0:00 slide 1	Hi, I'm Patrick Connolly, I'm a program director with IBM's InfoSphere and I work in the product marketing group. Today we are going to talk about integrating strategies to enable information governance. Information governance has certainly been a hot topic in the minds of clients that I work with, and I'm sure that in your organization focus has been placed on establishing processes, formalizing standards to be able to do more with less, to be more competitive, to reduce cost associated with integrating information with information technology, etc. So we are going to talk about the strategies to help in your journey to attain information governance in focusing primarily on the notion of integrating information.
0:47 slide 2	Let's move to the second slide here. If you think about your organization and how it produces and delivers information, it's often useful to think of the notion of an information supply chain as a good metaphor for your organization. You're talking raw materials that you see on the bottom left there, from external information sources, or maybe from applications that are running. And those raw materials or that data needs to be transformed and modified and also cleansed to be able to become a finished product, much like you would in a standard supply chain, and of course then, delivered in a timely fashion for the consumer of that information. So in this discussion, we are going to focus on the challenges of integrating and cleansing information. Virtually all information intensive solutions, applications, would required some degree of integration and certainly all of these types of initiatives can benefit data quality best practices to be able to deliver a trusted view of information. And that of course these activities, if done correctly and if done in concert working with your colleagues in business and IT, can yield better information governance. And there we are talking about some of the policies, some of the capabilities around information quality, lifecycle management and standards.
2:13 slide 3	So let's move to the third slide. Let's start with the baseline definition of information governance. It's useful often to look at sources like Wikipedia to see what one definition might be. And this is a pretty worthy one. Multi-disciplinary structures, policies, procedures, processes and controls. That's a lot to take certainly with this definition. I often heard it also described as the orchestration of people, process and technology, which I think is interesting, because we are really go to technology, we've certainly invested a lot of money there, we're pretty good with processes, but the people amendment is often a challenge within information governance. And certainly, as we look at information governance and the management in delivering information, those organizations that are competing more successfully, that are thriving in ours times are the

	ones that are calculating this as an enterprise challenge. So we're getting away from silo based or product line based or business unit based organization around information delivery. And the goal here to deliver common good for the enterprise. So we do the things we've been doing, but we do them a little bit differently and in concert with our colleagues in business and IT across different product lines, across different business units to achieve some benefit around increasing revenue, lower costs, and even reducing risk for your organization, whether you're a traditional corporation or a public sector enterprise. So that's a good baseline definition of what information governance is.
3:49 slide 4	Let's move on to slide number 4 and talk about why we are doing information governance. We need to start treating information as an enterprise asset. It lives in a lot of different places, but it has a lot of challenges. It's often duplicated across business units and product lines and solutions, and specially some of these enterprise systems. And it can live in a variety of different formats. And so the very nature of it is to have quality issues to be talked to the analyst. I believe Gartner estimates the data left in place decreasing quality at a rate of 2% per month. So unlike finer wine, data doesn't get better over time, and this bad data certainly can yield bad decisions. So to be able to make impact, to be able to gain some ground, we need business and IT to work together to deliver value. And the way to do this is to establish standards, policies, and processes for how data is originated, managed, transformed, delivered and consumed by the business process owners. So by tackling business initiatives like information governance, and that's the important thing as it needs to be in concert with business and IT. Organizations are better able to response to competitive threats, deliver value, and also face some of the challenges around regulatory requirements, which as been a big rational or a big driver for information governance.
5:20 slide 5	So let's move to the fifth slide here and let's talk about the role of information integration in a governance context. So if we look on the left side here, we see some challenges that are facing information technology. Our goal is that we need to somehow deliver trusted information from this potentially network systems and data sources etc. in light of requirements to be able to deliver information at the speed of business. And also contained with volume and variety of sources. Organizations typically are in a pattern of hoarding information without really looking at the lifecycle application, how long you need to keep a data, is it legal for your organization to keep

