OPEN DATA CENTER ALLIANCE℠ MASTER USAGE MODEL: INFORMATION AS A SERVICE REV. 1.0

TABLE OF CONTENTS

CONTRIBUTORS

Matt Estes – The Walt Disney Company
Moty Fania – Intel IT
Jason, Li-Yi Lin – Taiwan Stock Exchange
Ivy Lin – Taiwan Stock Exchange
Shawn Ramsey – The Walt Disney Company
Bala Rasaratnam – National Australia Bank
Mick Symonds – Atos

LEGAL NOTICE

Open Data Center Alliance[SM] Master Usage Model: Information as a Service Rev. 1.0

Executive Summary

The combination of Big Data (driven in part by mobility and social media) and cloud computing is exponentially expanding the amount and types of available data while at the same time growing business demand for near real-time information is compressing the time available to process and use that data. This presents both challenges and opportunities for solution providers and consumers of information. Information as a Service, defined as the ability to provide standardized and secure methods to create, manage, exchange, and extract meaningful information from all available data in the right format at the right time, is one of these opportunities.

The benefits of Information as a Service include:

- Dynamic ability to acquire information and access business insights through the orchestration of information delivery from multiple data sources and in multiple formats.
- Reduction of cost, time, and complexity of sharing data stored in multiple locations.
- Standard methods that enable data consumers and producers to interact easily, consistently, and efficiently.
- Separation between the consumers of data (end-users, applications, or other services) and the data that they use or produce.

Interestingly, the same forces that are driving the rapidly evolving data landscape are also enablers of Information as a Service. For example, organizations are using Big Data technologies in conjunction with existing data warehouse (DW) infrastructures to process greater volumes of data on shorter time cycles. These solutions have been fuelled further by the greater access to information enabled by mobile solutions. Cloud computing has the most dramatic effect on Information as a Service, because it has propelled businesses, consumers, and applications to interact with each other across organizational, geographic, and technological boundaries.

In this document, the Open Data Center Alliance[SM] (ODCA[SM]) explores the overall information architecture required for Information as a Service implementation, and what aspects of that information architecture can be enhanced by Big Data, mobile applications, and cloud computing technologies. Specific aspects of Information as a Service covered in this document include data management, tools and technology that support Information as a Service, governance, and service delivery architecture.

This Master Usage Model is designed to help the industry move toward interoperable Information as a Service solutions. It can guide organizations in their implementation of Information as a Service. It can also help solution providers understand enterprise requirements relating to Information as a Service features and functionality.

DEFINITIONS

Table 1 defines terms used throughout this document.

Table 1. Terms and Definitions

Term	Definition
Data	The lowest level of abstraction from which information is derived; for example, an address
Information	A combination of contextualized data that can provide meaningful business value or usage; for example, the postal address of a business client
Information as a Service	The ability to provide standardized and secure methods to create, manage, exchange, and extract meaningful information from all available data in the right format at the right time
Orchestration	The on-going ability to arrange, coordinate, and manage the automated deployment and configuration of one or more interrelated components required for delivery at a point in time

PURPOSE AND AUDIENCE

This Master Usage Model defines requirements, methodology, and governance for different scenarios of Information as a Service use cases–including Data as a Service (DaaS), which supports data acquisition and distribution.

This document serves a variety of groups, including but not limited to:

- Business decision makers looking for cost-efficient ways to maximize the value of information. For example, publishing information to internal and external stakeholders to increase engagement, satisfaction, and – where appropriate – create revenue-generating services with information.
- Risk management and security operation teams seeking a means to improve data controls for sensitivity, loss prevention, and data quality management, in complex information architecture ecosystems.
- IT groups involved in planning, design, operations, and procurement that want to improve solutions from the cost/time efficiency of project delivery and on-going operations perspectives.
- Solution providers and technology vendors seeking to better understand customer needs and tailor service and product offerings.
- Standards organizations involved in defining standards that are open and relevant to end users.

