

The SOA Magazine

Feature Article



Viral Data in SOA: An Enterprise Pandemic

by Neal Fishman

Published: August 30, 2009 (SOA Magazine Issue XXXI: August 2009)

[Download PDF](#)

[Digg This](#) • [De.licio.us](#) • [Slashdot](#) • [Technorati](#) • [StumbleUpon](#) • [Google Bookmark](#)

Abstract: A services-based IT solution such as "A Single View of the Customer", is typically deployed for use as an enterprise-wide business application. The general intent is that any corporate application (such as an order entry system or billing) requiring the use of customer information can interact through a services layer to the single view in order to obtain de facto information for any particular customer.

Should any information within the single view prove to be incorrect, all subscribing applications would have had access to the same incorrect data at the same time. Further, should a service that writes or manipulates data in the single view contain a bug, any element that is touched could be compromised. In this way, data can be viewed as having viral properties. In an SOA environment, the effect can be akin to a pandemic-an enterprise pandemic. This article has been adapted from the book "Viral Data in SOA: An Enterprise Pandemic."

Introduction

In the book of Genesis, Adam and Eve consume fruit from the tree of the knowledge of good and evil. Although the type of fruit is not explicitly mentioned, the apple has survived as a popular hypothesis. In the centuries that followed the death of Jesus Christ, the Old and New Testaments were translated into Latin. In Latin, the word for evil (as in the tree of the knowledge of good and evil) is malum. As a Latin word, malum is a homonym. An alternative meaning is *apple*.

To venture into understanding meanings in communication or to delve into linguistics is to study semantics. What one person means [to say] and another person takes away [as being said] can represent a significant challenge with both verbal and written styles of communication.

Viral data is concerned not only with verbal and written communication, but the communication and linguistics involved in sharing data between software programs-especially, those deployed as part of a service-oriented architecture. Viral data is a metaphor used to indicate that business-oriented data can exhibit qualities of a specific type of human pathogen: the virus.

By Itself, Data is Inert

In humans, the classification scheme for disease-causing organisms fit into five categories: viruses, bacteria, protozoa, fungi, and worms. Viruses are responsible for many of the great plagues and pandemics and are the most simplistic of the listed pathogens and parasites.

The word *virus* is also Latin and literally means poison or poisonous slime. In biology, all living things are cellular with the exception of the virus. By itself, a virus is an inert particle-tiny and lifeless and requires a host (a cell) to become actionable. Like a virus, data can be thought as being inert. Data requires software such as a service (or people) for the data to appear alive (or actionable).

If a piece of data is never used by a service or seen by a person, the data cannot have a negative or infectious effect. When data is consumed by a service or appears on a screen that is scanned by a person, that data has entered an actionable world. The inert biological virus is skilled at getting into a cell-the virus' version of an actionable world. Once the virus is inside a cell, the game of infection begins.

A virus outside of a cell is not dangerous; likewise digital data outside of a program is not dangerous and neither is printed data that is not viewed. To be a pandemic requires more than just a widespread disease. Cancer is a deadly widespread disease, but cancer is not a pandemic because the disease is, generally believed, not to be infectious. To be classified as a pandemic, the disease (or condition) needs to be widespread and also, infectious.

Within the boundary of an enterprise, viral data can be pandemic when a service-oriented architecture achieves high-degrees of interoperability throughout the corporate value chain while leveraging synchronous data stores.

A generation for humans is measured approximately by a 30-year interval. For some viruses, the equivalent measure is less than five days. Furthermore, for some bacteria, the time span from one generation to the next can be as little as ten minutes. As for data traveling through the real-time enterprise, the generational time from one data store to the next can probably be measured in fractions of a second.

Stock exchanges around the world, now permit brokerages to co-locate their servers within the exchange data centers to help reduce latencies-if only by nanoseconds. "And latency affects more than execution; it also impacts prices and the distribution of data inside an enterprise." [REF-1]

A necessity for gaining the upper-hand on viral data is to scrap some of the data-oriented dogma that has existed since the late 1960s and rethink many of the precepts associated with data. Solutions for viral data in SOA may not be unilaterally preventative in all business situations, but the more one understands about the causes and what symptoms to look for, the better the chance an enterprise has to control its pathogens.

Establishing Truth

Folklore has it that there are three degrees of falsehood: first there is the fib, then the lie, and finally, there is statistics. Ordinarily, knowledge workers in a corporation do not seek to record or manufacture information that deviates from the truth. However, misinformation can systemically happen in a business-the viral data can enter through sheer carelessness, a business rule, or a host of different ways.

What is believed to be true can be true, or false, or both true and false at the same time. Again, provenance is a concept that can prove helpful in establishing truthfulness in data-trusted information.

The world of sports is big business and often relies on statistics to recount an event. In a Major League Baseball (MLB) game held on April 28, 2008, the Baltimore Orioles beat the Chicago White Sox 4-to-3. [REF-2] The winning run came in the 14th inning and the Orioles pitcher, Alberto Castillo, earned his first major league victory. In all MLB games, one of the pitchers on the winning side is given the victory for historical and statistical purposes.

