Government Information Sharing & Advanced Insight and Analytics

The National Information Exchange Model (NIEM)
Government Agencies & The Information Sharing Challenge

Today’s Mission and IT leaders find themselves stuck between intensifying threats on one hand, and stovepiped information management systems on the other hand that limit their ability to respond. Clients tell us they have reached a tipping point -- they know they need to transform the way they leverage information. They need information agility, where they are able to easily share trusted information as a strategic asset across compartment and across agencies for mission optimization.

Perhaps nowhere is the information challenge more evident and acute than in the area of Law Enforcement and Homeland Security. In a post 9/11 world, agency leaders find themselves struggling to leverage and share the vital information they need to detect threats and prevent crime and terrorist attacks. While fictional TV shows and movies portray police, fire, and emergency management personnel as having ready access to all the information they need to make an arrest or prevent the terrorist plot, the truth is less comforting. Department, agency and mission leaders often struggle to leverage their own information in a meaningful way, let alone share it with their peers across the government ecosystem.

Recently this challenge was underscored once again, with the attempt by the Nigerian Umar Farouk Abdulmutallab on 25 December 2009 to ignite a PETN Incendiary device aboard Northwest Airlines on approach to Detroit’s Metropolitan Wayne Airport. What is evident from post incident analysis is that enough information existed across the Homeland Security and Intelligence communities prior to 25 December to raise enough red flags to prevent his boarding the plane. However, the agencies involved suffer from a sort of “Enterprise Amnesia” in which relevant data assets are scattered across multiple departments, applications and databases and essentially unknown or inaccessible to the enterprise. Simply put, current data integration and information sharing architectures make it tremendously difficult for agency leaders to understand the assets they have, and to establish and deliver the Trusted Information they need to spot the relationships and patterns obscured by the huge volume and complexity of formats in which the source data resides. This challenge is only magnified when the agencies attempt to share information outside their individual enterprise.

Quite simply, straightforward and seamless Information sharing within individual agencies and across the government ecosystem is at the core of improving our system of justice and defending our homeland.

To support our clients through the transformation, in early 2006, IBM launched its Information On Demand (IOD) initiative. Through IOD, IBM is focused on enabling clients to unlock business value of their information assets by freeing them from legacy database and application silos, integrating and cleansing the information to deliver Trusted Information, and then providing it as a service to the people, processes, and decision points that need it to optimize outcomes. IBM has invested heavily in support of its IOD initiative. The company has invested close to $12B in the acquisition of more than twenty (20) Information Management companies, has dedicated 10,000 industry practitioners to supporting our clients, and has developed more than three hundred (300) industry solution assets to help accelerate value, reduce cost, and mitigate client risk. Since 2006, we have acquired 2000 business partners, and have helped over 10,000 clients realize breakthrough value from their information assets.

Through IBM’s investment in Information On Demand, the technology platform to support real time information sharing has matured greatly. What has been missing in the Government space is a standardized vocabulary – a standard set of terms and definitions for the data that needs to be exchanged across the government ecosystem. The good news is that the National Information Exchange Model (NIEM) promises to provide this missing piece, by defining the language for government information sharing.
Making this transformation can seem like a daunting task and it is often difficult to know where to start. IBM understands that in today's environment, leaders need to deliver a rapid return on investment; they need light weight, low risk solutions to their tactical challenges, but within the context of a larger investment strategy for transformation and competitive advantage. IBM can help agencies get started with both the tools and a methodology to convert information into a trusted strategic asset which can be shared beyond your tactical business processes. This approach allows agencies to deliver value to the mission, and then leverage that investment for even greater levels of return. IBM has the experience it takes to guide this transformation and has invested in the capabilities to help you leverage the NIEM standard and join the information sharing community.

The National Information Exchange Model (NIEM)
In the post September 11th world, US Law Enforcement and Homeland Security agencies found themselves stuck between the intensifying requirements to share information across compartment, missions and agencies, while dealing with information dissemination rules, processes and systems that simply did not lend themselves well to this type of real-time collaboration. While mission rivalry and plain old territorialism certainly played some part in limiting information sharing, the most fundamental barrier to communication was the total absence of a common language with which to communicate. Each Agency is comprised of multiple lines of business, each with its specific set of information systems, vocabulary and data – each in the language of that particular mission. This lack of common language prevented all but the most rudimentary information sharing. Where agencies did achieve this rudimentary capability, they were forced to develop a shared vocabulary and language to do so.

