

Reducing Internal Costs and External Fines of Email Compliance Using Text and Graph Analytics: A Q&A with David Gebala of Teradata Aster

By [Ron Powell](#)

Originally published September 14, 2015

This BeyeNETWORK article features Ron Powell's interview with David Gebala, senior business development manager of field applications for [Teradata Aster](#).

What is the 10-letter word that causes many corporate headaches? Compliance! With the continued proliferation of email and government regulations, non-compliance fines can be enormous and easily eat into a company's profits. Many companies have hired a significant number of people to address compliance issues, but often have experienced limited success and high costs. By utilizing text and graph analytics to monitor communications, a company can easily identify potential compliance issues, reduce compliance costs and get resolution in a very timely manner. Additionally, because government regulations that require compliance are increasing, the need for an effective compliance solution becomes even more critical. In my interview with David Gebala, we discuss how text and graph analytics for email compliance can effectively reduce costs, reduce fines, reduce compliance staffing needs and increase overall profitability.

David, let's talk about email compliance. Can you provide a high-level overview of what email compliance entails?

David Gebala: Email compliance is really just one aspect of our digital communications compliance offering, with email being the biggest one that everyone understands. It has really emerged most recently after the financial meltdown of 2008-2009. It has always existed, but it has been thrust to the forefront now, especially around deceptive or misleading advertising and the unsavory parts of the financial world. They're finding that the requirements for monitoring communications, especially with clients or even potential colleagues or M&A targets can be fraught with pitfalls that are costing financial institutions a lot of money – in some cases that means it's costing the taxpayers a lot of money. It's becoming a much bigger problem. One of the things that we often cite is that the ease, ubiquity and frequency of communications has been facilitated by digital platforms like the Internet, email, texting and social media platforms.

The potential for abuse has also grown. A lot of corporations recognize that this is a big area of liability that they need to address.

Are there any particular industries that need to pay more attention than others?

David Gebala: Most certainly. We recognize that there are many industries that have restrictions, and the restrictions range from self-policing to government-related and legally binding. For example, pharmaceutical companies can't make claims of drug efficacy without certain side-effect warnings. There are privacy concerns in many legal communications with customers. Of course, there is also the whole privacy situation with electronic medical records and communicating between the care provider and the patient. Those are also potentially very expensive compliance liabilities.

The biggest industry for us is financial services where the financial crisis, as I said, has really pushed that to the forefront.

You mentioned government regulations. When you look at organizations like FINRA and many of the big laws that have been passed recently in Congress, the liability to a lot of financial institutions is from fines, right?

David Gebala: That's right. That's exactly the unsettling trend that we are trying to address. It's not about generating more revenue. This is to keep you out of trouble. We actually have figures that illustrate how big of a problem this has become. It has been publically reported in the press that JPMorgan Chase, the biggest bank in the country, had \$36 billion in legal fees since the financial crisis on non-compliance issues. That's legal fees as well as direct punitive fees from the government that they've had to pay because of non-compliance issues. Those are things, like you said, FINRA, insider trading, and the Foreign Corrupt Practices Act. They can wade into all of those legal murky areas, and any employee can put them there just by emailing or otherwise communicating with someone. That can potentially cost the company millions of dollars per incident. On top of that, once a non-compliant event has been identified – for example if the current monitoring of their email system indicates that someone offered something of material value to their counterpart in China – they have to be able to produce that whole email thread within a certain amount of time or the bank will actually face more fees, not only for the actual occurrence but for not having the archive and for not delivering it to the government in a timely manner. There are multiple layers of costs that the banks are incurring so it is easy to understand why they are implementing solutions that give them a better means of addressing it, reporting it and being able to comply with the law.

So you can actually be fined for non-compliance of non-compliance?

David Gebala: That's right! And there are extremely high financial repercussions – the \$36 billion I mentioned earlier. JPMorgan Chase has an army of 2500 people in their surveillance unit. That's just for internal surveillance of communications and practices that could indicate bad behavior or non-compliant behavior amongst their employees all around the world.

How is email compliance typically addressed today?

David Gebala: It is normally addressed through a fairly simple process. All of our digital communications go through a communications server so the logs from those servers are preprocessed. This is how it is done now: they scrub out the attachments and then they look at the text of the message, and they apply rules against that message. The rules are where you embody the knowledge of what you're looking for. The rule has to be a little bit clever to discern that something non-compliant is happening. Insider trading is a good example. But if someone says, "I know this price is going to go through the roof after their announcement in 3 days," that would definitely flag a rule and highlight that message for review by an analyst at the end of the process. That's the simplistic view of how email is currently screened. It is basically taking the server copy of the message, stripping out the attachments, looking at the text of the message and determining if there is anything that violates the rules. That's the machine screening part. Then the message is passed on to someone to be reviewed. That process can produce a lot of false positives where the machine thinks it is wrong, but it really is not.

