Directors and Data Rich Jones

August 2022

By Rich Jones

Rich Jones 00:00
So today we're going to be talking a little bit about data governance. And I know as directors that you've got a lot of governance issues that you have to face. But I think data is one of the critical areas that you really need to kind of add to your laundry list of things that you need to be governing as directors. And there's five things I'd like to address around data governance, that that you need to take into consideration.

Rich Jones 00:39
And understand, number one, that there's a tendency to kind of do software to solve data problems, buy a data warehouse, you know, buy all kinds of applications, spend money, hire a lot of people, and I've seen all of those tactics take place. But what I want you to kind of keep in mind that really, there should be a data strategy, before you even get to software.

Rich Jones 01:05
You don't really want to buy software til you know you're going how you're going to use it. You don't really want to buy software until you know what data you're going to have to access and how you're going to access it. You don't want to hire staff until you kind of know what the long-term strategy is, on how you're going to use that data organizationally, across the board.

Rich Jones 01:28
The next element is: data lies. And we have to remember that, that just because you see a report doesn't necessarily tell you the truth. There's a lot of variables, especially in credit union data about what is the real truth. And that those lies come from a couple of different elements, No. 1, and we'll talk about that in a minute.

Rich Jones 01:54
There's a lot of sources of data within the credit union. The credit union holds a lot of data about their members, about their products, about their services, about transactions, how the members are using their money, how they're spending their money, how they're borrowing their money. There is a lot of data there. And not all of these applications tell the story the same way. There are three things that cause data to go bad.

Rich Jones 02:20
No. 1 is definitions. How you define the object of the data is going to tell you exactly what you're looking for. But the problem is these different applications have a different label for the same object. For

example, one application may call a member a member, another application may call a member an account. And how do you reconcile that semantic difference between what this application is calling a member and what this application is calling a member and what you know to be a member? So those definitions are really critical. So definitions is one.

Rich Jones 03:06
No. 2 is queries. The logic within a query is going to tell you what kind of data is coming out that's less than or greater than, equal to those kind of logic points. If there's any variation, those large points, then the data is going to tell a different story. And we need to make sure that we understand that the quality of the query is just as important as the quality of the definition.

Rich Jones 03:35
And the third thing that can make data go bad and cause it to lie is timing. I already mentioned that there's several applications that you get data from, well, all of those typically run on a different cycle monthly cycle. And when that your file gets updated from that third-party file is going to dictate what the data is saying at that point in time. And that data point in time may be different from one application to another. And so, if the timing is off, if the query is different, or if there's a difference in definitions, the data is going to lie. So all those have to be synchronized and have to be acknowledged as equal.

Rich Jones 04:20
The third element that we need to really understand is, and I alluded this earlier, there are up to 50 separate data sources outside of the credit union core operating system that we get data from. Just think about it. Your bill-pay, your peer-to-peer payment system, your mobile banking, your your online loan origination systems. All of these are data sources, and they all have different rules, but also they're accessed differently. Some of them are very easy downloads that automatically go into core and update it but those are rare. Some of them, there's a PDF that sent some you have to go into a third-party website and extract the data. So how do you do this without being very complicated work, and without taking lots and lots of time? And it gets especially difficult when you're trying to compile a report from multiple data sources because you have to go to each one independently and individually, grab that, and aggregate those reports together. And you've got to normalize the data so that the definitions are matching, and so forth. So it gets to be very difficult. So keep in mind, that data is coming from a lot of sources. And they all manage data different differently, that they may have all the same rows and columns but they're going to label those rows and columns differently. That complicates the aggregation of data.

Rich Jones 05:58
The next element are the queries that I talked about earlier. Queries, queries everywhere. Without query control, then people can write reports that will tell you what they want you to hear. Yeah, it can be done. In fact, I've seen it done, I've seen where there was a lending officer in a credit union,who didn't want to get it let anybody into their data set. Because that that executive wanted to control the message completely, I think you can see as a director with the risk of that is. If they can manipulate data, you've heard it before, statistics can tell you anything you want them to tell you, depending on how you build that statistical file, the same thing as with data and reporting.

Rich Jones 06:52
So there has to be some kind of quality control on queries. And there has to be some third-party oversight over those queries. Because if you don't have that, then you can never really trust that the data that you're seeing in that report, the information that reporters giving you is actually accurate, and truthful. Not saying that anybody is intentionally lying to you. But sometimes they can build a report that tells a story that is a little more friendly for them than what they they initially been really what the data may say. What you're looking for here is a query that really tells you that one source of truth.

Rich Jones 07:35
The next element: It's not about spending more money. As matter of fact, there's a lot of efficiencies that can can be gained, if in fact, you strategically face data, instead of tactically face data. What I mean by that is, right now, if you go in and you do a survey of your different departments that are building reports, you're going to find a significant amount of of human time being invested in doing nothing more than extracting that data out of source sources, compiling that data, trying to clean up that data, aggregating that data, and then formulating all that into a report. The time consumption is amazing.

Rich Jones 08:31
There was one credit union that I worked with a couple of years ago that had a VP that was manually taking PDF file information, and manually inputting that into an Excel spreadsheet to build a report. Now think about that. By her own admission, she was spending over 10 hours a week in doing nothing but manual input of PDF info into an Excel spreadsheet. If that could be efficient, if that extraction could be made automatic, and could go into a data warehouse or a data lake where it could be easily grabbed without having to go through that manual input, it would have saved them a huge amount of payroll expense and benefit expense.

Rich Jones 09:23
So think about it's not just about spending more money. It's about why you're doing it finding those efficiencies. What are the data pain points that the organization is suffering from? And how can we, in fact, make those data pain points less of a pain?

09:42
I'll tell you what, right now I know credit unions across the country have data analysts or financial analysts spend more time in building reports, aggregating files, and putting things together than they do in analyzing the data. What you want is an analyst to really look at the data and say, what is the data telling me? How can I share this with the executive team, so they can make better decisions?

10:08
It's not about spending more money. But what it is and where this fits for the board of directors is they need to understand just enough about this to be able to make sure that management is doing their job to clean up those reports, make sure that the queries are correct, make sure that the information you're getting is correct.

10:33

So there needs to be a little bit of like governance oversight over what they are doing, to confirm and verify that the data that you're seeing in all of your monthly meeting reports and so forth, is accurate and telling the truth that you want to see so that you are in a position of strategically making the right decisions for the organization.