SCOPE

This Master Usage Model is predicated on there being a provision of data, possibly a Data as a Service (Daas) model. This Master Usage Model extends and adds value to the DaaS model, and may well identify specific requirements from it. The following topics are addressed in this document:

- **Overall information architecture.** A high-level description of capabilities to improve information architecture complexity such as a sample information architecture framework; enterprise data model; and catalog for data sources and information platforms.
- **Data management.** A discussion of standardized access to data through the data supply chain. This includes data sourcing and collection; data cleansing and enrichment; data delivery; and security, availability, redundancy, and backup/restore requirements.
- **Tools and technology.** An assessment of the existing tools and technologies that can support Information as a Service.
- **Governance.** A discussion of data change management, intellectual property management, data ownership, and accountability–particularly with respect to how these topics affect external data and cloud service providers.
- **Service delivery architecture.** A discussion of Service Level Agreements (SLAs) and their management as well as the delivery processes, standardized client interfaces, staff roles, skills and capabilities, and support structures between multiple suppliers, including incident monitoring and management.

FOUNDATIONAL ELEMENTS THAT ARE OUT OF SCOPE FOR THIS DOCUMENT

There are a number of foundational elements that are not directly part of this Information as a Service Master Usage Model, but are critical to the concept as a whole:

- Underlying "as a Service" facilities
- Data security
- Regulatory requirements and compliance

Many of these foundational elements have been addressed in prior Usage Models published by the Open Data Center Alliance℠ (ODCA℠).[1] Except for a brief discussion in the following subsections, this document does not reiterate the detailed information found in those Usage Models. However, for a full understanding of the requirements for Information as a Service, it is important to refer to those Usage Models as necessary. A thorough awareness and understanding of these foundational elements for Information as a Service will make implementing the necessary information architecture easier and more efficient, and will help to create a more comprehensive solution.

Underlying "as a Service" Facilities

It is presumed that underlying facilities such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) are understood. Although it is possible to construct an Information as a Service ecosystem without a foundation of cloud computing, the end result will not be as feature-rich, responsive, or efficient.

- **IaaS.** These cloud services offer configurable virtual servers and storage, and companies or individual users pay for the capacity they use, which can fluctuate as needed. Example: Amazon Web Services. A full discussion of this foundational element can be found in the "Compute Infrastructure as a Service" Usage Model. IaaS provides the ability to allocate compute resources to existing pockets of data, rather than moving the data to the compute resources. This is critical when addressing the volumes of data companies are dealing with today.
- **PaaS.** The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.
- **SaaS.** Here, the host provides everything – infrastructure, software, data storage, and the user interface. Because everything is in the cloud, users can access the services from anywhere. Examples: Web-based email services such as Hotmail*, Windows Live Mail*, and Gmail*. SaaS extends the information acquisition, storage, processing, and analysis beyond the systems and capabilities that organizations have deployed. SaaS also provides the option for organizations to use services only when they believe they can obtain adequate return on investment — avoiding capital- and time-intensive technology deployment.

Data Security

Another foundational element for Information as a Service is security. The ODCA has published a collection of Usage Models that deal specifically with the topic of security for cloud computing. We highly recommend that organizations attempting to implement an Information as a Service ecosystem refer to the ODCA's Secure Federation Usage Models,[2] which are listed in "Further Reading."

Regulatory Requirements and Compliance

Regulatory compliance is an important element when dealing with data and information. The ODCA has produced two relevant Usage Models in the area of compliance:

- The **Regulatory Framework Usage Model** helps organizations assess and monitor their regulatory obligations when acquiring cloud services. Even if companies do not choose to use many cloud-based services for their Information as a Service ecosystem, the Regulatory Framework can be of assistance in navigating the complexity of regulation across the global marketplace.
- The **Commercial Framework Master Usage Model** can help drive efficiency and value in the standardization of commercial dynamics involved in the cloud-based data lifecycle. This can be particularly important to the Information as a Service ecosystem if a company chooses to utilize one or more cloud service providers in delivering the capability. Additionally, this Master Usage Model can help establish a master service agreement between suppliers and consumers, a task critical for creation of such a complex ecosystem.

---

[1] See www.opendatacenteralliance.org/ourwork/usagemodels.

[2] www.opendatacenteralliance.org/ourwork/usagemodels#securefederation

INFORMATION ARCHITECTURE CHALLENGES

In an end-to-end information architecture, information delivery relies on the following basic components:

- **Data capture or generation point.** The data from critical front-line operations of an organization (such as customer origination) that is captured through web portals or generated by a specific operational process must be stored in data source systems in a timely manner. Also, appropriate quality controls for accuracy and security must exist, and the data must be made available for information creation.
- **Information platforms.** Appropriate information and storage platforms, such as data warehouses (DWs) and operational data stores (ODSs), are implemented to house integrated data and to transform/enrich it to produce valuable information for distribution to downstream and upstream processes. These processes include reporting and analytics. Where there are multiple information platforms, usage boundaries need to be established for each of the information platforms.
- **Reporting and analytics.** Appropriate information service delivery platforms (such as a reporting platform) should be established with appropriate tools and technology to present the information in many ways for end users so they can analyze and derive insights from it.

