Business Confidence: Trust

Event ID: 152908

Good afternoon and welcome to today's webinar, "How to deliver information your users can trust." I'm Mary [Windeshar], and before we begin here are a few brief announcements. This webinar is being recorded and will be available at www.ibm.com/cognos/techtalk. For an enhanced view of our presentation please disable any popup blockers you may have installed, then click the Enlarge Slides button located just underneath the slide window at the bottom of your screen. You can download a PDF version of the slide presentation by clicking the Download Slides button located underneath the slide window. Our presenters welcome your questions which may be submitted at any time during the webinar. To submit a question, please type it in the question box located in the lower left of your console and then hit the submit button. We will answer as many as possible during the final fifteen minutes of this hour long webcast. Any remaining questions will be answered via follow-up emails.

Now on to the webinar, "How to deliver information your users can trust." Speaking today is Becky Smith who is with the IBM Software Group information management team based in the UK. Her responsibilities include product marketing for the IBM Cognos Business Intelligence platform. Joining Becky is another UK team member, Steve Totman, who is responsible for foundation tools and industry models. And with us from the east coast of the United States is IBM Software Group technical specialist, Ernie Ostic.

Our first speaker today is Becky Smith. Welcome Becky.

Becky Smith: Hi, thank you Mary. So our objectives today – we are really going to walk through and share with you two key "How-Tos" to help you improve business confidence in information. The first one we are going to cover is around how you can build your business confidence by increasing information understanding through context and consistency of usage. So really giving better meaning to terms, better understanding to terms for business users.

And then secondly we are going to look at information trust and how you can increase business confidence by providing visibility into the quality of the data that is being consumed and how you can look at improving that quality over time.

So, to start this off, kind of one of the big questions is how much confidence do you have in your information that you are putting out to your business consumers through your BI system. And through a number of industry studies and IBM studies and also from customer interviews and feedback, there is some fairly startling statistics. From over half of users don't have confidence in the information they are using. And a lot of users, nearly 60% are saying that they miss information that might be of value to their jobs because they can't actually find the information they are looking for.

Over a quarter of manager's time is spent searching for the right information. And half of the information they do actually find has no value to them. And I am sure you are all familiar with the scenario of management meetings where everyone is turning up with numbers, the revenue margin for example, and the problem is the numbers are all slightly different because they may come from slightly different sources.

So we are going to examine today how we can actually address some of these problems and make sure that we can increase the confidence in information.

So, the big question is that we hear from organizations when they are looking and what they are trying to achieve is how can they ensure that everyone in the business is more informed, engaged and aligned in the decision making process so they can drive better business outcomes. And we see that this falls really into two distinct areas. Firstly the business challenge where we know people and IT are looking to address the diverse needs in the business making sure that the relevant information is provided to all users how, when and where they need it. So it is in the right sort of consumable form and giving the right information to the users. And Steve, can you just talk us through the system challenge?

Steve Totman: Okay, so on the system side, because these companies have grown through acquisitions and just natural growth, you end up with lots of fragmented silos. And you have all of these individual teams concentrating on point projects. So data ends up coming from all sorts of places and you have got across that heterogeneous enterprise, so you are going to have legacy systems like mainframes of (inaudible), you are going to have relational sources. You are also going to have applications with transactional data, so not just data coming in once a day through a data warehouse but real live transaction messaging over a message queue in formats like XML. So you have this massive diversity of sources, huge volumes – you are not just getting the batch but having to do with real-time as well. And the key is to make sure that the business gets the information they want and they need it structured in a way that they can easily consume it.

So you don't want to just present it as it is in the source systems. You want to get that 360 degree view of the customer and take out the critical things that the business is after.

Becky Smith: Perfect, thanks Steve. And so really from working with all of the organizations that as IBM we have worked through facing and overcoming these challenges, our conviction is that you need to put in place a platform for decision making that can deliver confidence in information to the business. And for this platform it is centered on empowering organizations with information, be it from coming at the bottom of having the hardware foundation to ensure delivery at a cost that is reasonable to achieve. Making sure that data and content management for document and process control is in place. And critically and something that we are going to talk about a little bit today with the trust is that the data integration is there to ensure that organizations standardize on a consistent view of information across their business.

And finally at the top layer, the business intelligence to provide business users throughout an organization with the information they need to understand, oversee and drive the business. And with IBM we provide market-leading capabilities in each layer for a modular approach making sure we have the openness to embrace the infrastructure standards that you have already got in place but also the integration to deliver the value when combined together.

So let's take a look at how we can help build business confidence, in particular the understanding and the trust side.

So first of all we are actually going to go and explore in a bit more detail the information understanding and how we can provide better context and consistency of usage to all consumers all for business information. And the first area of this that I would like to look at is how lineage in business terms can help users understand in business terms where data and reports come from. So if a user is able to see where information comes from they will have more trust in that information and more understanding in that information. And in Cognos BI, typically a place where people are looking at their reports for performing analysis, we want to make sure that the business can quickly find out where the data came from right from within their report and get the answer in business terms that make sense to them. After all, it is not really very useful to provide business consumers with all of the technical transformations and calculations as the data winded its way up to the report.

