | 2:55:46 AM

Re: Junk

@James its still best corporate communication tool (that's for me), especially in high technology and financial industry. Email is very important, it contained "precious data", in handling litigation around corporate meltdown (remember Lehman Brothers?), email plays a key role because it can solve a case, contained a missing links or data to solve a "corporate puzzle".



50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board



netcrawl, User Rank: Petabyte Pathfinder  
5/23/2013 | 2:46:27 AM

Re: It Can Be Junk For Someone And Treasure For Some One

@debera I agree with you, it sreally true email junk is a big headache when it comes to management, it consumed almost 50% of your inbox. It could ruin your office works. You need to junk them one by one, to remove them in your daily lists.



50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board



James M. Connolly, User Rank: Blogger  
5/22/2013 | 7:40:17 PM

Re: Junk

@ariella. There's real value in knowing when to leave an email thread and go to an inperson or phone conversation. In fact there are some things that are better off not done completely in email (personnel matters in particular). However, the people who want a record of the conversation aren't wrong. When you go to a personal or phone conversation for something sensitive, I'd argue that an email recap of the key points is valuable in getting both parties on the same page.



50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board



James M. Connolly, User Rank: Blogger  
5/22/2013 | 7:36:20 PM

Re: Junk

@Legalcio. I'm on record (and have the email to prove it) that I hate the idea of storing tons of stuff "just in case" it comes in handy down the road. I say that despite all the stuff clogging my toolshed and spare bedroom. That said, I'll make an exception and say that we should keep email "just in case". We don't have to run analytics against most of it, and there's good chance that those emails won't be called for in an electronic discovery order. But there's a chance that they will be called for. I don't want to be the exec who has to answer to a judge because a crucial email was tossed to save some disk space.



50% 50%

I can think of too many times when I've had to go way back in my email archive to find an item where an idea was proposed or rejected or a person said something that was contrary to what they later said. That's not even a legal matter, just corporate stuff.

Given today's storage costs, it's not even worth the labor cost of going through an doing a save/not save analysis for all email. Archive it, and have it available if you choose to run analytics against it if your lawyers come calling.

Reply | Post Message | Edit/Delete | Messages List | Start a Board

[More Video Blogs](#)

FOLLOW US ON TWITTER

**Big Data Republic** @BigDataRepublic 15 Apr  
AllJoyn: A Common Language For Internet Of Things - Open-source lingua franca aims to let IoT devices and services... [ow.ly/2FHbY7](#)  
Retweeted by Yakamoto  
Expand

**Hajime Ozaki** @hajozaki 22h  
BDR Top 20 #BigData David Smith David Smith List of top tweeters on Big Data from @BigDataRepublic [wp.me/p4Aavi-2](#)  
Show Summary

**Marcus Borba** @marcusborba 28 Apr  
#BigData #Analytics #BusinessIntelligence #FS @informationweek @AllAnalytics @BigDataRepublic @BigDataNetwork @BigDataClub

LIKE US ON FACEBOOK

**Big Data Republic**  
Like

4,168 people like Big Data Republic.

Facebook social plugin

ACCOLADES

Media Pioneer Award Winner



Debera Harward, User Rank: Exabyte Executive  
5/22/2013 | 5:50:16 PM

**It Can Be Junk For Someone And Treasure For Some One**

Almost 70 percent of the inbox usuall gets filled with the junk mails but the question is this that who sends these emails definitely these emails are the adds by the marketers . It ca be junk for you but it is not junk for them similarly you cant declare that these emails are junk on the contrary these emails are junk for you but they are not junk for everyone because if they would have been junk you wont have been receiving these mails on daily bais .

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



Ariella, User Rank: Blogger  
5/22/2013 | 2:25:07 PM

**Re: Junk**

<A lot of email could be eliminated by not using systems for personal communication, non business communication, or creating endless threads where a 5 minute phone call would suffice.> @legalcio I agree! But there are people who will avoid talking on the phone at all costs (I know some personally). Also some just like to have that digital record to have something or someone to point to if there is a problem down the road.

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



legalcio, User Rank: Exabyte Executive  
5/22/2013 | 10:41:36 AM

**Junk**

Most of it is junk. Mindless yes or no responses, threads about your kid's birthday party, laments about the final episode of The Office.

Then there is that one piece of mail containing a cryptic message about a package that needs delivery to some godforsaken part of the earth. Senses heightened by the scent of a big ol' smoking 357 Magnum, investigators pounce on the communication like an Apple junkie jonesing for the iPhone 6.

Which is probably why we need to keep everything when it comes to gov't communication and the CEO email from publicly traded companies. There's a reason e-discovery vendors are booming in the legal arena. Big Data can help target what is important in email communication, but it can't administer email best practices. A lot of email could be eliminated by not using systems for personal communication, non business communication, or creating endless threads where a 5 minute phone call would suffice.

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)

**BIG DATA REPUBLIC**

[ABOUT US](#) | [CONTACT US](#) | [HELP](#) | [REGISTER](#)  
[TWITTER](#) | [FACEBOOK](#) | [LINKED IN](#) | [GOOGLE+](#) | [RSS](#)



bigdatarepublic : /bigdatarepublic/section/2635 : /bigdatarepublic/section/2635/263442



UBM TECH

OUR MARKETS: [Business Technology](#) | [Electronics](#) | [Game & App Development](#)

Working With Us: [Advertising Contacts](#) | [Event Calendar](#) | [Tech Marketing Solutions](#) | [Corporate Site](#) | [Contact Us](#) / [Feedback](#)

[Terms of Service](#) | [Privacy Statement](#) | [Copyright © 2014 UBM Tech, All rights reserved](#)