These three basic components, however, have become increasingly complex due to a rapidly evolving data landscape; as a result high-quality, agile, and efficient information delivery has become a real challenge in the industry.

Rapidly Evolving Data Landscape

Gartner, Inc., one of the world's leading information technology research and advisory companies, states that a "nexus of forces describes the convergence and mutual reinforcement of four interdependent trends: social interaction, mobility, cloud, and information. The forces combine to empower individuals as they interact with each other and their information through well-designed ubiquitous technology."[3]

These forces are changing the IT landscape with respect to data and information, and present several notable challenges in developing information delivery:

- **Data sources.** New data sources are emerging, such as voice, text, and web, with rich customer information. With growth in Big Data, the sheer volume, staggering velocity at which the data is being generated, and the variety and complexity of data is overwhelming traditional technology. The new data sources and Big Data technologies are further increasing the degree of change that needs to be factored into information delivery solutions.[4] In other words, understanding of the difference between traditional data warehousing and new techniques such as NewSQL, NoSQL, Apache Hadoop*, and MapReduce are important for companies to decide when and how to best integrate these new technologies into existing data ecosystems.
- **Data location.** Cloud computing is another key change that is challenging some long-standing data engineering and data management methodologies. It has introduced a significant shift in the way that infrastructure is deployed, applications are built, and software is consumed. Application developers desire to access information through a service interface, where the physical location of the data is irrelevant to the applications consuming the data. The growth in cloud computing along with external service providers and packaged applications, is creating a "boundary-less information ecosystem" where data capture, enrichment, and distribution can occur in multiple locations, and they are no longer limited to the systems within an organization.
- **Data timeliness.** Data analytics is rapidly moving toward real-time or near-real-time analysis of events and interactions data. Businesses do not want to wait for "end of period" to analyze data, they want to make timely decisions based on customers' current interactions. This has brought a whole new dimension of criticality to data latency in the context of information delivery.
- **Dynamic data activities.** Businesses are increasingly demanding easy access to data to experiment with and search for unknown business values and actionable insights. These activities are performed in exploratory environments. As a result, the industry as a whole is embracing agile development methodologies, resulting in less time to integrate and manage data delivery.

---

[3] Source: "The Nexus of Forces: Social, Mobile, Cloud and Information," Gartner, Inc., January 29, 2013.

[4] See "Open Data Center Alliance Usage: Data Security Framework Rev 1.0," ODCA, March 2013, at www.opendatacenteralliance.org/library.

Heightened Data Security and Quality Risks

Greater data accessibility, variety, velocity, and volume, combined with broader location and multiple sources of data, increases the risks and challenges of data security. These risks include data loss, privacy and confidentiality, and misuse of data for fraudulent acts. Similarly, data accuracy, completeness, consistency, referential integrity, and reconciliation become extremely difficult. New approaches, techniques, and capabilities are needed to comply with the regulatory and end-user requirements for data security and quality controls.

Given the open and interoperable nature of the Information as a Service Usage Model, particular attention must be paid to the security between connecting elements and to the security of the consumers of the information. While Information as a Service deals specifically with delivering information that can lead to business insights, many of the principles surrounding data for traditional service-oriented architecture (SOA) and cloud computing apply.

Increasingly Complex Information Architecture

In the past, an organization's information architecture was fairly simple. It typically included a DW that acquired data from multiple transactional source systems, integrated, enriched, and distributed valuable information to a few reporting platforms for end-user consumption and possibly to upstream and/or downstream systems for further processing. However, as shown in Figure 1, over time the enterprise information architecture has become highly complex due to the rapidly evolving data landscape.

These complexities make it difficult for businesses to rapidly access the right information at the right time to make appropriate and timely business decisions. In a global marketplace the speed of right business decisions is a critical factor for competing effectively. In fact, Professor Eric Brynjolfsson, director of MIT's Center for Digital Business Research, conducted research on 179 large publicly traded firms and found that companies that use "data-driven decision making" are about five percent more productive and profitable than their competitors.[5] Contemporary organizations are more social, and matrix collaboration is becoming more common than a strict hierarchy of organizational structure. The ability for groups within a company to leverage the applications, services, and information generated by other groups and other companies provides a competitive advantage.