The business users really want to simply be able to see where the information comes from and we will look at this in a demo in a few minutes. And they want details such as who is responsible for the report.

Well who is responsible as the data steward for that information? And we can do this now with one simple right click from where the user is on the screen. But [whether your reporters are all], professional reporters want a bit more information, they wan t a bit more detail or perhaps the metadata model is in IT, we also want to be able to find a technical view within the reporting world, how the information was sourced – did it come from a warehouse – and also give a trail from the report field back through the modeling details to where the sources for the performance management layer. And this technical view is also available with one simple right click from within the Cognos BI environment.

The third area that is key of lineage which is key and is of interest to the more technical business community, typically people like analysts, report authors, perhaps [jurors] and also perhaps those involved with the development maintenance, the warehouse and also the modelers, that those that want to dig deeper into the source of the information, back through the ETL transformations and through the warehouse back to the original transaction source.

And with Cognos 8, again with one click, business users can have direct access from their reports, analysis or the modeling as they are developing reports through to full lineage in the information infrastructure. And we provide this by hooking up to data infrastructure that you have in place already as well as through pre-built integration to IBM InfoSphere Metadata Workbench. And we will look at these in a little bit more detail.

So Steve, could you just talk us through the Metadata Workbench a little bit?

Steve Totman: Okay, so the Workbench provides a window into the information server repository. So what you are doing here is you are allowing a user to see all the way through the process from the starting point where you are dragging data out of the source systems all the way through to the delivery tool like Cognos and then the reporting layer. And the key thing is it is providing this view across them, so a lot of times when we are talking about Metadata Workbench we talk about something called data lineage or providence. So showing to the business people where the information came from. Whenever you roll out a new application like a business intelligence application, the business has an existing way of getting that information. They will look at reports that they get from various sources and they will [run] it through Excel. And what you are really doing is when you roll out the new system they are immediately going to compare the results you are providing in the new system with their existing way of getting data.

What the Workbench does is it allows you when all of those questions come in about well how did you know, did you get this number, because it is completely different to the one I had before – you can right click and say well this is actually how it came through the entire system, this is the transformation rules that were applied all the way through, and actually what we have a lot of customers doing is they then challenge the business people and say well tell me how you go the information before. And you see them drawing little diagrams of [Sneaker Net] where they phone Bob in accounting. So you are providing that lineage view all the way through. This also helps a great deal when you are trying to assess the impact of a change. If you want a new column added in the source system or you want to add a new column into a report, before you even do it you can see how much impact that is going to have on you IT infrastructure. So you are already giving visibility into the system from start to beginning. And as compliance and governance becomes much more important now it becomes a key enabler of that.

Becky Smith: Perfect, thanks Steve. So really by using the Metadata Workbench not only do the data integration report authors have the understanding but data stewards in the business can actually manage and monitor compliance as well.

So the other aspect of understanding information is really making sure that users can understand what the terms mean. And this can be achieved by having a common enterprise vocabulary in place which again can be easily accessible with one-click access. And the key thing about having a common vocabulary is it provides the business with consistency. It gives the consistent use of terms and descriptions for information that is consumed in all reports. And because there is a common vocabulary for each and every report element it means whoever is running a report or whoever is viewing a report, they are seeing that same terminology that explains to them. They are seeing a definition of what a term actually means in business terms.

And again, one simple right click from any element provides visibility into those common definitions. And we will look at this in a demo in a bit more detail as well. So now if we could actually move on and have a look at the demos.

So here we have got a (inaudible) report where we can actually see we have a spike in unsatisfactory product returns for outdoor protection range. But before I make any decisions or actions, I would like to understand more where this problem comes from. So first I want to know where this comes from, what the data, where its origin is. So I drill directly into the business lineage that is part of the Cognos 8 version for environment. And here I can see some business view details such as information about the report, and the path that the data has been through. I can also go through a more technical view where

I can actually graphically see where this item in the report has come through from the data source all the way up into the report item. I can see there have been filters on the information. And I can see that they come through and even look at things like the SQL and see the SQL that is being requested directly from the database. And in this instance I can see that my information comes from my go data warehouse.

So now I have a better understanding of where my information has actually come from. I would really like to understand a little bit more about what unsatisfactory product means and actually perhaps what defective product means. So again I can simply right click from my Cognos 8 report and go into my InfoSphere Business Glossary for Cognos which again is a web interface. And I can see here I have a quick description on both items I selected, my unsatisfactory product and my defective product. And I can drill down into any of these items and get some more information simply by clicking on that information.

So we can see here I now have a long description. I can actually very importantly see who the steward, who owns this definition from a business terminology. And I can see things like examples and related terms.