Finally, in 2005 DOJ and DHS came together around a common mission to develop a standardized language for communication. Mission leaders would no longer be forced to develop a specialized language ad hoc, every time they wanted to share information. Rather, once they mapped their own internal vocabulary to the new standard – thus “translating” their own language to the standard – they could communicate seamlessly with any agency who had also done their own mapping. The intent of NIEM, then, is to become a sort of universal information translation engine – a Rosetta Stone. This common language accelerates the business value of sharing, while reducing the cost and risk of building one-off integrations. Missions no longer need to work out formats and terms -- NIEM has them predefined. The separation of terms, definitions, and formats from the way data is stored in individual systems gives NIEM its information sharing agility.

NIEM provides a consistent, extensible, maintainable set of XML-based data elements and types that represent the business requirements of the communities involved. It also provides longevity to the NIEM data model because the data is actually stored in a database and is thus not tied to XML schema but can be used to generate new data model formats (like the Web Ontology Language) if required.

The NIEM data model consists of a core set of components and a set of domains that extend the core: -- Immigration, Emergency Management, Justice, Person Screening, Intelligence, International Trade, and Infrastructure Protection. These represent the specific communities of interest, or extensions of the core, through which the domain communities want to communicate domain-specific concepts. For example, the justice community includes law enforcement, prosecutors, courts, and corrections agencies at all levels of government. Each domain community has its own governance process, and thereby evolves their domain model independent from the NIEM Core.

Since its initial development and piloting in 2005, the NIEM standard has undergone a rapid set of enhancements – including over a dozen releases and numerous homeland security and justice pilot projects. In July 2007 and with the release of NIEM 2.0, NIEM transitioned from pilot projects to an operational program. This transition has brought a new
level of momentum to the standard. NIEM 2.1 was released this year and both refined some existing domains and adding some new domains including a healthcare domain.

**IBM Capabilities and the IBM NIEM Solution Framework**

The IBM Information On Demand initiative consists of three platform layers that support the full lifecycle of structured and unstructured data. The first layer is a set of integrated tools to support the management and integration of both structured and unstructured data and content. The next layer consists of a unified platform to help mission leaders create, manage, deliver and analyze trusted information – that is, information that is Accurate, Complete, In Context, and which delivers real business insight; this platform allows information to be managed outside of the scope of individual applications. The third layer is a platform of integrated tools that allow business users to gain new insights into their data via through business intelligence nad predictive analytic models. The breadth of the IBM Information On Demand Solution portfolio is considered by many, including Gartner and Forrester, to be the most comprehensive and cohesive solution in the market today.

For Government Agencies, especially those involved with regulation enforcement, law enforcement and national security, information sharing is a critical component of their information management responsibilities. In 2005, the Office of Management and Budget released Version 2.0 of the Federal Enterprise Architecture (FEA) whose primary purpose was to “enable information sharing and reuse across the federal government via the standard description and discovery of common data and the promotion of uniform data management practices.”

As a framework for information sharing the underlying Data Reference Model (DRM) provides a model and pattern for government agencies to create an information sharing architecture. Figure 2 depicts the key components of that information sharing architecture.

![Figure 1: The IBM Information Management Solution Space](image1.png)

![Figure 2: NIEM and the FEA DRM](image2.png)
As depicted in Figure 2, the FEA DRM architecture is composed of three subject areas: description, context and sharing. The diagram was used by OMB to brief the DRM to congress as part of their Expanding e-Government report to congress. The red marks on the diagram highlight how some of the DRM concepts translate into specific implementation practices, software components and artifacts. Not all of the DRM components are mapped as, in this document, we want to focus on those practices that culminate in the development of a NIEM Information Exchange package which is a formal definition and description. The DRM is an abstract model of how an organization's data architecture can be focused to deliver and increase information sharing. In the description area, the data assets of an organization must be formally described and those descriptions captured in either data models or content management systems. The red circle around the structured information descriptions shows where enterprise data modeling plays a role in the information sharing architecture. An enterprise data model harmonizes the many physical data models (derived from physical databases) into a logical data model and further abstracts it into a conceptual data model. Those models must be captured in an enterprise metadata repository which brings us to the context layer of the DRM.