Of particular note during the Orioles and White Sox game of April 28, 2008, was that Castillo had yet to start his MLB career. In fact, Castillo and a handful of other players were not on either team's rosters on that date. Rain had temporarily halted the game in April and the game was not restarted and concluded until August 25 in Baltimore. Officially, all the statistics for the game are attributed to April 28 in Chicago. As such, even in business, statistics can, on occasion, distort the truth.

Without the provenance to go along with the statistics, decisioning on the truth might become problematic. Statistics or other derived or aggregated data become part of the semantic landscape for communicating; whether verbally, in writing, or as a message exchanged between services.

Service-oriented architectures and cloud computing build on the concepts of distributed computing and modular programming. Services are often grouped as interoperable packages. In a general sense, interoperability in software is a "capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units." [REF-3] Essentially, this is an ability for two or more programs to exchange and use exchanged data.

During each interchange, data can be further manipulated, persisted, or decisioned on like data in any traditional programming environment. Services that are orchestrated to interoperate in a real-time mode are on the one-hand extending the capabilities that information technology departments can bring to the business and on the other hand, capable of causing unimpeded havoc.

The ability to transform disparate or standalone functionality into a seamless string of orchestrated business processes or to continually distribute the same data from a common data store to disparate processes, creates for viral data, a perfect storm, [REF-4] a perfect opportunity to miscommunicate with ubiquity and simultaneity—a service-oriented pandemic reaching all corners of the enterprise.

A starting point in the search to control viral data in SOA is to begin with the semantics associated with the lingua franca. A subject matter expert from the business may know that preferred customers are offered a standard 15 percent discount, or that Federal Express delivers all plastic kumquats from suppliers located in Tennessee, or that excessive humidity can cause quality problems on the production line, without knowing exactly why these facts are true.

From an epistemological point, a subject matter expert can know that a certain something is true without fully understanding why that certain something is true in the first place. The functioning enterprise naturally allows conflicts in its vocabulary and is generally capable of putting into place exception criteria to manage undesirable behavior on a whim without needing to fully comprehend the ramifications. But then, sometimes, what we believe to be true, turns out to be false.

For example, the Federal Reserve System is an intricate part of monetary policy and the money system. Some textbooks on economics suggest that the Federal Reserve System was created by Congress in 1913 as the central bank and monetary authority of the United States. [REF-5]

However, the Federal Reserve System is not federal and there are no reserves. Further, the Federal Reserve Banks are not even banks. The Federal Reserve System is a privately owned corporation and its ownership, in the U.S., has control over all things that deal with money. The initial plan for the Federal Reserve System was drafted at a secret meeting held in 1910 at the private resort of J. P. Morgan on Jekyll Island. At Jekyll Island, the meeting attendees conspired to: [REF-6]

- Stop the growing influence of small, rival banks.
- Make the money supply more available.
- Pool the meager reserves of the nation's banks into one large reserve.
- Should the cartelization approach lead to a collapse of the banking system, shift the losses from the owners to the taxpayers.

The Jekyll Island story and the factual formation of the Federal Reserve System is not as well known as its fictional tale. In business, some truths, for one reason or another, are hidden. Other truths are concealed because of a lack of information or because of the way the information is presented.

Implicit information hiding is why some people have suggested that PowerPoint® has become responsible for destroying our ability to adequately communicate. Professor Edward Tufte believes that PowerPoint played a significant role in the 2003 Columbia space shuttle tragedy. [REF-7]

From school, to government, to work, speaking in abbreviated sound bites and bullet points, has become a natural way for us to communicate both simple and complex issues. In a story about the Iraqi war, General Tommy Franks remarked that "It's quite frustrating the way this works, but the way we do things nowadays is combatant commanders brief their products in PowerPoint up in Washington to the Office of the Secretary of Defense and the Secretary of Defense... In lieu of an order, or a fragmentary order, or plan, you get a set of PowerPoint slides... [T]hat is frustrating, because nobody wants to plan against PowerPoint slides." [REF-8]

Furthermore, Colonel Andrew Bacevich has quipped, "To imagine that PowerPoint slides can substitute for such means is really the height of recklessness." [REF-9] As an analogy, imagine an automobile mechanic that uses a manufacturer's glossy sales brochure in order to figure out how to repair an engine.

Consider observing a trend in services deployed for use in business intelligence that helps push knowledge through a series of dashboards. One wonders if the brevity and semantic stumbling blocks of PowerPoint will be repeated in the use of management dashboards.

What this means and what that means; trying to sort out what information is important and what information possibly represents excessive noise plays into a culture faced with information overload. Semantically, we can pose a question, what is a book? On the surface this benign question may seem like it should have a simple answer. The first edition of the book *Viral Data in SOA* has an ISBN, but not just one:

ISBN-10: 0-13-700180-0

ISBN-10: 0-13-703565-9

ISBN-13: 978-0-13-700180-4

ISBN-13: 978-0-13-703565-6

Answering the question, what are the royalties due to the author? can pose a difficult semantic question depending on how the question is asked, how and where the information is stored, who has access to what information, and what, if any, assumptions were implied in the way the question was initially phrased.

