

## What's Your Information Agenda?

**Mahesh H. Dodani**, IBM, U.S.A.

### 1 APPLICATION ARCHITECTURE

“As the information age rolls forward, our businesses, markets and societies are being transformed into adaptive, connected networks. The Internet of today only hints at the ubiquitous communication infrastructure of tomorrow. The construction of this brave new world requires a new kind of architecture, focused on digital structures of information and software rather than physical structures of bricks and mortar. As we spend more time working and playing in these shared information spaces, people will need and demand better search, navigation and collaboration systems.” – [Institute of Information Architecture](#)

Over my last four articles, I have laid a foundation for a Service Oriented Architecture (SOA) as the enterprise architecture of the [globally integrated enterprise](#) and focused on how to define and establish the business side of the enterprise through a well defined [business architecture](#). In this article, I continue our journey into the IT architecture side and focus on the information architecture. To reiterate the importance of the information architecture, we look at the key role information services play in the SOA capabilities reference model shown in Figure 1 required to implement comprehensive, enterprise wide SOA. The Development Services are used to implement custom artifacts that leverage the infrastructure capabilities, and Business Innovation & Optimization Services are used to monitor and manage the runtime implementations at both the IT and business process levels. At the core of the SOA Reference Model is the Enterprise Service Bus (ESB) which delivers all of the inter-connectivity capabilities required to leverage the services implemented across the entire architecture. Transport services, event services, and mediation services are all provided through the ESB. The Model also contains a set of services that are oriented toward the integration of people, processes, and information through Interaction Services which provide the capabilities required to deliver IT functions and data to end users. Process Services provide the control services required to manage the flow and interactions of multiple services in ways that implement business processes. Information Services provide the capabilities required to federate, replicate, and transform data sources that may be implemented in a variety of ways. The services in an SOA are provided through existing applications via the Access Services, in newly implemented components via Business Application Services, and through external connections to third party systems via the Partner Services. Underlying all these

capabilities of the SOA is a set of Infrastructure Services which are used to optimize throughput, availability and performance. IT Service Management Services include capabilities that facilitate the management and security of the deployed services, composite applications, and hardware/storage/network resources. These services also provide the capabilities to monitor the deployed environment to collect both technical and business KPI metrics and present these via appropriate “dashboards” for inspection by appropriate personnel and to help them take the necessary actions to optimize the managed environment or the business service/process.

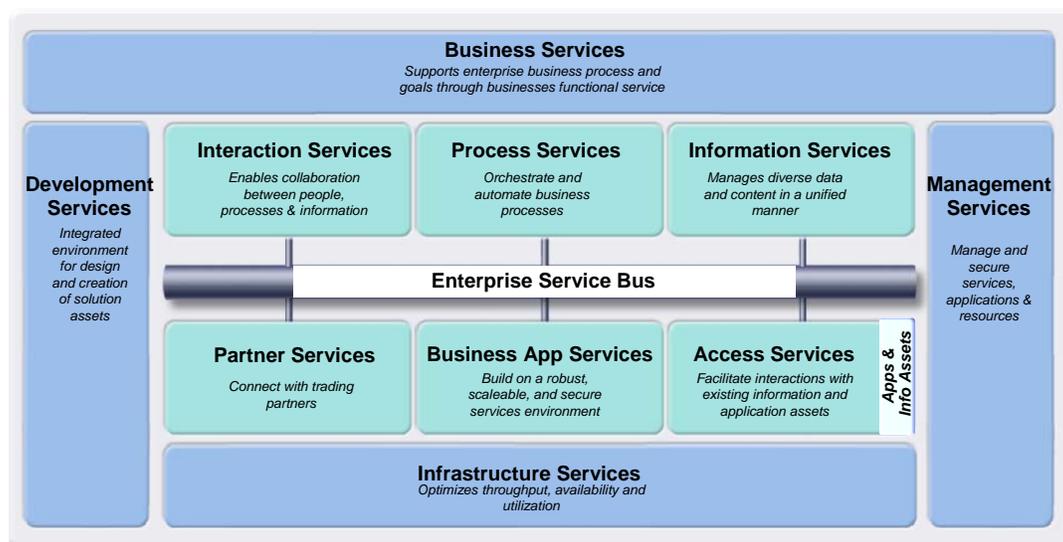
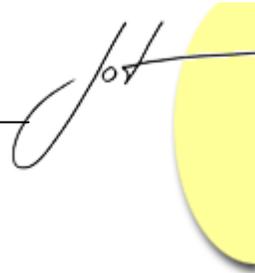