The context layer, therefore, becomes the focal point for information discovery and revolves around the population and search of a metadata registry/repository that houses an organization's data asset inventory. The data asset inventory can be browsed via a taxonomy to connect producers and consumers of a particular type of data. In the context of NIEM, those data assets are the source of information that will populate the exchange messages that will be sent to trading partners. It is the creation of those messages, populated with data from data assets, where NIEM comes into play. NIEM provides a framework and reusable set of components for rapidly generated exchange messages and documenting those messages in Information Exchange Package Documentation (IEPD). In this light, and at its most simple form, NIEM represents a method for satisfying the output side of the information sharing architecture. Effectively, NIEM is the visible part of a much larger architecture and set of processes required to deliver that output. This relationship between NIEM and the complete information sharing architecture is depicted in Figure 3.

In Figure 3, NIEM manifests as the visible part of the information sharing iceberg. Below the surface a non-exhaustive set of activities and artifacts is depicted in producing or supporting the production of those NIEM exchange messages. Some of these are obvious, for example a data quality process is necessary to guarantee that the data that populates those messages can be trusted. Secondly, a glossary is the means to define the fields of the data in the data assets populating those messages. Internal standards help in integrating data sources across the organization and a NIEM message may be from multiple sources to create a single transaction. As an XML exchange message, a NIEM message may be processed by a service in the organization's SOA architecture. A NIEM message may be composited on the organization's master data (to get a single version of the truth).
or even be used in a federated query across business units or across agencies.

Thus, we see that the key challenges for an organization in creating NIEM-conformant exchange messages is more than just the expertise to assemble the messages but in establishing and mastering the “rest of” the architecture (the iceberg).

To assist federal, state and local organizations to succeed in establishing a complete, NIEM-compliant information sharing architecture, IBM is focusing and integrating its product offerings into a NIEM Solution Framework. The IBM NIEM Solution Framework takes a holistic approach to create a sustainable, repeatable and flexible enterprise information sharing environment. There are three axes to the IBM NIEM Solution Framework:

- **End-to-End Information Sharing Architecture** – the Infosphere trusted information platform offers an integrated set of capabilities that enable mastery of data from profiling it at its source, to describing it, enriching it with metadata, cleansing it, transforming it, integrating it and delivering it to its destination. In that framework, the NIEM-conformance component comes in at the tail end of this end-to-end process. This solution framework focuses on the development of the complete, integrated set of capabilities for sustainable information sharing.

- **Streamlined IEPD Development** – this solution framework will enable customers to more easily develop Information Exchange Packages by lowering the barrier to entry for data architects new to NIEM. The solution framework breaks IEPD development into a fine level of granularity (even below the level of the IEPD process promoted by the NIEM PMO) and addresses each sub-activity and the tool support required to streamline the effort.

- **Information Sharing Analytics** – in the beginning of the section, we saw that the Information On Demand initiative includes the end-user functionality that directly unlocks the business value of information in the form of better decisions, better insight and better efficiency. Following that model, the NIEM solution framework must also focus on the exploitation and unlocking the business value of information sharing by providing tools to better exploit the shared information. IBM has taken the lead in business intelligence, predictive analytics and business rule engines and non-obvious relationship processing to deliver just that.

**End-to-End Information Sharing Architecture**

As discussed in the previous section, the assembly of a NIEM compliant message actually takes place at the tail end of a broader information sharing architecture. In order to create a sustainable information sharing process, an organization needs to address the complete architecture. Accordingly, the IBM NIEM Solution framework leverages the holistic and unified InfoSphere platform for establishing, managing and delivering Trusted Information. The InfoSphere Software portfolio brings together a set of four key building blocks, unified through a common Metadata layer, and represents the industry’s most comprehensive information integration and sharing platform. The platform, depicted in figure 4, supports the ability to harvest, enrich and disseminate all data in an organization no matter where it resides, its format or its quality. Taken as a whole, the platform provides a line-of-sight solution addressing the entirety of the information supply chain, from the data sources all the way to the information exchange package and every step in between.