If establishing meaning can be difficult even when the information present is correct, the latency involved in uncovering incorrect information can be outright painful. A recent mistake in assessing the property value on a single house in Valparaiso, Indiana, caused the county's computers to automatically increase the property's tax liability. By the time the typo was discovered, the Valparaiso school district, and government agencies faced a financial shortfall and were forced to cutback budgets by \$3.1 million. [REF-10]

Trying to Physically Align IT to the Business Can Result in Shortsightedness

The International Standards Organization (ISO) established the ISO 4217 standard as a means to manage currency codification. For example, the ISO symbol for U.S. dollars is USD and the Danish krone is DKK. Listed among the ISO currency codes is the codified value XXX. Although XXX is a valid code value within the standard, XXX does not actually represent the currency of any nation. Additionally, the code value XTS is reserved for testing, while XAU and XAG are codes for precious metals. Codes CLF, USN, USS as well as a number of other codes are provided for bonds and other fund types. At the semantic level, ISO 4217 covers more than currencies.

Many companies conduct business in England. However, England is not a sovereign country. The sovereign country of which England is a part of is the United Kingdom of Great Britain and Northern Ireland. However, England, as well as Scotland, Wales, and Northern Ireland are regarded as countries when participating in sporting events such as soccer, rugby, and cricket. This is not the case in all sporting events. For example, for the Olympics, the United Kingdom enters a team under the moniker GB (Great Britain).

The latitude afforded the United Kingdom would be equivalent to Germany entering teams from East Germany, West Germany, and Bavaria, or the United States of America entering teams from the Union, Confederacy, and Texas.

Any movement toward an IT solution that overtly aligns itself with the business unnecessarily constrains the business and, over time, may ultimately result in promoting further misalignment. By recasting the desire to align as an ability to continually react, may begin to get closer to the type of relationship information technology needs with the business. Reacting to temporal needs of a business requires adaptive, accurate, and timely solutions. Pro forma costs associated with reacting must also be cost-efficient. Overall, the ability to react is much better than the ability to establish an alignment. Organizing to be reactive poses a far loftier and prudent goal for information technology.

The paradigm associated with SOA is extremely important because its architecture can readily be deployed to react. The innate ability to compose services of different levels of granularity and externalizing the orchestration of how those services interplay are also vital ingredients to create form-trusted information-an antibody to viral data.

Conclusion

Overall, viral data in SOA has the capacity to become an enterprise pandemic and disable a company. Service-oriented solutions that incorporate interoperability, reusability, layering of abstractions, and loose coupling can serve as perfect hosts to propagate misinformation: That is the knife's edge of SOA.

Whether viral data in a services-oriented solution is being created, discovered, remediated, conditioned, or

inoculated, chances are that representatives from all facets of the enterprise are involved. Those included may be business personnel and subject matter experts as well as participants in a governance body, architects, analysts, developers, and database administrators. Crafting trustworthy information necessitates a coordinated common cause across each of these specialized communities.

References

- [REF-1] Crosman, Penny. Lehman, *NYSE, CME, Forex, Capital Pursue New Latency Killers*. Wall Street & Technology, January 2008.
- [REF-2] Viera, Mark. *For O's, April Showers Bring Split in August*. Washington Post, August 2008.
- [REF-3] TC: Technical Committee ISO/IEC JTC1. *ISO/IEC JTC1 SC36 WG4 N0070*. Geneva, Switzerland: International Standards Organization, 2003.
- [REF-4] A *perfect storm* connotes a particularly bad situation arising from a large number of negative and, often, unpredictable contributory factors. As a metaphor, a perfect storm is normally attributed to a fierce storm arising from a rare combination of adverse meteorological factors.
- [REF-5] Baumol, William, Alan Blinder. *Microeconomics: Principles and Policy*. Mason, OH: South Western, 2009.
- [REF-6] Griffin, Edward. *The Creature from Jekyll Island: A second look at the Federal Reserve*. Westlake Village, CA: American Media, 2002.
- [REF-7] Pigeon, Steven. *The Devolution of Communication*. Milwaukee, WI: Milwaukee Journal Sentinel, September 13, 2007.
- [REF-8] Ricks, Thomas. *Fiasco: The American Military Adventure in Iraq*. London, England: The Penguin Press, 2006. [REF-9] Ricks, Thomas. *Fiasco: The American Military Adventure in Iraq*. London, England: The Penguin Press, 2006. [REF-10] Ruethling, Gretchen. *A One-House, \$400 Million Bubble Goes Pop*. The New York Times, February 15, 2006.

THE PRENTICE HALL SERVICE-ORIENTED COMPUTING SERIES FROM THOMAS ERL



[Home](#) [Past Issues](#) [Contributors](#) [What is SOA?](#) [SOA Glossary](#) [SOA School](#) [SOA Books](#) [About/RSS](#)
[Legal](#)

Copyright © 2006-2009
SOA Systems Inc.