And again I can drill now directly into the Business Glossary browser. And here again I have got my descriptions. And you will see when I click on the steward it is actually a hyperlink and I can go and get more information about the owner of this information, Becky Kerr. And I can see she is located in Ontario. And I'm over here in the UK. So perhaps what I would like to do is send her an email because I would actually like to discuss a little bit; I'm not quite happy or really understanding the description that is in the Business Glossary. So I would like to have a quick chat with Becky Kerr who is the steward to understand a little bit more of whether we can enhance the description to make it more meaningful for other users who might be consuming this report, or indeed other reports of using unsatisfactory product.

So I have just simply created a quick email that I am now going to send off. So if you would like now, if you would just like to imagine that I have changed hats and I am now Becky Kerr the data steward. And I'm here in my Business Glossary for Cognos environment. I'm in my information server environment. And I am going to, having had a chat with the user, I'm going to have a look at the description and update it. So the first thing I have done, I'm going to search for this item simply by typing in the search. And you will see now it will return the unsatisfactory product for me. And I can select this unsatisfactory product and edit the term. And there are a number of different things that I can actually address here. You can see I could change things such as its name. It also belongs to a category which is if you like web

organizing my terms – you have got the short description and the long description. And the user asked me to make the short description a little bit more customer focused so it was very obvious to anybody consuming and wanting to understand the meaning of the information. So I am just now mending the short description. And I could do things like change usage, maybe enhance the examples, update the status, so it could go through a workflow process.

Other things I can do in my information server is do things like amend the terms, categories, work with stewards. But I can also very importantly import and also export terms. So I can import terms for example from my Framework Manager that I may already have setup.

So having made those changes, I have just gone back and rerunning the report so I can actually make sure that I am seeing that new description, that new update that has come in place. So by using a combination of the glossary and also the lineage I am able to much more easily as a user see and understand what the terms mean, know where they come from. So it helps me increase my confidence in using my business intelligence information.

So we have had a quick look at the lineage integration, the Cognos 8 lineage and how both from a business user they can see, the user can quickly and easily see who owns the report, who is responsible for that information, also the technical view. And we also looked at the Business Glossary for Cognos where we are able to actually look in and see what terms mean. So just to give a little bit more information on the Business Glossary, it can be a really good starting point for designing and implementing information governance into your organization. It is a completely web-based authoring, sharing, and management of business metadata, so there is no need for any clients on local desktops.

And by using the Business Glossary it really helps to align both the goals of the business and the efforts of IT together, enhancing the collaboration which not only can result in accelerated project delivery because we are information sharing, but also helps with acceptance as the business is engaged in the role of stewards the whole way through the process. And because the Business Glossary can actually define the relationship between business definitions and IT assets, this makes communication between business and IT much easier when for example the users want changed information and are requesting changes or additions from IT, is that linked [conjoin-age] between what IT may call a term or what terms are called in source systems and how the business would actually refer to them.

It also helps to increase information trust across enterprise applications because although it can be used within the Cognos 8 version 4 application, it can actually be used across any other business application

as well to allow the Glossary to be taken broadly across all applications. And very importantly it helps to establish accountability. And it is accountability not just from IT but also accountability across the business because with the creation of stewards and the assignments and responsibility is that sense of ownership which helps to build up the understanding and then creates confidence and trust in the information that is consumed.

So what are ways that we can actually get started today in helping to increase understanding? Well there are number of key areas and probably the most important is to look to work with key business users to develop a good collaborative process to describe the terms and descriptions to be used across the business in everyday reporting and everywhere that the information is consumed.

And the first place to start is from across the different sectors and areas of the business to identify some key business stewards who are going to manage and own these data definitions, starting of course in heavy collaboration with IT. If you are using Cognos 8 BI of course a good place to start is to make sure that you update your Framework Manager models to include descriptions of all the metadata. And again, don't forget engage the business in defining these descriptions, especially because the Business Glossary can import the definition that you have already got in your Framework Manager models.

Also look at how you can explore the value of creating full end-to-end metadata lineage, all the way from source systems to the user. And you can do this and start creating this information by importing your Cognos metadata into your metadata workbench environment. So it is bringing in not only from the source systems transformations that you perhaps put in place but bringing in everything that is happening to that metadata within the Cognos world.

And then finally bringing back to the Glossary again, stop building your dictionary of terms. And this can actually start if it is not in your Framework Manager models, even in the spreadsheet. And have them ready to import into your Business Glossary.

So with that I would like to move over to starting the poll.

Mary Windeshar: Thank you very much. Our first poll question today is "What do you see as the most important driver for increasing the understanding your users have in the business information they use from BI and Performance Management applications?" So your choices are four. The first is determining where the information came from. Second is understanding the terms and definitions that describe

information. Third, identifying business stewards to take ownership of business information. And IT and business collaboration to ensure effective description of terms.