What are some of the problems with the current approach?

David Gebala: One glaring problem is that the process actually screens out the attachments, and sometimes as much as 70% of the message is in the document or documents enclosed. Obviously, a clever person could circumvent the screening process just by putting everything in a document, attaching it and sending it as a blank message. Then it would essentially be scrubbed out. That preprocessing generally is acknowledged to have a big blind spot that is not scrutinized.

Another problem is that the actual text of the message that is preserved in the current process is actually a lot bigger and a lot more redundant than it needs to be. The current solution doesn't use text analytics to whittle a message down to its core essence. What I mean is there are tons of signatures and disclaimers that are "canned," such as the writer's name, company, contact information, etc. That should be scrubbed out. In all the emails that are sent, even though it's a small percentage of the actual message, over all of the millions of messages it actually adds up to a lot of unnecessary text.

That's also true for disclaimers – the notations at the end of many emails. Those also have to be detected and screened out.

The third problem is echoes. Every time someone replies to all of the recipients in the email thread, it creates a lot of redundancy. When a company uses our text processing to clean that correctly and boil it down to just the relevant part that is new to the message, we've been able to reduce the amount of text stored by an order of magnitude – from 100x to 10x or even 1x of what it used to be. Because you're shoving through a lot of noise into the screening process and applying rules against it, every one of those messages gets flagged as potentially in violation. And then that results in the need for an army of analysts that have to actually go through and screen it. It's a lot of non-value-added work up and down the chain.

How does the Teradata Aster offering differ from the current practices?

David Gebala: The things that I've highlighted give you an indication of how our text analytics does a far superior job of scrubbing out all of those undesirable parts of the message. On the flip side, we actually run all of the email attachments through a screening so we don't have that blind spot. Because we use graph analytics, we can tell when a message is going out in bulk. Basically, if there is a huge amount of receivers on an email, it's probably not someone trying to do something non-compliant. Because we can identify that by understanding the graph of the network of where things are going and where they're coming from, we actually can apply rules to whittle down the amount of messages that have to be screened. And, as I said, we actually look at the attachments. On top of that, we also are able to feed back the actual non-compliant verified message rule type and update the rule set. The open-loop system we discussed before just goes through and screens against known rules, but we are able to update the rules in some continuous fashion over time.

You mentioned in the current practices the issue with regard to attachments. Does Teradata provide greater protection as well with attachments?

David Gebala: We actually process the attachments the same way that we are able to process the email message. To us it's all the same. We can process any digital file the same way against those rules. We actually screen the attachment for non-compliant behavior as well.

How does Teradata's email compliant solution coexist with legacy email monitoring?

David Gebala: It's actually the same process that we go through, but because we are able to scrub out using our advanced graph and text analytics, it actually feeds right into the existing screening, monitoring and verification infrastructure that most of these places have. We just allow it to be a much more efficient process, and we have some quantified examples. We took the same email archive from one particular bank and ran it through their process, and it ended up flagging 47% of the emails as being potentially non-compliant and in need of manual review. That exact same archive was run through the system with our analytics in place, and it actually reduced it down to just 4%. You can imagine the workload fatigue of the end-of-the-line analyst was greatly reduced and costs were much lower through using the intelligent system that removes the noise before you start screening and monitoring.

Wow – just the savings in people time is huge.

David Gebala: And also there is the – I don't know if you can quantify this – but a system that cries wolf that often is just not going to be believed. How useful is it when almost half of the messages are being flagged for review. That's just too much. I should point out the review revealed that the other 43% were just garbage – not truly non-compliant – because the screening process was not intelligent enough. So reducing that down from 47% to 4% is huge in confidence-building as well because we're reducing it to real threats, not just a bunch of noise.

David, it's been a pleasure to learn how email compliance using text and graph analytics is important for enterprises today.

About the Author



Ron Powell, an independent analyst and consultant, has an extensive technology background in business intelligence, analytics and data warehousing. In 2005, Ron founded the BeyeNETWORK, which was acquired by Tech Target in 2010. Prior to the founding of the BeyeNETWORK, Ron was cofounder, publisher and editorial director of DM Review (now Information Management). Ron also has a wealth of consulting expertise in business intelligence, business management and marketing. He may be contacted by email at rpowell@wi.rr.com.

*All Rights Reserved. Copyright 2004 — 2015, TechTarget
BeyeNETWORK™, a TechTarget company*