![Figure 4: The IBM InfoSphere Information Server Platform](image)
The IBM InfoSphere Information Server Suite is based on a unified, service-oriented architecture (SOA) that provides the complete set of capabilities needed to understand, cleanse and extract, transform and deliver information from across heterogeneous systems to any target. InfoSphere automatically analyzes source data, applies data quality and complex transformation, rules and algorithms, and then delivers the result via the widest range of formats and models. Its unique, metadata-driven design helps to align business goals and IT activities, providing a consistent understanding of what data elements mean, capturing business specifications and using them to automate development tasks, and providing deeper insight into data by tracking its lineage. IBM InfoSphere Information Server has native connectivity to the broadest range of data sources, including databases, applications, files, and more on any platform. Its parallel processing foundation allows it to easily scale to meet the needs of even the most demanding environments, working consistently across batch and real-time processing. InfoSphere Information Server improves the productivity of development teams working on data integration and data sharing projects, improving the collaboration across these development teams, and between the developers and the business users who set the requirements. In addition, the common metadata repository maximizes the reuse of data models, ETL and cleansing and transformation jobs within and across project teams, to provide ongoing value.

By eliminating the duplication of functions, the architecture efficiently uses hardware resources and reduces the amount of development and administrative effort that are required to deploy an integration and sharing solution. By using the IBM InfoSphere Information Services Director, Government Agencies will have the ability to publish data integration tasks as reusable services in an SOA environment. The IBM WebSphere Service Registry and Repository enables sharing, publishing, finding, managing, and governing services and policies across your entire enterprise environment.

- Publish web services without coding
- Execute data integration tasks in real time
- Improve accuracy and completeness of all data
- Create shared services for on-demand data cleansing from any application, business process, or portal
- Ensure consistent enforcement of governance across all uses of data
- Utilize the best data available in the enterprise for operational decision-making

In the context of the information supply chain, the IBM InfoSphere Information Server establishes, manages and delivers trusted information through four primary steps:

- **Understand** or profile organizational source data from both technical and business perspectives as it exists within current systems and databases. The IBM InfoSphere Information Server Suite supports business and data analysts in understanding the content, quality, and structure of data sources by automating the data discovery process. With the IBM InfoSphere Information Analyzer module, Government Agencies mitigate risk by uncovering missing, inaccurate, and inconsistent data early in data integration project lifecycle.

- **Cleanse** the source data. The cleansing process not only includes the repair of bad data, but also the standardization of terms and the elimination of redundant records by using our industry-leading, probabilistic matching engine. The IBM InfoSphere Information Server Suite will deliver accurate, trustworthy, standardized, and unique data to business users across your enterprise.

- **Transform** data to meet the formatting requirements as established by organizational standards. The IBM InfoSphere Information Server Suite transforms and enriches data to ensure that it is in the proper context for new uses. With IBM InfoSphere DataStage, Government Agencies have access to hundreds of pre-built transformation functions to combine, restructure, and aggregate information on a visual design desktop workspace.

- **Deliver** the new information via whichever format or data model the target system requires. It is in this stage that InfoSphere Information Server would create the NIEM compliant message. This capability ranges from loading quality data to multiple systems and data stores, to data virtualization and changed data capture in real-time to
support disaster recovery and Continuity of Operations (COOP) initiatives.

The IBM InfoSphere Information Server Suite provides Government Agencies the ability to virtualize, synchronize, and move data to the people, processes, or applications that need it. The platform provides direct, native access to a wide variety of data sources, both mainframe and distributed. It provides access to databases, files, services, and packaged applications, and to content repositories and collaboration systems. With the InfoSphere Federation Server module, Government Agencies can create a consistent, single view of enterprise data that resides in multiple sources in order to support key business processes and decisions. The delivery capabilities provide support for many different data sources and data targets. Whether the initiative is creation of a distributed operational data store on any relational database management system, or the requirement is to populate a Data Warehouse, or simply to transmit a NIEM compliant message, the IBM InfoSphere Information Server components will deliver the data where it needs to be and when it needs to be there, even for near-real-time applications. IBM also provides the analytical tools needed to ensure that end users can extract business value from this new, Trusted Information, utilizing tools such as IBM Cognos Business Intelligence.