	hanging up to that data. And if you do some degree of archiving or lifecycle management, what are the best policies for managing that information and making the critical information available. So taking these sources of information and combining, filtering, aggregating, summarizing, integration this information, then to the users of information so they've got the facts that they need in a timely relevant fashion to be able to make the right decisions. And that's really the goal of integration information that whether we were talking about delivering data to analytic applications, whether we were talking about other types of products, and we are going in some details about some of these. So our goal is to make sure that everyone is more informed, confident and aligned to build to achieve better business outcomes.
6:54 slide 6	Let's move now to slide number 6. We've already seen some challenges in integrating information, in light of a current mantra of information technology which appears to be do more with less. But also, if you think about it, some of these things have been around for a long time. I know in my first computer science course, one of the things that they talked was this notion of garbage in garbage out. So bad data in yields bad outcomes yields bad information. And that certainly is the case today, and perhaps more so with the wide variety of sources, transformations and targets that aren't helping, and they are also different perspectives on the data. And of course, the notions of volume variety, we're dealing with huge volume of information in a variety of different formats, so it's not just about data that lives and are hosted in columns of a relational database. Unstructured information potentially makes up roughly 80% of an organization's data. So we're facing significant cost pressures, again do more with less. There's been far too much focus on individual projects, rather than looking at the good of the enterprise, eliminating duplication across information assets. There is also been far too little emphasis on data quality. We need to make sure that the users or the consumers of information can trust the data that they are looking at, that it's been standardized, planned, and represents a consistent single version of the truth. There is also a profound lack of understanding of information. So we don't know where it came from, certainly we have issues or challenges identifying the owners of information and whether or not they can trust the information.
8:33 slide 7	Let's move on to slide number 7 and talk about some of the types of projects that can benefit from information governance. So, as I

	mentioned before, better information governance, that is delivering trusted information, can certainly benefit in projects like BI implementations. And those can be smaller based BI implementations or large warehouse projects. Those clients seeking to deliver a single version of the truth with Master Data Management (or MDM) solutions, or CDI, can certainly benefit from better data quality, better integration practices, and better management of information. Applications, so we are talking about large scale ERP or CRM solutions, where even home grown applications can benefit from better information governance. If your organization is looking at providing real-time information, or what's often referred to as information as a service, there is certainly benefits to enabling some of the processes, some of the best practices, to be able to better govern that information, that's driven that access and that's delivering information. Enterprise Business Glossary, we'll talk about a bit later in the context of metadata, but the notion that if we do things correctly, and we gather metadata and expend on the definition of that metadata, we can actually create a shared vocabulary that spends our IT assets all the way out to the business representation. So more on that just a little bit later. And certainly, if you're running new applications or implementing new systems, you can benefit from better information governance.
10:00 slide 8	Just move on to slide number 8. Let's talk about a rough process for effectively integrating information. And so, the process that I have talked about here are certainly an iterative approach, so this happens time and time again, for projects, for large scale efforts, etc. But the notion is that first important we need to understand our information asset. We often refer to this notion of helping client get their arms around their data. So applying solutions for data discovery, profiling, analysis, and data mapping, can go a long way towards ensuring the success of your initiatives. But one of the important things of each of these processes is the big era of going down into the shared metadata repository. So as we're going into the process of data discovery and profiling, we are able to harvest and get a hold of that metadata, which we can then enrich and also leads to a deeper broader understanding of where the information came from, what happened to it, who owns it ? So once we have a detailed understanding of the data structure content and relationship across the entities, we can clean up our

	information. Data quality includes processes around standardizing data, verifying whether we are talking against business rules or something as simple as verifying an address with postal information from the post sign up in over 20 different countries. Best inquest matching capabilities, so applying advanced probabilistic matching algorithm speed to identify what records potentially one another across different systems. And then figure out the best to survive that information, whether it's creating a best of breed record, like an MDM would use, or perhaps creating a link to view, or even cleansing and getting rid of the data that is no longer used. Again we're capturing metadata throughout this entire process to create a broad understanding everything that happens to your data from source, to transformation, to report. And last but certainly not least, transforming and delivering information. So that's the notion of aggregating and combining and summarizing data for the various information intensive solutions that we talked about. And this can happen traditionally in batch processes, but there's a growing interest and desire to have more real-time capabilities. So that can include true real-time access, as well as change data capture or CDC delivering of information. And we can also take these cleansing and transformation processes, wrap them up in a web service and deliver them for Services Oriented Architecture. And that's great if you think about it, because it allows us to define a standard and a policy, wrap them up in a web service that is centrally maintained and the rules behind it are centrally maintained and shared across the organization, shared across different solutions. That's sound a lot like what we're trying to achieve with governance. So there's some great benefits to these types of approaches.
13:17 slide 9	Let's move on to slide number 9. Let's talk a little bit more about this getting your arms around your data. If you look at the various stakeholders in our information processes, we have a lot of challenges. We don't know what is the information we need, we don't necessarily trust the data, we're not leveraging what we've got. There's also a gap between the business users and the technology providers, IT and business. There's also a need to better understand the data across various sources and the relationships across those information assets. So we need a better of getting our arms around our data, to better