Figure 1: SOA Capabilities Reference Model

Note that the cornerstone of SOA and the supporting capabilities shown in Figure 1 is separation of concerns. SOA's initial focus has been on applications and as a consequence the separation of concerns principle has only been applied to the application architecture. The nature of web services made it very easy to expose application logic as services. Extensive tooling became available early on to wrap application functions and to make them accessible through SOA for a broad range of consumers such as business processes, portals, and other applications. As shown in the Reference Model, even the integration of the services is focused on routing of requests through the ESB to various applications and to orchestrate the execution of services through processes. In this approach, the application is the gate keeper to information and there is a 1:1 relationship between an application function and its corresponding persisted data. “Information services” are still viewed as simple data storage and retrieval mechanisms.

This approach forces a tight coupling of information to application which leads to various challenges:



- **Reuse:** Some information needs to be accessed by many applications, not just a single application. Different applications view the data in different contexts and not necessarily in an enterprise wide scope. Encapsulating all data within an application limits the use of the data to the context of that single application.
- **Accessibility & Distribution:** Many applications or consumers need to access information – structured and unstructured – from a variety of sources. We need to provide those consumers with effective mechanisms to transform and integrate distributed data and content.
- **Single Version of the Truth:** In many situations where data is distributed, we have islands of systems with incomplete and conflicting data. The consumer needs to see the single version of the truth through the service interface, not a set of conflicting options.
- **Access to Analytical Data:** Analytics is often done by looking at historical data after the fact, like looking in the rear-view mirror. However, in order to provide a more agile and flexible business architecture, we need to embed the analytical insight more directly into our processes and connect them closely with our applications.

The goal of information architecture in SOA is the design of comprehensive services that provide accurate, consistent, integrated information to business processes and people from existing legacy, inconsistent and diverse data. In IBM, we have established the Information On Demand (IOD) Architecture as our approach to help clients set their information agenda and to align information architecture with SOA. In particular, the IOD supports SOA by defining information services that expose trusted and integrated information from structured and unstructured sources to a broad range of consumers. Furthermore it improves overall (SOA) governance by aligning the approaches to include data governance as well. SOA, in turn, can support the Information Agenda for the enterprise through IOD by providing the principles and approaches to limit and control the information access so that it can be decoupled from consumers which results in better reuse. In the following sections, we explore the design of information services through a well defined Information Reference Architecture.

## 2 INFORMATION ON DEMAND REFERENCE ARCHITECTURE

Figure 2 shows the Information on Demand Reference Architecture which provides the foundation to design the information services needed to support SOA. The Reference Architecture supports SOA by defining information services that leverage appropriate capabilities and patterns. The Reference Architecture also provides a capture of the best practices in delivering these information services. In this section we will work our way up the Reference Architecture layers by discussing the business challenges that each component of the architecture is addressing, looking at the key capabilities and solutions

that are provided to address these challenges, and elaborating the benefits that these capabilities can deliver.