Finally, with the InfoSphere Metadata Workbench module, Government Agencies will be able to explore a visual Web-based browser of metadata generated and used by the IBM InfoSphere Information Server Suite. IT staff are able to explore and analyze technical metadata about sources of data, business metadata for data meaning and usage, and operational metadata that describe what happens within the integration process itself.

Using this end-to-end architecture, we have the ability to translate the understanding of your agencies data assets into outputs suitable for packaging via the NIEM standard. This brings us to the second pillar of the IBM NIEM Solution Framework, an integrated process and tool support for streamlined Information Exchange Package Documentation (IEPD) Development.

Streamlined IEPD Development

Government agencies can develop and document an information exchange through either a “Top Down” or “Bottom Up” approach. In the former, the agency would start with the information sharing requirement and develop a message from a clean sheet of paper; in the latter, the agency starts with the information that is currently available in one or more existing data sources, and creates the message upon this foundation. Both approaches have certain practices in common, especially in relation to the mechanics of construction of the message that will be shared between the organizations. In fact, most organizations focus their IEPD development on the message construction phase as it sets aside thorny issues like data ownership, quality and access. Additionally, this is the process that is best documented by the NIEM Program Management Office as depicted in Figure 5.

![NIEM PMO IEPD Lifecycle](image-url)
The NIEM PMO’s IEPD lifecycle diagram, known as the “fishtail diagram”, offers a top-down process for defining information sharing requirements, mapping those requirements to the components available, building a message, documenting and publishing the message for reuse. This lifecycle is “organization-agnostic” as it cannot and does not want to assume any reliance on the data architecture of the organization that wishes to create an IEPD. Unfortunately, this lack of understanding or leveraging of the organization’s existing data architecture is a critical flaw and Achilles’ heel to IEPD development.

The IBM NIEM Solution framework addresses this directly that by recognizing the importance of existing architectures and available data components, and leveraging this understanding in its augmented version of the IEPD development process. Figure 6 depicts the four phases to the IBM IEPD Development process.

**Component Discovery and Selection**

Component discovery refers to the ability of an organization to rapidly relate four things together: their current architecture and data elements; the needs of their exchange partner and themselves in processing an exchange; and the availability of NIEM data elements that match both those needs and their existing data elements. By far, this is the largest hurdle for organizations developing IEPDs.

The IBM’s solution strategy for this is depicted in Figure 7, and joins the Top-Down discovery process with the Bottom-Up, thus enhancing the value of each. Top-down discovery would start with NIEM components and the raw requirements as define in the NIEM PMO IEPD process. A bottom-up discovery process would focus on mapping existing data elements from profiled data sources to NIEM data elements. The IBM approach brings together the NIEM components and existing data elements in the Infosphere Metadata repository and, and then selects the pairs based on transactional needs. In other words, the transaction is defined in the middle at that point where requirement is matched to available data, and from this definition the message is formed.

The four phases of the IBM IEPD Process recognizes the complexity of source data architectures and the difficulty in transforming existing data into IEPDs. The IBM process is specifically designed to tackle and improve the difficult areas of IEPD development. We will discuss each phase in detail and how IBM tools can be are applied and in some cases enhanced to support the streamlining of this process.

**Figure 6: IBM IEPD Development Process**

**Figure 7: IBM’s Middle-Out Component Selection**
A number of tools supporting the “Understand” step of the IBM InfoSphere Information Server can be applied to support the component discovery process. Two modules in particular are focused on understanding data from the bottom-up and top-down. The bottom-up analysis is supported by Information Analyzer, which accesses enterprise data sources and profiles them to understand data elements, their quality and business rules. Top-down analysis is supported by Business Glossary and Business Glossary Anywhere. The Business Glossary module imports business terms and categories from a number of different sources, including XML Schema like the NIEM Components. Figure 8 is a view to the Business Glossary and a notional mapping of its term to data elements in a physical data model.

