

**Keynote address by Don Donahue, President and CEO,  
The Depository Trust & Clearing Corporation  
Insurance Data and Analytics Summit  
The Harmonie Club, 4 E. 60th St., New York, NY  
April 25, 2012**

---

## **The Gold in the Data Hills: Mining it, Refining it and Using it**

Good morning, everyone. It's a pleasure to be here. I'd say it actually feels a little like a family reunion. At DTCC, we know almost all our cousins on the life, retirement and annuity side of the business.

In fact, for the last 15 years we've been working closely with them. Our Insurance & Retirement Services unit offers a set of services that help foster growth, reduce paper flow and lower costs by bringing the scalability and efficiency of automation to the markets and to our clients. Those of you who know our business model know that our clients are also our members and frequently our shareholders, and they represent the whole range of companies serving the life and retirement insurance industry... including carriers, distributor broker/dealers, vendors, clearing firms, and third-party solution providers.

Now today, I have the opportunity to meet the other cousins—your side of the business. So I'm delighted to be here.

The topic on the table this morning is data: mining it, refining it and using it. The insurance industry, of course, has always embraced the most advanced ways of working with data. That's why you have organizations like ACORD. Data is your bread and butter. It's what actuaries feed on. They're voracious. They're the original "data-vores". And today they use data not just to measure and calculate risk, but to define market segments, gauge returns on investment and monitor leverage.

## **The Moneyball Factor**

In the last few years, however, much of the rest of the business world has also become a lot more aware of—and savvy about—the value of data. When Hollywood starts making movies like Moneyball, which, as you probably know, is about the business of refining and using data to get a competitive advantage in the world of professional baseball, it's clear that data is the new gold. What happens in Moneyball is that the quants in the back room, rather than the old-line managers, start calling the shots based solely on data—and the team improves dramatically.

One reason there's more interest today in analyzing data is probably that there is simply so much more data available. As you know, because your business is data-based, we're witnessing a huge explosion in data collection. Courtesy of today's technology, we are able to collect and store more data today than ever...vastly more. We have it in such volume that we can group it into "meta data" or "big data".

We sit on a wealth of data about everything from financial transactions and exposure to casualty rates...to weather patterns, property values, credit histories, coffee consumption, Facebook traffic, and so on and so on. The question is how

to use all this data to improve the information flow to and from our customers, our investors, our markets, our regulators and the broader marketplace. How can we use the data to help understand the kinds of risks that are evolving in our markets? What kinds of new products and services can we offer our clients based on the data to help them in their own businesses and activities?

## Data Explosion

These questions seem straightforward enough, but the problem is that mining our growing stockpile of data is becoming far more complex. About two years ago, an article in the Economist talked about the rate at which data bases are growing. We're all familiar with megabytes and gigabytes as measures of quantities of data, and I'm sure many of you know what a terabyte is. But how many of you know that a petabyte is a thousand terabytes? Or that an exabyte is a thousand petabytes, or a million terabytes? Let's go one more – a zettabyte is a thousand exabytes, or a one followed by 21 zeros. The Economist article estimated that the total amount of data in existence in the globe at that time – two years ago – was roughly 1.2 zettabytes, a staggering amount.

For those of us who don't think in terms of zettabytes, perhaps a more understandable way to grasp the challenge is that the data that represents was, at that time, estimated to be about **67 million** times the contents of all the books catalogued in the U.S. Library of Congress. Fast forward to the end of this decade and, according to IBM Chairman Sam Palmisano, the data we will have accumulated will rise to 35 zettabytes – or several hundreds of millions of times the entire contents of the Library of Congress. We will be vacuuming up everything in sight.

But data without understanding, data without analysis, is just raw material. To make data useful the way the Brad Pitt character uses it in Moneyball to remake a team, the data has to be analyzed, thought through, and turned into information – with intelligence about what it means and what conclusions it suggests. The sheer volume of data now accumulating, and the sheer amount of detail needed to understand today's financial markets and instruments, represent a real burden on and challenge for us and our clients.

Turning data into information, and finding ways to help our clients and their customers understand what the data means and how to use it, is becoming a major adjunct to our business. In fact, while DTCC has always been a transactions business, the value we add now lies increasingly in how we leverage the data we derive from those transactions... and in how we mutualize access to and the costs of core data bases in order to deliver to our clients market intelligence, risk control, compliance solutions and cost reductions.

So what I'd like to do today is talk about the role we play as an infrastructure organization in helping the financial industry to handle its massive data management challenge. I will focus on the kinds of data we tend to gather, how we standardize its collection and transmission, and how we capture and warehouse this data in ways that serve the entire financial services industry spectrum, including your companies, not to mention regulatory agencies. As today's conference suggests, we're hardly the only business working harder and harder to capture and digest data. In today's world, accepting that challenge is a given. But perhaps some of our experience will give you food for thought as you ratchet up data capture, analysis and distribution in your organizations.