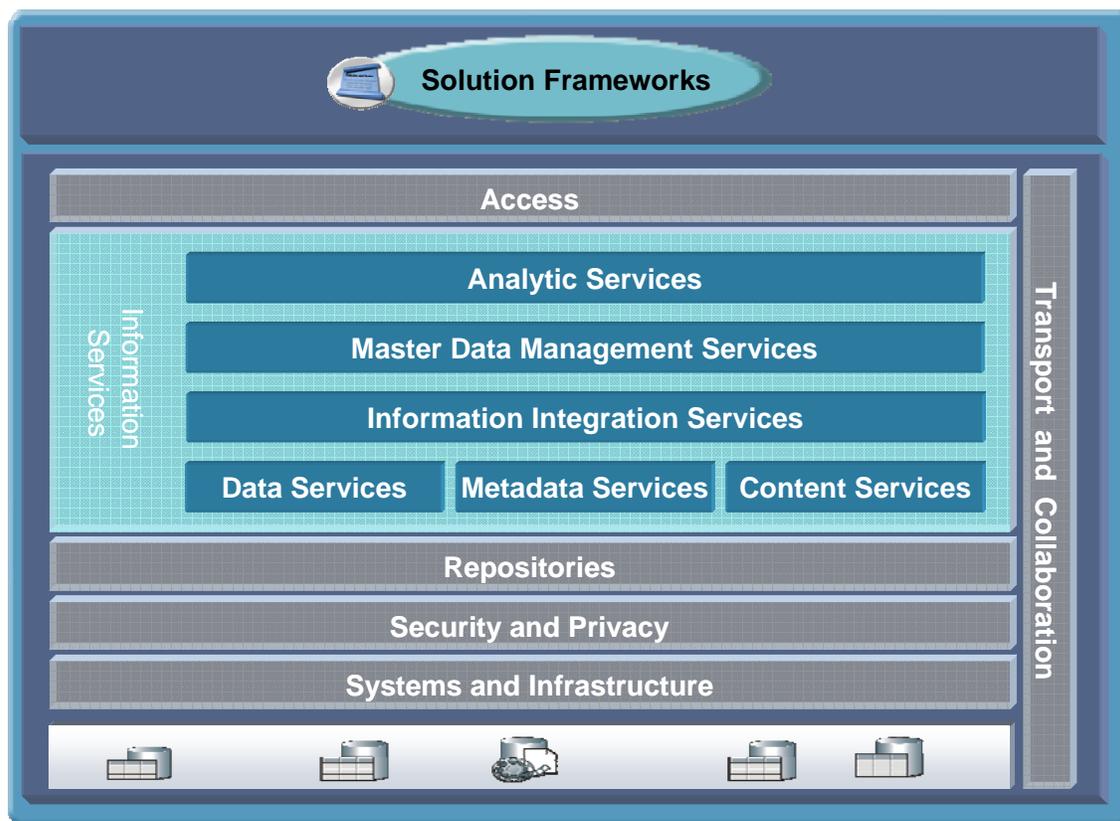
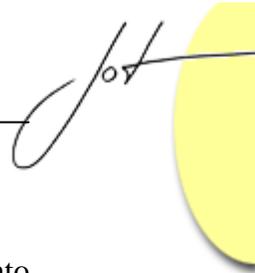


Figure 2: Information On Demand Reference Architecture

**Content Services:** Most enterprises have multiple content repositories that hold unstructured information. In most cases this diversity is not appropriately controlled. As a consequence, most of the enterprises' time is spent on searching for the information – around 70% according to some studies – rather than allowing personnel to perform their job. Many enterprises still rely on paper and manual processes to deliver some of their critical functionality. They lack formally specified processes that are automated in areas where it makes sense and that can be systematically controlled. They struggle with the amount of paper documents that are difficult to manage and cannot be easily queried to automatically extract information.

Content services decouple content consumers from the variety of content providers through a single interface for trusted and consistent access to unstructured data. This single interface can be registered in the service registry to facilitate access across and beyond the enterprise. These services also provide content-centric processes and



---

workflow technologies to most effectively manage unstructured data and to embed it into the overall business process.

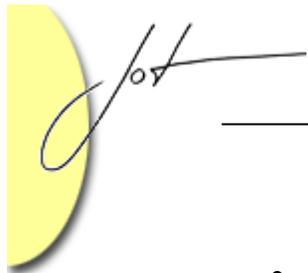
Applying these capabilities can lead to more efficient processes to manage unstructured information that are also aligned with the overall business process. This approach leads to an end-to-end optimization of business processes across various divisions, technologies, etc. Processes can access these services through a single interface that hides the underlying implementation and heterogeneity of content. Access to trusted content can be reused in a consistent way by multiple applications. As a consequence, a change to the underlying technology (e.g. adding new repositories, replacing them, consolidating them, etc.) has minimal impact to the consumers of the service.

**Metadata Services:** Various activities in SOA rely on metadata, such as data, process, service models, etc. Often, those artifacts are only available in proprietary formats requiring architects to spend a significant amount of time trying to find the information. Tasks are not aligned due to missing metadata information and become more and more inconsistent over time. Metadata services provide a foundation to manage metadata consistently and to share the metadata across relevant tools. Furthermore, in the context of SOA these metadata capabilities or the data itself can be exposed through a service interface. Applying these capabilities can lead to an improved and common understanding of the data, which in turn can lead to significant improvements in collaboration and worker productivity.

**Data Services:** Enterprises need to persist service transaction data in its native (e.g. XML) format. The challenge is how to uniformly and holistically manage, process and query enterprise data that is stored in different formats (e.g. relational and XML.) Data services combine the strength of traditional database technology that provides performance, integrity, protection and scale with the flexibility of XML when persisting and retrieving data. The concept is to manage both types of data in one uniform repository and to provide unified access and management for that data. The unified access to and management of XML and enterprise data leads to important business benefits including significant reduction in development time, significantly less lines of codes to manage and better performance.