Figure 8: IBM Business Glossary maps Terms to Data Elements

In this view of Business Glossary (top left), we see a hierarchy of categories and a business term on the right (“Balance Sheet Net Assets Approach Analysis”). Three key aspects of the Business Glossary product enable it to support robust component discovery. First, the module allows business users to create and maintain the glossary of terms -- business users execute the mission of the organization and therefore drive the need for information sharing exchanges. Secondly, the Business Glossary Anywhere Tool provides the ability for those business users and managers to access and use the enterprise business glossary from anywhere in the organization - right from their desktop or handheld. Third, the Business Glossary tool relates IT assets including business intelligence reports, data elements from various data sources, and even other terms (like those from logical data models). Currently, IBM is constructing a Proof of Technology (POT) to demonstrate importing and mapping NIEM business terms to related IT assets profiled by the Information Analyzer tool.

In addition to the Infosphere tools, related assets like IBM Optim provide tools that can aid in the discovery, understanding and mapping process. Additionally, IBM labs are participating in the Open Integration project (http://openintegration.org) to provide advanced mapping and visualization tools.

Once the components have been selected for the transaction via both the bottom-up and top-down approaches discussed above, the team can move on to the next phase of IEPD development – Message design.

**Message Design**

Once a government agency has selected a messages component parts, it can design the actual message. The chief complaint organizations report for in this area is that the process is labor intensive and takes too long. The driving factor for this is the need to understand the systems, on both sides of the exchange, that will process the message. This is further complicated in the event either of the organizations has implemented a Service-Oriented Architecture (SOA). Further, the interaction might require data integration prior to being able to populate the message. Given this, message design requires a focus
broader than the simple data elements, taking into account many other parts of the entire enterprise architecture. This notion is reflected in the Federal Enterprise Architecture as depicted in Figure 9. In the Federal Enterprise Architecture, the Data Reference Model interacts with the other four reference models. Most important of those interactions is its connection to the Business Reference Model, via business processes, and the Service Reference Model via services.

Simply put, message design involves layers of the enterprise architecture beyond the DRM, because messages are necessarily part of business transactions. They do not exist in isolation. As part of business transactions, they serve a purpose in a business process. The business logic for those business processes is often implemented in a service component which is part of a service. A service has a message input and a message output. These are the very messages that need to be designed. To design these messages effectively requires an understanding of these components their interaction. Once again, the breadth of IBM’s vision and portfolio of assets helps address this challenge. IBM has tools for Business Process Modeling, Service Oriented Architecture Implementation, data integration and Master Data Management. Whether you have IBM tools for these parts of the architecture or not, the transaction design must include these multiple perspectives. For example, Figure 10 depicts how InfoSphere Data Architect can import and model XML Schemas.

InfoSphere Data Architect can import NIEM as well as perform modeling functions for UML, ER diagrams and Business Process Models. Visualizing models is important to better understand them and how they relate to other models. Message design goes far beyond just selecting the components because agencies must arrange them in patterns like request/response, command/action, event/response and many others. Additionally, agencies must design the messages for cascading actions, dependencies and transaction metadata.

Once the message is designed it can go into production.

**Message Production**

Message production can be split into three sub-parts:

- **Pre-production (or testing)** - message testing is an important part of pre-production. Message testing involves creating dummy messages, testing corner cases, and testing processing (including SOA) interfaces. To facilitate testing, the IBM Optim component has the ability to generate anonymized test data.
• **Message population** - after successfully testing the message design, the messages are ready to be populated with real data. This is the critical reason why the component discovery process cannot be just a top-down approach -- in the end the messages must be populated from the existing data assets. As discussed in the first section, this is where IBM’s end-to-end information sharing architecture provides distinct value, in that the agency now has a complete profile and understanding of all available data assets; now it is merely a matter of selecting the right elements. Pulling information from these data assets is easily performed by Infosphere’s DataStage module as depicted in Figure 11.

**Figure 11**: Infosphere DataStage transporting and transforming data

As depicted in the diagram, the DataStage module lets you graphically design data flows with various aggregation, calculation and merging steps between the inputs and outputs. The data sources are the inputs to the flows and there can be many different data sinks, including a web service that produces NIEM-based XML. Data Stage provides graphical icons for different steps like merge, sort, code lookups, aggregation and many others. Additionally, in addition to transformation processing, agencies can add QualiStage processes into the flow for pattern matching, standard lookup, address correction, and many others. The Infosphere DataStage and QualityStage modules are the cornerstone of a robust message generation process by way of extracting, integrating and transforming data from existing data sources.