	provide better information, better outcomes for these different stakeholders in the organization.
14:00 slide 10	Let's move to slide number 10. Virtually all information intensive projects require that you get your arms around your data. It's pretty safe to say that you can't manage what you don't understand. And the challenges will guide to data across different applications and databases, without necessarily understand the relationships between the data entities and maybe duplicate representations across the enterprise. That's why we each have different definitions of information, so a person looking at a report for a business context has a very different vocabulary than the IT providers or the database administrators, who are concerned with the correct feeding of that information or that data. We also don't understand or have very much insight to the quality of information. And often, data models and designs are out of date and applications can manipulate or transform the data without the degree of control that is necessary for better information governance.
14:58 slide 11	Let's move to slide 11 and let's talk through an example. In this case we're looking at building a data warehousing application and the steps that we can go through to better understand the information that's driving this information intensive solution. So first thing we need to do is figure out what's important to us. So figure out what are key performance indicators so we want to track whether data warehousing and BI exploitation. And also figure out what the associate business terms could be for that. So right of the back, we really try to orchestrate between business and IT. A consumer needs reports who are talking to IT to create this end to end view of where the data comes from, what happens to it, where it goes, as you see in the box in the middle of the screen. And that's actually a screenshot from information servers, Blueprint Director. We will talk just a bit more about information servers a bit later. Once we have defined those KPIs and terms, we can start looking at where data exist and what are the relationships between data entities. Next, we assess the quality across the sources. So we figure out what type of structural anomalies we might have, what kind of quality issues. And we can also test against business roles to identify the overall health of information that we are trying to transform and deliver.

	Once we've done this, we can create a data warehouse model, using those identified sources as inputs, orchestrated around the KPI we want to deliver. IBM also has some industry models so you don't have to start from scratch and allow best practices models to be able to jump start your data warehouse implementation or your specific industry. Next, step 5, create business transformation rules, so figure out what we need to do with to the information. We need to consolidate the information into data warehouse. And last but certainly not least, we want to verify and report on the data lineage so we document where the data came from. Because of this shared metadata approach that we are taking, each step of the way we are collecting and updating metadata, we have the ability to have detailed data lineage and we can do detailed impact analysis. And we certainly want to measure the quality implication overtime. Data quality is not a once and done activity.
17:17 slide 12	So let's talk a little bit more about data quality as we move to slide number 12. Our goal is to ensure quality or ensure trusted information across the enterprise. But we've got some challenges whether you're from the business side of the house or the IT side of the house. IT of course is trying to do more with less. The business doesn't necessarily know how bad the data is or how much it costs to fix it, they just know they don't trust it. They have very little confidence in the information. IT understands that there are or may be problems and anomalies in the data, but they really don't know what's important to the business. And so between these two groups, we really don't have a plan in place to be able to positively impact the quality of information that is driving our business processes. So we need to clean up the information. We can see there's some challenges in collaborating between business and IT.
18:12 slide 13	Let's go ahead and look at a process on slide number 13 for applying data quality. I'd like to propose a process for data quality that can be useful and that applies to a variety of different industries and projects. As we mentioned before, the first step is getting your arms around your data. So identifying who the owners are, doing some source data analysis and profiling, as well as figuring out the relationships the entities. And then measuring and creating our baseline for our data quality. That baseline is critically important because over time,

	as we compare our results to that baseline, we are able to see how we are progressing. Is the data improving? Or are there situations that need our attention? The ability to red flag situation. This better understanding is going to help us enforce and apply data quality policies and standards. Standardization, cleansing, verification, matching and then applying our linkage or our survival of records from all those duplicate records. And then that critical monitoring step. Again, comparing the baseline to our results to identify our progress. This is going to go all way towards also validating the success of your data quality efforts. It can help definitely show results to executives and focus on the line of business, so they know that the things are getting better.
19:37 slide 14	Along with cleansing information, you can also think of cleansing, as on slide number 14, as a step of transformation. But the notion that we want to be able to filter, aggregate, summarize, and transform data from sources to various targets for all of these different information intensive solutions that we talked about earlier. From a business perspective, they're seeing more information and more sources. They want to be able to handle expanding of velocity or higher velocity of information, but also provide consistent integration across systems, so we get to that single or correct version of the truth. So the solutions that we need to be successful in governing information like this include easy to use user interfaces for designing these types of processes, tools like our Blueprint Director that we saw in the warehouse example, to be able to provide end to end project visibility, whether you are representing business or IT, it provides a blueprint we can all work from. Of course, dealing with the volumes, variety and velocity of information, we use solutions that can support massive scalability and exploit parallel processing with multiple processor environments. And also, solutions that can transform data from virtually anywhere and deliver it to just about anywhere, in batch or real-time, in a variety of different flexible capabilities.
21:11 slide 15	Once we've done this transformation, here on slide 15, we will talk about delivering information, so the actual timely, reliable delivery or movement of data, regardless of where it came from or where it's going to. On the business side, we see challenges with the cost of capturing