## A Look at DTCC

DTCC is structured somewhat like a mutual insurance company. We're owned by our members, and we operate on an at-cost basis. We're also the largest post-trade infrastructure company in the world for the post-trade processing of transactions in securities and, increasingly, other types of financial assets. As soon as trades are confirmed, the trading markets step out of the picture. They leave all the tedious bookkeeping and all the transfer of money and securities up to us. Transaction data flows into our systems from more than 50 different trading platforms for equity securities alone, as well as from thousands of participants in the over-the-counter markets. We provide services that cover just about every asset class...from equities and government securities...to over-the-counter derivatives and money market instruments...to mutual funds and mortgage-backed securities. We are, in essence, a huge data processing business involving the safe transfer of a trillion or more dollars in securities and funds under tight deadlines every day. Messaging traffic

alone on our systems runs into the many millions a day, and our data bases answer queries more than 2.4 billion times each day...or about 28 times a second. That's a lot communication involving a lot of data.

Even so, we are not the largest data processing business. Visa, for example, handles more transactions per day than we do. There's a difference, though. If you buy a new TV from Amazon and put it on your credit card, Visa is not responsible for paying Amazon the very next day, nor is it involved in the process of getting the TV delivered to you. If you buy a million dollars' worth of government securities, however, we're responsible for making sure payment gets made the very next day, and for recording the securities in your account simultaneously. Not only that, we guarantee the transaction will be completed.

Last year, we settled securities transactions valued at just about 1.7 quadrillion U.S. dollars. Put another way, we turn over an amount equal to the entire, annual U.S. GDP just about every three days. In the course of doing all this, we gather vast amounts of market and transaction data, and then we store it, massage it, standardize it, and maintain distribution networks so that our clients can access the data for processing, networking, compliance and other purposes. It is, after all, *their* data. On behalf of our clients, in certain circumstances we also make data available to regulatory agencies, both domestic and international.

A key concept behind this approach is that we're able to centralize access to critical data for our clients around the world. In effect, we operate like a data "storage cloud," providing our clients on-demand access to this shared resource. As a result, our data management tends to be more cost-effective because we "mutualize" the expense of the data base all across the industry and reduce the complexities for firms dealing with commercial data providers.

## **DTCC Analytic Reporting for Annuities**

Frankly, there's nobody in the industry positioned like DTCC Insurance and Retirement Services to extract and provide metrics to the market based on actual transactional flows. We currently receive and process some 2.4 billion reports on annuity contract positions a year, plus millions of other kinds of insurance and annuity transactions, which means we have at our fingertips a huge amount of insurance industry information and market intelligence.

With the baby boomer population aging, annuity product sales are increasing. To remain competitive in this growing business, companies need more sophisticated and robust market analytics. As one example of what we're able to do with our mountain of data, we recognized the opportunity we had to put an innovative, elegant and easy-to-use client-interfacing "wrapper" on the vast amount of annuities data aggregated through our daily processing routines, and to make this wrapper available specifically for our insurance clients.

Last year we launched DTCC's Analytic Reporting Service for Annuities as a fully hosted, web-based solution – available online anywhere, anytime – to provide accurate, timely data and much desired transparency to the annuity businesses. The service, naturally, also includes insight on financial inflows and outflows that are processed through DTCC. Because the data in our Analytic Reporting Service is based on actual transactions – specifically, the millions of annuity transactions centrally processed by DTCC (39 million last year) – firms can make more informed business decisions and identify key trends about sales, sales effectiveness, product management, marketing and compliance.

Analytic Reporting provides insight on information such as:

- Cash flows by product, product type, carrier and distributing broker/dealer
- Market shares
- Fastest growing products
- Fastest growing client types, and
- Detailed trends by carrier, distributor, product, account type, and more.

Clients can also rank and benchmark their business, comparing information and trends from their own transactions with the aggregated data of all participants, which can help firms understand similarities and differences, and identify further opportunities. Data is refreshed on a monthly basis, and is generally available in the middle of the following

month – a turnaround time about one sixth or less of the reporting lags of three months or longer that the industry used to endure before we started the service.

## **Compliance Solutions in the Data**

This approach – centralizing, consolidating and helping to interpret the data – can also be particularly effective for clients who need ready access to satisfy more stringent regulatory and compliance requirements. As you are all well-aware, insurance firms face a growing number of regulatory requirements at both the state and federal levels.