• **Message processing** - once messages are populated with real data and flowing between systems, the business logic that processes incoming messages (and responses) can be tested (and even stressed) on the sending and receiving systems. Besides interface testing this would also enable the ability to test identity management, message security and message privacy.

Once agencies can reliably produce and process messages in a production environment they are ready to support the mission. In supporting the mission, the information exchanges and the instances of those exchanges need to be managed and allowed to evolve for changing conditions.

**Message Evolution and Management**

After the message is in production, agencies must manage the message flows and allow the message schema to evolve for changing conditions. Once again, IBM’s assets assist in both parts of this phase. IBM has robust products to store and query XML messages like DB2 PureXML and DataPower XML appliance. Additionally, the IBM Metadata Workbench allows agencies to capturing and view lineage information for message consumers that want to know the pedigree of a message. The IBM Metadata Workbench can provide queries and visualization of lineage as depicted in Figure 12.
In addition to base level message management, agencies may need to evolve messages as conditions change. IBM’s XML Schema Repository (XSR) and Rational Clear Quest enables straightforward change management.

We have seen that the IBM streamlined development process consists of four phases; component discovery, message design, message production and message evolution. Additionally, we covered how IBM’s broad suite of products can help organization successfully execute each phase. We have also discussed in some instances how IBM is continuing to evolve their product suites and conduct Proof of Technology demonstrations to improve how their tools can support the process.

In addition to streamlining the development of messages, IBM’s products extend even further to the analysis of information exchanges to improve the exploitation of that information. Such improved information exploitation leads to better decisions, the detection of non-obvious relationships, real-time business rule execution and even predictive analysis.

**The IBM NIEM Solution Framework & Advanced Analytics**

Core to IBM’s Information On Demand Strategy and the concept of “Trusted Information” is the transformation of disparate, inaccurate and seemingly conflicting data elements into valuable information, providing new, incremental or even breakthrough insight. Information On Demand is about understanding and cleansing agency data, and bringing it together in a way that reveals relationships in the base data elements that were invisible or at least obscured before simply because the data was spread across the enterprise -- in disparate applications and databases, in a wide array of formats and semantics. At its most basic level, Information On Demand is about “connecting the dots” between disparate data sets to reveal hidden relationships and patterns.

While IBM’s early focus centered on providing new insight through the delivery of Trusted Information within a single enterprise, the development of information sharing models like the NIEM standard allow government agencies and the government ecosystem as a whole to open the aperture well beyond the single enterprise, to reveal relationships and patterns within the totality of the information that is shared. The emerging ability to share information across the ecosystem promises breakthrough – the more information that is shared, the more relationships and patterns are revealed, the more accurate and “Trusted” the information becomes.

Consider this – By the time Umar Farouk Abdulmutallab purchased his ticket on 16 December 2009 on KLM and Northwest Airlines from Lagos through Amsterdam to Detroit, on which on 25 December 2009 he attempted to ignite a PETN incendiary device, no fewer than 5 US and UK
government agencies and 3 private entities (the airlines and the Yemen university from which he had recently dropped out) possessed derogatory information that taken together might have raised a warning. Were leaders at each of these agencies derelict in their efforts to protect the lives of those on Flight 253? Hardly – the men and women who serve in these positions are largely diligent and steadfast. Rather, each organization suffers from the challenge of “Enterprise Amnesia” in which they fail to understand the volume of data they have, let alone recognize the relationship of these data elements to others inside their enterprise or across the ecosystem.

Critical to addressing the “Enterprise Amnesia” and larger “Ecosystem Amnesia” challenge, and thus provide critical insight through information sharing, is the concept of persistent semantic context, through a centralized and common metadata repository, such as the common meta data layer provided by IBM's Infosphere Platform. This metadata repository enables discovery and new insights, while allowing the information “stewards”, whether a department within a government agency, or another government agency altogether, to retain ownership of the data elements themselves, and therefore policies regarding access. This persistent semantic context allows events and observations to become discoverable, because they are in context with other events and observations.