	changes from data sources and of course delivering the outputs of those changes. Also lengthy batch processes or long batch windows often hold up business workflows. So we need solutions that can perform and deliver information at the speed of business. Certainly the timeliness and also the trust or the quality of the information is very important as well, as we are delivering information. So the capabilities we need to look for are solutions that don't add adversely effects, source systems from a cost perspective or even a processing perspective. That can shorten that loading process times, providing for that high performance data movement, regardless of whether it's batch processing or in real-time. And then of course deliver information to a broad range of sources and support a broad range of data delivery styles.
22:24 slide 16	So this is all supported on slide 16 by a variety of different platform services or capabilities, that allow for this timely delivery of trusted information. So as we are coming to terms with multiple desperate data source, we've got direct and easy access to all that business-critical data. We are going to play shared understanding of information to have more confidence in the quality of data, and of course that's for measuring the quality of that data, we can identify how we are doing over time. And then of course supporting and expanding business objectives and growth, whether we're talking about opportunities with mergers and acquisitions and coming to terms with that acquired data, or new business ventures. We need to be flexible, we need to be adult and adaptable to be able to provide trusted information. So the capabilities required for this include connectivity to a variety of different sources and targets, native connectivity whenever possible to be able to provide unmatched data scalability, and flexibility in deployment and reuse of information tasks. I've talked about that earlier with the notion of wrapping up any sort of functionality around quality or transformation in web services, and then delivering them across your organization. So you can make those standards and policies available to web sites, to other applications, all through a flexible modular integration approach with service oriented architectures. And of course we need to be able to support the big picture, the information governance and integration initiatives, by making available common metadata infrastructure, that as we talked about before, allows the ceding of that critical information over time.

	And then of course scalability, to build with the growing needs of data and also exploit processing power that exist within organizations today and tomorrow.
24:20 slide 17	Here we are on slide number 17, let's talk about metadata. Put simply, metadata is defined as data about data. Well that's kind of a strange definition if we think about it. A good analogy I think is if you look at the information on say a product label, like a can of soup. Things like the name, the ingredients, the volume, so in other words, the data about all the attributes about that thing. So as we break it down a bit further, we think about data in a couple of different contexts, from a business perspective, or the consumers of that information, or from a technical perspective. So there is definitively two sides of the coin, the business users have their own terminology and definitions etc., and then the technical users typically talk about data sources or databases, rows, columns, and other technical attributes. These two groups need to collaborate. We need to create a shared understanding of our information assets that span the technical representation all the way out of how it's used by the business user. So these two sides need to work together to create that shared vocabulary.
25:34 slide 18	Slide number 18. What's the big deal with metadata? If you talk to the analyst, if you do some reading and research on metadata, it's definitely a hot topic. There's a lot of business drivers for better managing this data about data, or metadata. Because without that, we can't communicate and collaborate across application silos, product silos, within the organization. We need a shared vocabulary, shared understanding of what data is, where it came from and what happens to it as we seek to deliver trusted information. So all of these different driver can have significant impact for metadata management.
26:16 slide 19	Slide number 19. Organizations need tools that are based upon the role of the user that we talked about earlier, put simply the business and IT users. But let's dig down a bit further into the types of roles that can benefit from the shared metadata approach, whether we're talking about business users that are looking at reports, they need to be able to trust the information they're looking at. So by having this shared metadata repository, we can provide powerful tools like our Business Glossary, that allow a user to look at something in a report, maybe they're looking at a report and they see a term called "current revenue" and they don't understand what it is or where it came from, or maybe it's just something they don't trust the number behind it.