So this is a poll where we ask you to choose your answer right there on the slide itself. Click the circle right next to the answer you choose and then hit the submit answer button. Once again the question is what do you see as the most important driver for increasing the understanding your users have in the business information they use from BI and performance management applications. Determining where the information came from. Understanding the terms and definitions that describe information. Identifying business stewards to take ownership of business information or IT and business collaboration to ensure effective description of terms.

We want to give just a couple more seconds so that everyone can weigh in on this question. And by doing that I will remind you that you can ask a question at any time during our event as well. All you have to do is type it into the question box which is located in the lower left hand of your console and then once again hit that submit button.

I think almost everyone has weighed in. Let's take a look at our results. The question is what do you see as the most important driver for increasing the understanding your users have in the business information they use from BI and Performance Management applications. And here are the results. D is a big winner. Almost 40% are saying IT and business collaboration to ensure effective description of terms. So that means that our second place is almost a tie, about 26% of you are saying determining where the information came from and 24% of you are saying understanding the terms and definitions that actually describe information itself.

Let's go ahead and take a look at our second poll today. That question is how are you addressing the need for a business dictionary of terms today. So, first choice – spreadsheets, documents or my favorite, post-it notes; leveraging separate metadata-based solution; relying on business analysts, report authors, IT to answer questions; relaying on data descriptions in the Cognos metadata model; or, well maybe this is my favorite, it's not currently being addressed. Not trying to sway you there. We just want to know from you how are you addressing the need for a business dictionary of terms today. And I know you can read but I am going to go ahead and tell you the answers one more time. Things like spreadsheets, documents or post-it notes; leveraging separate metadata-based solution; relying on business analysts, report authors or IT to answer your questions; relaying on data descriptions in the Cognos metadata model; or, this is actually not currently being addressed in my world. Please chose your answer by clicking right next to the answer you like on the slide itself and after you have done that hit the submit answer button.

This question again is about a dictionary of terms. And I do want to remind you at the end of our event we will have a question and answer session so you can ask any questions you would like to ask and our experts will answer them for you.

Alright, it is probably time now to take a look at the results of our poll. The question again, how are you addressing the need for a business dictionary of terms. And the results are, well it looks like 28% of you are saying using spreadsheets, documents or post-it notes. 30%, that's our winner, relying on business analysts, report authors, or IT to answer questions. And about 17% of you are saying we are not really addressing that at this time. Becky I would like to turn it back to you to continue your presentation.

Becky Smith: Thanks so much Mary. So we have had a look and thank you very much for answering the poll questions for us. We have had a look at how you can increase the information understanding by increasing context and consistency, using glossaries and lineage capabilities. And now I would like to hand over to Steve Totman who is going to talk about how we can improve business confidence in information by increasing information trust with visibility into the quality of the data and how you can improve the quality of that data over time. So Steve, over to you.

Steve Totman: Thanks Becky. So the key thing that we have found in an awful lot of customer engagements is that the most significantly underestimated factor in any of these deployments, any data-centric project is actually the impact that data quality can have. And it ranges from Business Intelligence all the way through to MDM. And the key thing that we have found is that if the users don't trust the system or the data itself, they won't use it.

Now, we talk about agility here – so agility is impacted because if your users don't trust the reports you have run for them, they will go back and find alternative methods to getting it. So you can spend a huge amount of time and money building the system, but if the first experience the end users get with this is that they get a report they don't agree with and they find bad data, it is very difficult to get them to come back to it. And to this end there is a direct linkage between quality and your ability to understand risks. We have seen a lot of banks all across the world now that have problems because they didn't understand their risk exposure because they didn't have a good understanding of their customer data.

And a customer recently was actually pitching to their CIO a data quality project and they had a little cartoon where instead of calling out ROI as Return of Investment on the particular project, they pointed

out that in this case ROI meant the Risk of Incarceration for the CIO. Now, data quality is not just something that costs your company money now.   People can actually get into real trouble and get pursued through the courts over this. So it is a very different type of problem now. And it has always been there but now it has just got much more visibility.

And when it comes to the data governance component, now organizations are really waking up to the fact that you need a good governance policy, you need a strategy around this. And one of the first steps is getting a handle on your data quality. So the stewardship that Business Glossary is bringing in kind of identifies who the people who own the items are, but then actually getting people to pick out the valid values, etc. And the climber is a good example here. If, when you are beginning a mountain climb your very first experience is that your rope breaks, a bit of rock breaks off in your hand. You are very unlikely to continue the climb. And in the same way when you first roll out these business intelligence decisions/systems to users, if they first experience they have is that the first report has bad data in it, it is going to be very difficult to convince them to come back again. So it is a real critical problem going through.

So why does this problem exist. Well everyone understands that enterprises have grown up. With the present economic climate we are also seeing an awful of consolidation. So where there were once 4 or 5 companies, there are now 3. So you have to be able to adapt to growth not just naturally but through this acquisition strategy. And of course every different group has, for example the marketing team will be very interested in their specific information. And if IT is not moving quickly enough as a group, they will do an individual project to solve their problem.