As an example, consider the following two diagrams, in which 3 disparate agencies produce three separate observations about three seemingly different entities. In the first diagram, reconciling two observations in the metadata store reveals no relevant relationships. However, when a third observation is added to the persistent context engine, new relationships are revealed – three seemingly separate entities are recognized as one. In this way, The IBM Infosphere Solution Framework provides need-to-know capabilities where the data “discovers” the user through the registration of user profiles and persistent, context-based searches. The search results are updated as new metadata is added to the common metadata repository. Data finds data, and the relevance finds the user – in an autonomic fashion.

Beyond the new insights revealed through development and delivery of Trusted Information, government agencies can also leverage advanced business intelligence and predictive analytics tools to glean additional knowledge and insights and improve overall mission effectiveness. Within the context of IBM's Information On Demand strategy, this capability is captured in the topmost “Business Intelligence & Performance Management” layer. IBM continues to invest heavily in this area with the acquisition of two companies, Cognos and SPSS. While Cognos represents a best of breed business intelligence tool providing state of the art “sense and respond” analytic...
capability, SPPS provides a robust set of analytic algorithms to enable agencies to “predict and act”. These tools are in wide use across the government ecosystem today, and leverage the same common metadata repository at the heart of the Infosphere Trusted Information platform. This allows government agencies to ensure they have constant visibility to the latest and most accurate information available across the information sharing ecosystem, and to review the sources and lineage of any mission critical information prior to taking action.

Achieving Transformation: The Information Agenda

Since the launch of our Information On Demand initiative in 2006, IBM has had the privilege of working with literally 1000’s of leaders across Defense and Civilian Agencies, as well as private sector companies, in transforming to information based enterprises in which information is leveraged as a strategic asset across the enterprise. As a result, we have learned a great deal about what it takes to be successful.

Transformation does not imply the replacement of existing systems and information sources; It requires only that those systems are able to expose insightful information with speed and flexibility. Making this transformation can seem like a daunting task and it is often difficult to know where to start. What is required for success is a combination of proven technology, and a proven approach to transformation – what we call the “Information Agenda”. With the right tools, the right processes, and the right environment, creating an Information Agenda is straightforward.

The key to building an Information Agenda is to bridge the gap between the information end users in your enterprise and information suppliers; between line-of-business (LOB) Mission Leaders who are facing enormous business challenges and IT executives responsible for providing information needed to succeed. It is critical for success to get these leaders in a room to start conversations on common understanding of business strategy, vocabulary and semantics. To facilitate and capture this conversation, you’ll need a prescriptive methodology and set of tools that:

- Help you identify which information is important to your organization;
- Create a common set of vocabulary to define this information for use across the enterprise;
- Allow you to make information available to the people and applications that need it;
• Assist you in managing your information;
• Help you create governance practices and processes that are needed to put an information management plan into action.
• Develop a roadmap for transformation that matches investments to Mission priorities

Through our Information Agenda approach, IBM helps our client get their arms around their information assets, develop a process for governance, define strategy and priorities, and plot the roadmap for transformation. We understand the pressures that government agencies face in moving quickly, while positing for long term optimization. We can assist agencies in bringing together the people, processes and technology necessary to navigate the new information exchange environment. The Information Agenda can help with broader agency information management needs and objectives as well, far beyond the NIEM standard.

The development and refinement of the NIEM standard promises to rapidly expand the level of information that is shared across the law enforcement, homeland security and emergency response environment, and opens up a breakthrough opportunity for agencies to enhance mission effectiveness. IBM has the experience and an integrated solutions to help agencies not only meet NIEM requirements, but also leverage the newly accessible information to improve overall mission effectiveness. Through innovation, government agencies can harness the new intelligence permeating the planet today—for smarter, more successful organizational outcomes.

About the Authors
Mike Daconta, Founder & CTO, AIM & Former Director, Enterprise Data Management Office, US Department of Homeland Security

Tim Paydos, Executive Consultant, Federal Government, IBM Information Agenda

Sujatha Perepa, Federal IT Architect, IBM

© Copyright IBM Corporation 2010
IBM Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
March 2010
All Rights Reserved

IBM, the IBM logo, ibm.com and InfoSphere are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml.

1 Office of Management and Budget; The Federal Enterprise Architecture Data Reference Model, V2.0; November 17, 2005.