**Information Integration Services:** Most organizations have disparate data sources with inconsistent and conflicting information. Bad information leads to bad decisions, and this is exacerbated when companies are not even aware of the data quality problem. If they are aware of these challenges, they often spend too much effort to manually correct the data. Information integration services provide capabilities to address these challenges in three steps:

- Understanding the enterprises' information is the first step in addressing the problem.
- Cleaning up the incorrect and conflicting information is a next logical step.

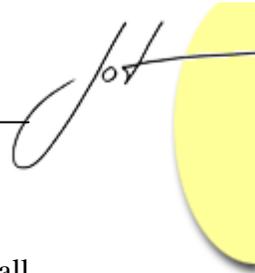


- Transforming and delivering trusted information as a service is the final step to provide this single consistent interface.

Applying these steps can lead to a significant increase in worker productivity and in the quality of decisions since trusted information is accessible to every user when they need it. Let us look at each of the steps in turn.

Understanding the enterprises' data has three major elements: consistent definition of business and technical terms, data assessment and profiling, and data modeling that is aligned with service and process models. In many cases, the business community (LOB owners, business analysts, etc.) and technical community use different terms and have a different understanding of its meaning. This problem is addressed by establishing a business glossary that allows us to define the terms and to share them across business and IT. We are often uncertain about the degree of data quality within the scope that has been defined. Do we really have consistent keys across systems, how is the data connected and how should it be connected, what are the data types that are defined and what data types do we actually have when we look at the data. Once we have a common understanding of some key terms, we also need to better understand our data and if it satisfies the integrity constraints and rules that we have specified. The data assessment or profiling performs this task: we try to discover inconsistencies, inaccuracies, anomalies in the data before we start realizing the services so that we're better prepared on how to provide trusted information and what cleansing transformation rules are required. If there is a conceptual or logical data model across various applications, it is most often isolated from the service and process model. That means, services are often specified in isolation and the input and output of the services are specified implicitly in message models that are not consistent with the conceptual data model. As a consequence, transformation may be required to map the data in the format of the message model into the format of the underlying systems. One of the outcomes of the data assessment is typically recommendations on how to better structure the data at the logical and system-independent from which we can build services. The purpose of the data model is to relate the data formats that exist in various sources with the data format that is required by the service.

Data cleansing improves the quality of data within and across databases but can also be leveraged as a service by applications and activities in business processes. These services address and resolve data inconsistencies and data redundancies. In traditional database data cleansing approaches we specify cleansing rules that define standardization formats, how to identify duplicates and how to resolve conflicts and enrich the information. Those rules are executed against data that is extracted from the sources and then applied against the same databases or new databases such as a consolidated warehouse or the master data management system. In the SOA context, we run the same cleansing rules against input that is coming from a service request and return the standardized data as the output of a service response. That means, that you can run the same cleansing rules against customer data that is stored as database records or through a



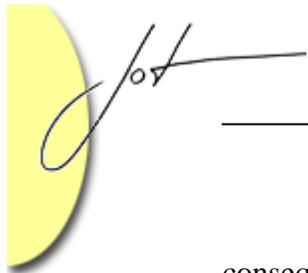
---

service that may be called 'cleanseCustomerData'. This service can now be used by all applications across the enterprise.

Transforming your data (in traditional terms extract-transform-load) fits into SOA by populating a consolidated data store from which data can be exposed as services and by aligning the process of when to populate the consolidated store with business processes. Data often resides in disparate data stores, complete information is not available due to a lack of integration, and the integration is implemented "manually" within the application which causes slow response times in particular for large data volumes and when complex transformations are required to merge the data into a consistent format. In the traditional database context, we apply the data consolidation and population pattern to extract the relevant data from the sources, to transform and merge the data into the integrated format of the target and then to apply or load it into the consolidated database. In most traditional use cases, it is an operational data store, a warehouse or data mart. Traditionally, the data population and consolidation process is triggered by a time schedule. In the SOA context, we are not just building a consolidated data store for database applications but also exposing some of this trusted information as a service. In many cases, a warehouse or integrated data store is the only place where an organization has established trusted information by aggregating the data. This allows the data consolidation process to be more aligned with a business process that then invokes the population of the target. The major contribution of this capability is to establish trusted information that can now be leveraged through a service by a wide range of consumers.