So these very complex systems have grown up. You have got duplicate data all over the place. You have got data moving around between these systems. The diagram on the right is actually a customer's architecture diagram and it is slide 1 of 4. And that is not a particularly big customer either. So what we are seeing is that there is this huge complexity of systems and on top of that the data inside the systems naturally degrades. So Gartner has a quote about how data degrades at 2% per month. And that sounds like quite a scary number until you actually think about how just in a customer name and address table you have got people getting married, you have people moving address. So that data is going to naturally change, and that is happening all across the systems. And then you have got the duplication and the synchronization that just adds to that.

So there is no sort of single database that everyone uses. All of these different systems and applications are interacting. And it is a really big problem that is sort of growing as well. Every year enterprises get more data and this simply multiples.

So just to give you some examples of some really common data problems, so if you look at the top here we have got some simple misspellings of Catherine Roberts across these different systems. And these are all perfectly valid entries in the particular source system. But as you can see, recognizing these are all one single customer and bringing them together is a really difficult problem.

Now everyone uses the name and address example because everyone gets it very quickly. But also look at the cable example. So it is not just your customer names and address, it is the product systems. So where you have got product codes, etc. There are some really interesting customers here in the UK where they have multiple different systems and it being, it was actually a defense contractor. So they weren't just storing cables. They were actually storing shells and ammunition and tanks, etc. And what they found is that they had massive duplications, across three different warehouses. And of course this isn't like a cable – when you are storing a shell there is a much higher degree of security and much higher cost of storage. They had massive duplication across those three different warehouses.   So just by getting a good handle on their data they had a massive impact on their bottom line of storage costs.

And I like the last example the most really. So this is IBM. These are all different variations, spellings of IBM. And this is something similar to what we see on our ledger system when we have customers sending us accounts at the end of each quarter. And even something simple like IBM gets spelled so many different ways. So even we are a good example of this. Now what we have found through this whole process is that the companies that are really successful with data quality projects, firstly they follow a clear methodology. So you have to follow a process through. The tooling is a huge enabler but you have to put that process in place. And we have a very well established, well proven methodology here.

Also it is very critical to get high level sponsorship and ownership of this. Data quality is something that has to have visibility at the highest level because you are going to have to point through multiple different systems and actually say well here is where the problem lies or these areas, and then have someone who can help support you when you go to change the source systems. There are two approaches. You can kind of take the approach of leaving the source systems where they are and trying to fix the quality on the way through or the much better approach is to fix it on the way through and then go back to the source systems and try and fix the problems there as well.

So you must establish what you want to improve the quality of. So it is very critical that you don't try and go after everything at once. You pick a specific project which has high visibility and will have a lot of impact and use that as the driver to then broaden the scope of the data quality initiative. You also have

to go through and analyze all of these source systems. It can seem like a kind of scary thing at first because everyone knows the spider web of systems that exist out there but we have some very good technology that can help you with that. But you do have to pick specific data sources and go through them and assign valid values for the various entries.

And, going forward you really need to recognize that data quality is not a one-off event. It is like your garage. You can't just clean it out once a year. You have to clean it out on a regular basis otherwise it continuously gets messy. So data quality is the same thing, you can't just do one massive clear out and then everything is great and your garage would always be tidy. You have to continuously repeat the process. So it is not a one-off event; you have a repeat, a repeat and a repeat.

You should also start looking at not just the data quality of the source systems but look at the data quality and the systems that are feeding the information to the business users. So data warehouses and data marts are great candidates for checking data quality as well because as part of the process of moving data from the source systems across, you may actually introduce further errors.

And also when it comes to sort of the publicizing of this, make data quality something that business users care about. I have seen some great customers where when they deliver their reports every morning to their users they ring a bell and they write up on a white board the percentage of data quality in their system. So they get the business to understand that not only is data quality something that is important to them, it is important to the IT group as well and that they understand that the improvement of quality is an ongoing thing that the technical team is striving for.

So, within the IBM suite of products we have several pieces that can really help with data quality. The first one is the ability to understand and monitor data quality, a component called Information Analyzer. This is a really critical step. When most people start looking at data quality in assessing their source systems what they will do is they will hire in a graduate, take a bunch of data out of a source system, put it into Excel and hand it to the graduate and have them look through the system. Well, if you can imagine some poor graduate sitting down with a huge Excel spreadsheet, they start with column one and they write down well this is customer name, by the time they get to column ten they are kind of bored so they hit page down a couple of times and say well this is a customer address. By the time they get to column 50 they are really bored so you are lucky if they don't just hit end and look at the bottom few rows in the data. By the time they get to column 100 they are bored out of their minds. So you are lucky if they even page down through the data. And it may be that the most important data that you have about your customers is in the later pieces of that spreadsheet.