	With powerful Business Glossary, they can look at that term, identify the data steward, in other words who owns that information, and also look at the quality attributes of that data. So what data quality rule apply to that, what are the maximums and minimums, are there a red flag on that information? All of this is enabling that shared view or that shared single metadata repository. This can also extend to subject matter expert so the focus is to understand data quality rules for example, providing powerful tools that not only allow us to understand the quality of data and cleansing, but also a better understanding of that. So if we profile some data we got from an acquired company, and we see that social security numbers is not a valid primary key for that data, in other words it's not unique, we can collaborate on this data to be able to make better decision about how we cleanse it, transform it and deliver it, from the subject matter experts to the architects, and the data analyst that are usually doing this type of integration. Then we have a better understanding of the requirements, we have a better understanding of where the data came from, and what happen to it from source to transformation to target. So this unified metadata management capability can go a long way towards simplifying a streamlining integration tasks, in other words reducing the risk of these types of projects. You can also go a long way towards helping with change management. One of the requirements for information governance is that notion of agility and adaptability, to be able to meet the requirements of the business. Because we're applying data quality best practices, we can also go a long way towards increasing the trust and confidence in information, and making sure that we're complying with the regulatory standards.
28:56 slide 20	Let's move to slide number 20. So, we can get some great benefits from more effectively integrating information, with applying this notion of information governance to what we're doing with understanding, cleansing, transforming and delivering information. Organizations can respond quickly to compliance requirements. And this is a big deal, several years ago, and it still is today. But this was a big catalyst for driving organizations to grow their information governance initiatives to become better at managing information across the enterprise. Of course, applying these best practices can help with increasing revenue opportunities, allow organizations to make the most of upturns in the economy, but also deal with challenging times, and rescue runaway costs so we're more effectively managing information, we're getting rid of duplication of effort, redundancies and process, and of course risks that could affect your organization

	very significantly.
30:00 slide 21	So the solution that enables this is IBM InfoSphere Information Server, as we see on slide 21. It's an information integration platform that can enable information governance. So it provides everything you need to basically integration heterogeneous information, data from a multitude of different sources, transform, cleanse and deliver it wherever and however it's needed. So it's a platform that includes those core capabilities of understanding, cleansing, transforming, delivering, provides a shared metadata repository across this entire process, has built-in scalability to be able to come to terms with those huge volumes of information. Data quality can be native to the fabric of any data integration processes. We can also provide 24 by 7 access and high availability and failover capabilities to make sure that your information gets delivered without fail. By leveraging these shared services and common metadata, we can go a long way towards enabling collaboration and aligning business and IT objectives. And the shared services also include some optimized connectivity to be able to access more data faster, from a variety of different data sources and deliver it to a variety of different targets for all those information intensive applications. So Information Server is a unique offering from IBM that allows clients to effectively enable information governance, as they're integrating and delivering information.
31:35 slide 22	Let's move to slide number 22 and our summary here. Information integration and the processes that we've been talking about in this session can be a great catalyst for information governance for your organization. It can help leverage information as an enterprise asset, and we need to start treating information as such to be able to improve and ensure better business outcomes. Any information governance initiative requires a deep broad understanding of your organization's information assets. So we need to get our arms around our data, figure out the structure content quality. And while we're talking about quality, we need to make sure that we're doing data quality and applying best practices for data quality, and the best way to do that is to get to the root cause of the data anomalies. that's going to be your best cost implication, the best back for the buck so to speak. Information integration done correctly can deliver trusted information from a multitude of different initiatives, and we talked about some of the types of projects that can benefit from this approach.

	And last but not least, the importance of metadata cannot be emphasized enough. It could make a significant difference in how your organization understands its information assets, is able to contend with change over time, and identify whether or not you can trust the information you're looking at. So it's very very critical for information governance initiatives.
33:00 slide 23	Slide number 23, how can you get started? First off, make sure that you're collaborating between business and IT. So if you're an IT, go talk to your business users, ask them if they trust their information, ask if they understand where the data came from. I think you might fine some surprises. We need to make sure that they do trust the information running in the systems, in these reports, in these applications, they were providing as IT professionals. The next step, convene a working group. that could be a cross functional group representing business and IT of just a couple of folks, or a larger group within your organization. Once you had those initial discussions, it certainly wouldn't hurt to have some management sponsorships. So it might be a good time for recruiting executive or identify a management sponsor for your project, whether it's simply a data quality initiative, or if you're applying some of the more advanced capabilities around integrating information that we've been talking about in this session. And of course, contact your IBM information management sales representative if you have questions, if you want to be able to see some of the solutions that I've been talking about, such as InfoSphere Information Server and the industry leading capabilities that it provides. Of course, for more information, go ahead and check out our ibm.com/software/data/infosphere.
34:21 slide 24	So, slide 24, it's my last slide, I want to thank you for your time and listening to this webcast. If you do have any questions, definitely contact your information management sales representative, and we'll be able to hopefully help you more effectively apply best practices around information governance and integrating information for better business outcomes. Thank you.