Like nearly all the financial service industry, the insurance industry needs to prepare for what this escalation of regulation will mean to its business. We all need to understand how to manage these new regulatory expectations, and firms need to ensure their sales and services fully comply with the regulatory changes sweeping the industry.

As we know from past experience, mitigating risk and increasing transparency are critical for sustainable growth in today's more demanding regulatory environment. And we believe that bringing greater automation and standardization to the insurance industry is perhaps the best way to adapt to the ever-changing regulatory landscape. When the cost of technology development is weighed against the potential cost-savings of automation, firms need to think strategically over the longer term. Cost savings from automation now can generate more capital for future technology development.

Just as the initial release of the Analytics Reporting service was developed with the active input of an advisory group of our member firms, we currently have a working group of firms continuing on the project. They're tasked with thinking through how the Analytic Reporting service might be further enhanced beyond the research and marketing intelligence to also provide the kind of in-depth transactional reports that are required by industry regulators.

This is a real paradigm shift in business analytics and data: clients do not have to store or manage the data, nor do they have to develop applications. As a neutral, industry-owned compiler of insurance data, DTCC is the source, the host and virtually, the data cloud.

In our role as an infrastructure organization, we have been tackling two other data initiatives intended to create more uniform and globally recognized data bases or approaches to data. Both of these will have an impact not only on securities firms but eventually on many companies throughout the world, including virtually all public companies.

## **Legal Entity Identifiers**

The first of these initiatives reflects the growing, global consensus on the need to create a worldwide, uniform, system to identify companies active in the financial markets – in industry jargon, a “legal entity identification” system. This idea for a global system came up, as you can guess, in the aftermath of the 2008 market meltdown. Regulators found they couldn't consistently identify the counterparties across the various global payment and settlement systems, and so they couldn't really evaluate the nature of a particular financial counterparty's global exposures.

Standards for identifying counterparties are, of course, not a new idea. But the crisis forced regulators to the conclusion that there was an urgent need for a single global standard legal entity identifier – an “LEI” for short – and that this was a job for a global utility like DTCC. Quite simply, the regulatory community realized it can no longer afford to tolerate a hodgepodge of proprietary and third-party codes to identify counterparties in individual payment or settlement systems. In order to monitor systemic risk across global markets, regulators need to match up what the counterparties are doing across these systems. They need to be able to match the company in column A with the company in column B.

U.S. regulatory authorities initially pushed this idea, but it soon caught on elsewhere. Jean Claude Trichet, then the head of the European Central Bank, endorsed the idea, and more recently, the Bank of England also said it thought this was a much-needed innovation. In fact, moving ahead with a system for identifying legal entities is an agenda item for the G-20 meeting in Mexico in June.

In preparation for the mandate we anticipate will come out of that meeting, the industry has already done a great deal of organizing to create a consortium of entities to manage the process of building and managing an LEI database. Centralizing the creation and on-going operation of this kind of system is a huge task – something you'd need a global utility to take on. Given our at-cost business model at DTCC, and our ownership of Avox, a global leader in company reference data, we were tapped to lead this effort. The job, however, will be done by a consortium that includes DTCC as the facilities manager to maintain the primary database, plus the Belgian-based banking cooperative SWIFT and other parties throughout the world to help validate and federate some of the information going into the central utility.

At this point, there's even a draft international standard which calls for the use of a 20-digit alphanumeric code as the legal entity identification number. A common identifier would not only enable systemic risk analysis, it also reduces the need for you to commit resources towards creating and maintaining your own, internal identifiers of companies or firms that you transact business with.

Assuming the G-20 endorses this approach – and I'm confident they will – I expect the industry, in partnership with the regulatory community, will move forward to implement a vast database of legal entity identifiers, a data base that will include most of your companies as well. And when that database is a reality, we'll wonder how we ever managed to get along without it for so many years.

## **Talking in XBRL**

Meanwhile, even where we have accumulated data – and we've all accumulated vast amounts – one of the problems is that nearly all of it is “unstructured”. Maybe only five percent is in a format that permits a computer to analyze it. And when the data comes in massive amounts that involve urgent matters, the need to have it a structured format so that real communication can take place is very acute.

One of the best ways to do this, we think, is through the use of XBRL – or eXtensible Business Reporting Language. With XBRL, data is no longer simply a block of text or numbers; it is expressed in a form that is readable, searchable, moveable, and extractable for easier analysis and reporting of information. As you probably know, some 8,000 public companies and another 8,000 banking institutions use XBRL today to submit their financial statements to the SEC and the Federal Financial Institutions Examination Council (FFIEC). Probably all of your companies have to file your quarterly and annual reports that way. Working with financial statements formatted in XBRL makes it easy to compare different company financial reports. It makes the data much more manageable and useful.