So what Information Analyzer is, it is a data profiling tool. It's kind of like the Terminator – it never gets bored, it never gets tired and it just charges through the data and does its analysis. So it starts with columns then goes to tables and then goes across tables. And it spots relationships you would never spot as a graduate going through the data itself. Now if you imagine a real customer environment with hundreds of columns across hundreds of different sources, if you want to spot a relationship between two sources, it is very tricky. But of course Information Analyzer is just doing this based on a frequency distribution. So it will spot those relationships.

When I was a child we used to play this game where you put a bunch of cards down on a table and have to turn two over to make a match. It is a very difficult game when you play it with just 100 cards, but in a real customer environment it is like the cards are all over the table, all over the ceiling, all over the floor. What Information Analyzer does is it spots those relationships on two completely different tables and says well there is a relationship here or there is duplicate data. So it is a great tool for getting that initial view across your data systems.

And Information Analyzer also has a rules component so you can build in much more complex business logic.

Now once you have spotted all of these issues using Information Analyzer you can then use Quality Stage to resolve them. So Quality Stage is a data cleansing and matching engine. It works as well on names and addresses as it will on product codes, etc. So it has three steps, the standardization step will it will take any free form text field and put it into a standard format. So if you think of my name, my name is Steven Totman; Steven can also be spelled Stephan and I have seen my surname Totman spelled every different way – Toutman, Tottyman, etc. And what it does is it will take those different spellings, standardize them, and then match them. So it uses public record linkage to do the matching and then it will gladly do survivorship. So when it brings records together it will actually allow you to choose the best data to survive. And when we do the matching it is not a pure match and based on the fact that it actually uses the recurrence of the data to choose the match. So in the UK a name like Fred Smith, which is a very common match, will be much less than Steven Totman which is a much more obscure name here in the UK.

So, now –

Ernie Ostic: Thank you for taking the time for a brief look at InfoSphere Information Analyzer. As a complement to Cognos, Information Analyzer will help you examine the content and structure of your

data, assess its quality and help increase user confidence in their reporting solutions. Information Analyzer has three major capabilities. The first is column analysis which will go through all of the rows of all of your data and examine columns for their data types, and also for various patterns and trends in frequency distribution. Primary key analysis looks for unique keys amongst the columns in your set. Foreign key and cross domain analysis compare the values between sets and baseline analysis allows you to establish an initial base where future analyses can be compared against to measure quality over time.

The home page that we are looking at of Information Analyzer has many graphs that show the progress of your analysis project, how well you are doing, what you have analyzed so far and the anomalies that you may have found. There are many reports that go along with the detailed graphical user interface that we are looking at now. Let's take a closer look at column analysis. Here we see the high level summary page, what we have analyzed so far. You can see that we have analyzed 100% of our columns but now we are going to go into review.

So let's open column analysis and go to the next level of detail. Information Analyzer runs on a parallel framework and so can examine huge volumes of data to try and assess the quality and examine the trends that exist in your data sources.

Here we are looking at the summary page. We see a list of columns; lots of details here but in particular we want to point out the red flags that identify some of the initial inferred differences that were found when we looked through the data. In this case a data type as it discovered was different from the actual defined data type set. There is a note, this yellow icon on the zip field – so the analyst has already made some discoveries that we might want to look at further.

This brings us to the next level of detail. We are initially seeing a frequency distribution for all of the data that was in the set that was analyzed, but we can also examine properties, data class, domain and completeness in the formats of an individual column. What is particularly interesting here in frequency distribution is we can look at it graphically with a histogram giving us a visual view of what the distributions are. And perhaps we might be interested in knowing why are so many of the postal codes null? Is that a problem for us? Perhaps it is.

Properties examines the percentage of values that fit into certain data types, link data type precision, scale, all of these things are graphically reviewed and may be important to you for certain data types in certain columns.

If we take a closer look at the format tab, and let's particularly do that for customer number, we see how it categorizes into various formats that were discovered in the data. And perhaps for our purposes we might want to look at this and say hey, a customer number of only one digit is in fact a violation. So I will flag it as a violation which populates the right hand grid and this right hand grid can then be used to generate reference tables for use in other applications.

Let's now take a look at cross domain analysis. Cross domain analysis allows us to compare the values between tables possibly discovering joins and relationships that we might not have realized existed. In this case my candidates for analysis are customer ID and customer number. We found a relationship between them so let's look at what cross domain analysis actually discovered for us.

When it runs it tries to do a relationship comparison between the two sets. So it gets us a summary and tells us that we did in fact have some intersection. If we look at the further details we get a visual diagram that illustrates the two sets and shows their intersection. So in fact in this case we do have a relationship between the two. I like to also call this predictive join analysis. Cross domain analysis is helpful for the joins it discovers as well as perhaps identifying situations where you thought there was a join but in fact we might see that there are zero intersections.

So that is just a quick view of Information Analyzer. Back here to the home page we can summarize by saying again that Information Analyzer will help examine your data, its structure and content and increase your user's confidence in their reporting solutions. Thank you.