Data federation is another mechanism for delivering information aggregation focused on real-time data integration to expose integrated information as services. Traditionally, we have applied a data federation approach to provide a single system image of data. This data federation server provides data virtualization, which means that an application can access distributed data through the data federation server as if it was all in one place. The data federation takes the request and then calculates the required sub-operations that need to be sent to the various sources and merges the data from the sources into the integrated format that is then returned. In the SOA context, we can leverage this approach to most effectively implement a service that needs to access data from different places by exposing federated data through a service. The result of this approach is that you can most effectively implement real-time access to distributed data through a service which guarantees time-to-market and high performance. This approach is in particular flexible when you need to add new data sources.

**Master Data Management (MDM) Services:** Master data are the core entities that are most critical to the success of an organization like customer, product, account, etc. In most organizations, master data resides in almost all applications in an inconsistent and incomplete manner. As a consequence, the organization lacks the single version of the truth for the most critical information, i.e. master data. Furthermore, we have many "consumers" of master data that try to access the master data from these heterogeneous repositories in different ways with redundant and inconsistent logic and technologies. The

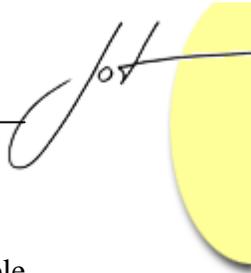


consequence is high cost to access the data that is most critical in an enterprise and also the cost to maintain the data and the access.

Master data management services use information integration capabilities introduced above to address the challenge of data inconsistencies and incompleteness. It provides a trusted source of information through a repository. It provides services on top to manage and query the information effectively and to keep it consistent with the existing legacy systems. As a first step, the master data entities that are stored in an inconsistent and incomplete manner in the various legacy systems need to be consolidated into a master data repository. An important aspect in this data consolidation is the cleansing of information and the specification of cleansing rules that can be applied across the enterprise. The MDM services provide consistent and extensible services to access and manage trusted information which can be accessed in a highly concurrent and scalable fashion by multiple consumers. It has basic services to secure the information and to provide consistent access methods. It leverages the cleansing rules to ensure that updates to the master data repositories are checked against the same rules in order to avoid degradation of the data quality when updating the data. It has event management capabilities to trigger other applications when certain events occur on master data elements. It includes hierarchy and relationship management because master data entities do not have a flat structure but may have more complex relationships like organizational hierarchies. Note that the MDM system is not a read-only consolidated repository but is accessed in an operational fashion. This allows it to also be used to make changes consistently in one place. The transactions to the master data and the data itself need to be synchronized across the various legacy systems and the MDM system itself.

MDM provides trusted information in a consistent and scalable platform that can be leveraged from multiple consumers. It improves the organizational effectiveness and allows the entire organization to leverage and reuse access to consistent, trusted master data.

**Analytic Services:** Organizations want to access and analyze heterogeneous data sources in real-time in order to predict market trends, understand customers better, increase operational efficiency, ensure compliance to regulations, and derive new insight into their business. The overall challenge is a lack of business-related insight into key metrics (that are either not available or not consistent), of early warning indicators, of identifying conflicts and uncertainty. By exposing analytic insight through services, we allow this information to be more tightly integrated into various applications and processes. These services are defined on a scalable foundation based on warehousing technology and leverage the information integration services to pull the information together. The analytic services then enrich the information and allow this information to be accessed by various consumers. Dynamic warehousing makes it easier for IT organizations to support business requirements for actionable information. They provide information (including unstructured information) with analytical intelligence to help people take action and make decisions. Unlike historical data warehouses and business intelligence approaches, dynamic warehousing delivers immediate, integrated information to empower users by



unlocking the value of information and analytic capabilities. Through a flexible architecture, it can serve multiple applications and lines of business for both strategic planning and operational purposes.

In summary, the Information Architecture is key to aligning IT to business which is in turn an integral part of ensuring successful SOA adoption in the enterprise. It is important for an enterprise to have a holistic Information Agenda that is supported through a well defined Enterprise Information Architecture, rather than looking at the information needs of their applications. The Information Architecture must address all aspects of the enterprises' information agenda from how to access and manipulate heterogenous data sources and formats, to integrating this information to be accessed in a consistent way, to managing master data across the enterprise in a consistent way, and being able to analyze the data in real-time to support the business. We have discussed the key components of the Information Architecture, and the capabilities it needs to provide to support SOA. In the next article, we will complete our journey into the IT side with a focus on the Infrastructure Architecture.

### About the author



**Mahesh Dodani** is a software architect at IBM focusing on Cloud Computing. His primary interests are in enabling communities of practitioners to design and build solutions that address complex business needs and deliver value. He can be reached at [dodani@us.ibm.com](mailto:dodani@us.ibm.com).