When data involves something time-critical – all of the financial flows behind a complex new security, for example, or the announcement of a merger between two companies – all the more reason why we need to be very active in promoting the much wider use of “structured-form” data in order to communicate. If a company whose properties you have insured announces that it is shedding some of its subsidiaries, that's material information for you. The sooner you know that, the faster you can act to react to this change in your business. But if the announcement isn't made in XBRL, you might not get that critical piece of business information in time, or understand it clearly enough, to protect your interests properly. Simply comparing and scrubbing the data in the announcements for accuracy is a major and expensive undertaking.

The solution to this – and it's a solution we've been pushing on U.S. industry for several years now is to communicate these announcements in XBRL. This approach, I'm pleased to note, is finally gaining some traction, and the securities industry is eager to see this happen. It knows the dangers that can come from misreading corporate action announcements that typically come flowing into the market from many different sources.

Our aim is that, ultimately, when issuers announce a corporate action using XBRL, the information can be immediately reconfigured into messages that can then be distributed and understood worldwide. If we can achieve that, I'm confident it will drive significant changes in how a major part of the business of the financial industry works worldwide. Not only will it cut risks, it will cut costs, and it will certainly cut headaches.

## Letting in the Sunshine

Cutting risks and costs was the driving force behind another major data control project we undertook for the securities industry. Ultimately, transparency also became a key issue.

The project involves OTC credit default swaps, or CDS. About nine years ago, our members first came to us for help in clearing up their huge backlog of unconfirmed OTC credit default swap contracts. Since standardizing and automating manual processes is at the core of much of what we do, we looked around to see how we could address the problem and realized we were able to repurpose some existing software for this task. By mid-decade, we had cleaned up a multi-billion-dollar contract backlog and automated the whole confirmation process.

Guess what happened next? Trading volumes that had been perhaps one or two thousand tickets a day zoomed to five and ten times that amount, as automation made the market far more scalable. But this innovation led to another. As we cleared up these backlogs, the logical next question was what to do with the contract details and where to store them. Since we operate a securities depository, it didn't take long to realize that creating a depository-like registry to store and service these contracts was the answer.

So we built what we called the Trade Information Warehouse, loaded all the outstanding and new CDS contracts into it, and quickly found out how valuable it was. The global industry now relies on our Warehouse records to handle a broad range of downstream activities, and the major operational risks associated with these markets ten years ago have been almost entirely resolved. The interesting thing about this activity is how clearly it illustrates the growth of a truly global marketplace.

Today, our Warehouse holds records on more than 2.25 million credit default swap contracts worldwide. These contracts involve some 1,700 financial institutions. The counterparties to the contracts operate out of 62 countries around the world, doing business with each other and supported by a global infrastructure exactly as if they were just across the street. The CDS contracts they trade involve assets originating from more than 90 countries. So it is literally the case that someone sitting in Europe will routinely trade with someone sitting multiple time zones away in Hong Kong in an asset that originates from the Americas. And all that activity is recorded in and supported by a global infrastructure.

It didn't take long for regulators to realize this was also an answer to their prayers. Instead of chasing down buyers and sellers across the world for records of various transactions, regulators began to look for more and more access into our warehouse files. To accommodate them, we not only make new data public every week, but have now built a special "regulators' portal" that gives regulators online access to CDS records involving the firms and jurisdictions they regulate. We already have about forty different regulatory authorities in the Americas, Europe and Asia using the portal. We let the sunshine in.

Since constructing the first global trade repository, we've been asked by the industry, on the basis of competitive bids, to build four more for other types of instruments. We have had repositories in operation in the U.K. for equity swaps since 2010 and for interest rate swaps since last year; just ten days ago we began taking the first contracts into our new Netherlands-based repository for commodity swaps, which will be fully operational later this year. In addition, we are now building a repository for foreign exchange swaps, which will complete the picture. Since we're dealing with global markets, we recognize that these need to be supported by a global infrastructure that can support appropriate public disclosure of critical market data and give the appropriate regulatory authorities access to the records they require.

The moral of the story for all of us here today, however, is that what we at DTCC started as a way to codify and store transaction data for one segment of the industry has blossomed into a global data business. And the more we learn about and operate this data business, the better we become at it, which means we hope to be able to apply the lessons learned here to yet other areas.

So I would say to you that, in a world where data is accumulating as fast as snow in a blizzard, certainly one way to cope is to make use of infrastructure organizations like DTCC and their large-scale, centralized approach to data aggregation, sharing and storage. You don't have to mine every vein yourself. We've already doing much of that work. And with your direction and encouragement, we'll do more.

Thank you.