Steven Totman: Thanks to Ernie for that short demo there. Now IBM offers a number of serves that can really help you here. So the first is a Data Quality Assessment. So we have a service that comes in using the tools you have seen and provides you with a detailed report of your present quality situation. It includes all of the investigation steps, it looks at all of the aspects of the data like storage, data types, etc, and it is really useful because it provides an independent audit of your data quality status. We find customers really like this because it means that you have got someone independent coming in and doing the audit so you are not pointing fingers internally. And it means that you can then go back to your management and say well this is where the issues are with the data quality system and this is where we should concentrate things. And it will give you an idea of where you are going to get the most payback because we all recognize in this sort of economy you can't do these massive all-encompassing projects. You have to focus them and pick the specific areas which are going to have the biggest impact to the business.

Now, the poll questions?

Mary Windeshar: That's right. It is now time for our next poll.   The question is: How confident are you in the quality of your data used in BI and performance management systems? So your choices are: All data consumed is clean and accurate; most of the data consumed is clean and accurate; we know we have a data quality issue, but don't have visibility into how big or small that problem itself is; or, don't know. And once again you are going to hit your answer right there on the slide itself and once you have done that please click submit answer. And once again the question: How confident are you in the quality of your data used in BI and performance management systems? And your choices are: We've got everything clean, accurate – just like Steve's garage; most of the data consumed is clean and accurate; we know we have a data quality issue, that would be my garage, but don't have visibility into how big or small that problem itself is; and, we don't know. So hit your answer on the slide itself. Just click the circle right next to the answer you like and then hit submit answer after you have done all of that. I will go over that question one more time. How confident are you in the quality of your data used in BI and then performance management systems as well? Please choose just one answer and let's go ahead and take a look at the results.

Well, we know we have a data quality issue but don't have visibility into how big or small that problem itself is is a very slight winner over most of the data consumed is clean and is accurate. So there are the answers to that poll. Let's take a look at our next one. That question is: How do you address the quality of the data used in your BI and performance management solutions? So we are giving you the choice of: Rely on the quality of data in source systems; Implemented data cleansing and quality processes; plan to implement data cleansing and quality in next 12 months; or, no data cleansing and quality processes are currently planned.

The question again is: How do you address the quality of the data used in your BI and performance management solutions? Well, some of you rely on the quality of data in source systems. Some of you have implemented data cleansing and quality processes. Some of you have some good plans, you are going to implement data quality and cleansing in the next 12 months. And some of you are saying we don't have data cleansing and quality processes currently planned.

Once again, question about how you are addressing the quality of data used in your BI and performance management solutions. This is a good way to see not only what you are doing and let us know about that, but to see what everybody else is doing as well.

And don't forget also to submit questions that you would like to ask our presenters by typing them into the ask a question interface and we will get to them as soon as possible. The question for this poll is: How do you address the quality of the data used in your BI and performance management solutions? Almost 30% of you are relying on the quality of data in the source systems. And then the largest amount

of you have implemented data cleansing and quality processes. So we will move on with our presentation. Back to you Steve; I'm going to give you that slide and you can go ahead and get started.

Steven Totman: Thank you. It's nice to see so many people are making progress. A few years ago data quality was something that you talked about and everyone would kind of say it wasn't a problem. Now people are really recognizing it is.

So how can you get started today? Well the first thing is to make sure that you pick what project you are going to go after. So don't try and swallow or eat the elephant all at once. And take ownership for it; you need people in the organization to step up and say it is important. You need to start understanding your enterprise topology, so understanding how these systems interact. And really make trust and data integrity and quality something that people talk about freely and actually rank systems on. I mean if you are going to deliver reports out you should make sure that you are using the best available, most trusted source for that. And really start looking across the different systems for duplication, etc, so look at a profiling technology to help you with that process.

And all through this data quality is not a one-off. You have to do it in a repeated manner, so just like with the garage you have to continuously clear it out. And make sure that people understand the visibility of data quality. That simple idea of the white board and telling people the percentage of quality each time you deliver the reports is a really good way of making them understand it.

And then the final slide, just passing back to Becky on this one.

Becky Smith: Thanks Steve. So really we have give a very quick walk through in covering about how we can help to increase both understanding and trust and make sure that from an IT perspective you are delivering data in terms that the business understands, owns and trusts. So we had a quick look at the InfoSphere Business Glossary for Cognos and how you can leverage a centralized glossary by creating business definitions and terms, also add annotations and really enhance the core items in the systems and work with that collaboration between IT and business to set it up.

We looked at how Cognos 8 Lineage can help both the business user and the more technical user gain an understanding of where terms come from and how with the InfoSphere Metadata Workbench you can provide and give that traceability on the origin of the information and anything, any stages, any transformations, calculations that have happened to that information through its entire lifecycle.

And then we talked about trust. Steve took us through some of the areas of trust and how the regular monitoring and reporting of data quality is critical and how having metrics and trends in that information can really help build confidence. We looked at the demo that Ernie gave how with InfoSphere Information Analyzer we can achieve this.

So before we move onto Q and A, just to recap we really wanted to cover these two areas that we have talked about and really help you with Cognos software and InfoSphere that we can address your needs, be they modular – we can help you address immediate pain points. So perhaps if it is a glossary that is required or the need to analyze your information, can bring in Information Analyzer for that.

And it is important to bear in mind that as part of IBM both Cognos and InfoSphere software are open. We make sure that you can utilize your existing investment in IT standards and infrastructure that you have in place.

But very key, we are integrated. So for organizations, if you are looking for a more complete solution, the combination of Cognos software and InfoSphere software from IBM can help to provide that solution.

So with that I would like to hand it over for the Q and A session.

Mary Windeshar: And just a quick reminder to our audience, be sure and type your question into the ask a question interface, and Becky you can go ahead and start out with the questions themselves. If we don't get to all of the questions during our event, will you be able to answer some of them after the event as well?

Becky Smith: Yes, certainly we will. We will follow up on email with the questions that we are unable to answer.

Mary Windeshar: Great.

Becky Smith: So just to start off, we have a couple of questions in of whether the Metadata Workbench and also the Business Glossary of whether they are available as part of Cognos BI or you have to buy them as a separate tool.

So in both instances they are actually a separate product and license. And they are integrated with the Cognos 8 version 4. So they are both available but they are an add-on, something that you can choose to really enhance your existing or a new Cognos environment.

Steve, we have a question in that has come in on whether Information Analyzer can perform analysis on the ETL processes as well and how we would go about addressing analysis as part of both processes from a data quality standpoint?

Steven Totman: Absolutely. What we are finding now is that the new rules component in Information Analyzer is a really good way of actually auditing on your integration processes. So rather than just auditing your source systems, you actually then audit your target systems as well to check the correct transformation has been applied throughout. And we are also seeing a lot of customers buy in data that they get in from a third party. And it is useful to actually just profile or analyze that to even decide if you want to load it into your downstream decision support systems. So before you even start the integration process, profile it up front, see if there are loads of nulls in there that you might want to remove. And we have actually seen customers that have used that to then go back to the supplier and having cleaned up the data, having run it through something like Quality Stage, sell back the clean data to the supplier. So it goes in a complete circle.

Becky Smith: Perfect. Thanks Steve. So we got another question that I would like to put to you Steve is when a warehouse is not clean and not actually perhaps managed by the person running the BI environment, what would you say is the best practice to address and resolve this issue from a business or if you like political standpoint without creating a negative situation. So would you use Information Analyzer to help create reports and things to help engage the business or perhaps something like the Data Quality Assessment?

Steven Totman: I think both would be very applicable. The Data Quality Assessment is a great way in a short period of time getting a good sort of overall view of where you should spend the most time on your systems. Information Analyzer being used on an ongoing basis and it generates a huge amount of very easy to consume reports by the business. Just that, these are automatically published out through the web console means that the business can kind of go in and be sort of self serving in terms of analyzing the data.

Now we have actually seen that some customers because IA makes it so easy to get to the data and sort of view through it, well actually restricts it from the business because they don't want them poking around too much initially. So there are various levels of control you can apply. But yes, all of those are a great way. The key thing is to make sure that the business is engaged. If IT is simply trying to push it through you probably won't get the funding or the sponsorship.

Becky Smith: Thanks for that Steve. The third question is what version of Cognos 8 BI introduced the lineage functionality. The lineage functionality that comes as part of Cognos 8BI was introduced with the Cognos 8 version 4 release. So you need to upgrade to that version to get those features as well as the integration with the Business Glossary and the Metadata Workbench as well.

And very quickly there was a question how is InfoSphere configured with Cognos 8? So basically it is very simple configuration to direct the Cognos 8 to your InfoSphere, your Business Glossary or your Metadata Workbench.

Mary Windeshar: Becky, I'm afraid we are nearing the top of the hour. Would you like to summarize now?

Becky Smith: For those questions we haven't been able to answer we will make sure that we do come back to you in email and answer your questions. So thank you for everyone who has submitted questions and there are still a few moments to submit a few more.

So today we hopefully have given you an understanding and a little bit of background on how you can increase business confidence, especially through information understanding and also information trust with a combination of Cognos and InfoSphere software.   And you can see the slides up we have got at the moment. We have some links there for you plus also white papers that are available and of course this recording is going to be available for you to watch at a later date and come back and pick up recordings.

I would just lastly like to thank Steve and Ernie for taking the time to participate in today's conference. And hopefully you have found the information useful and do submit a few questions and we will come back to you. So thank you very much.

Mary Windeshar: Thank you Becky and also to Steve and Ernie. We have reached the top of the hour and would like to ask our audience to complete the following brief satisfaction survey before logging out. To see our calendar of upcoming events please visit www.ibm.com/cognos/techtalk. Thank you all for participating in today's webinar.