The complexity of today's information architecture reflects the rapidly evolving data landscape. Big Data technologies, new frameworks such as Gartner's "logical data warehouse" (LDW), the needs of mobile applications, and cloud computing mean that organizations will be bringing more–not fewer–technologies together; they need a way to create a cohesive and comprehensive Information as a Service ecosystem that can deliver outcomes for the entire organization's needs.

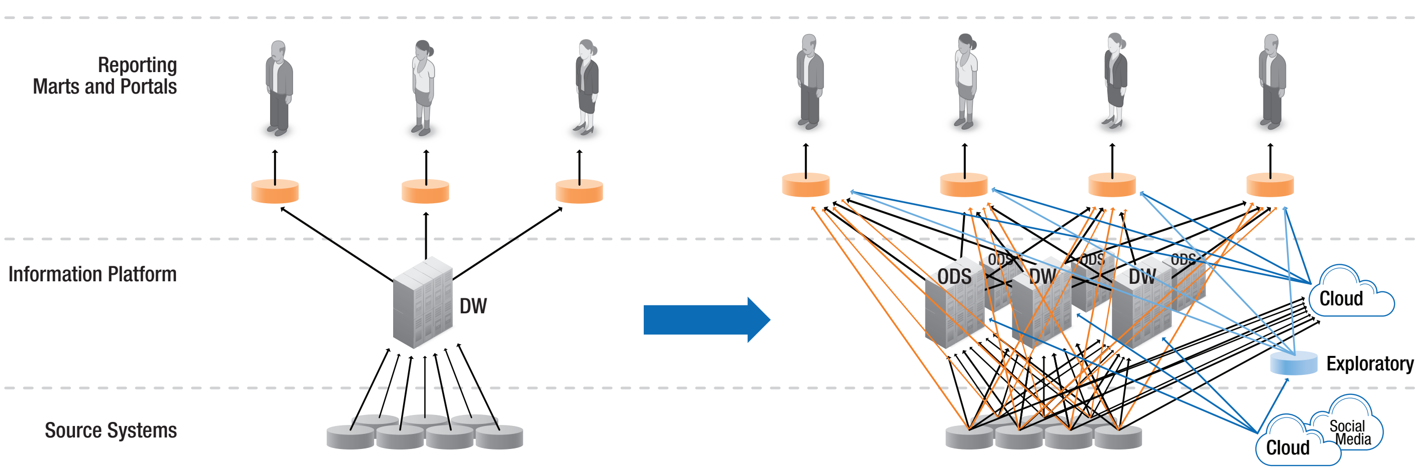


Figure 1. The Growing Complexity of Information Architecture

[5] Source: MIT Sloan Experts: Commentary on today's business issues, February 14, 2012. (http://mitsloanexperts.mit.edu)

Establishing Information as a Service

Information as a Service is a newly evolving concept to decouple the information consumer from the underlying complexity of the data landscape. It is a delivery mechanism to orchestrate the transformation of data to information and derive actionable business insights in a structured and streamlined manner, regardless of the type or location of data. In essence it removes the complexity of the data landscape from the information consumer–providing a positive experience. Although Information as a Service can be useful in traditional DW and ODS environments, it has even more to offer when used in conjunction with innovations provided by mobile technologies, Big Data, and cloud computing. For example it can leverage the ability of mobile applications to acquire and distribute data and information. Big Data presents new types of information for delivery, and cloud computing can establish cost-efficient service offerings and cutting-edge orchestration patterns.

Properly implemented, Information as a Service, supported by appropriate information architecture, technology, processes, and service delivery architecture can help organizations overcome the challenges of a complex data landscape as described above. Combining this with mobile applications, cloud computing, and Big Data solutions will unleash the power of data and information for organizations to outperform their competitors in the market.

Transforming Data into Information and Deriving Business Insights

Data–raw numbers in rows and columns without context–is virtually meaningless. Going beyond traditional data analytics and business intelligence (BI) means not only transforming data into information, but the ability to deliver that information at the right time to empower meaningful business insights that a business can use to drive profits, increase productivity, and enhance innovation. To “deliver the right data at the right time,” Information as a Service can leverage the benefits of combined traditional data management, Big Data, and cloud computing.

The transformation of data into information requires deep knowledge of business goals and needs. For example, if the business goal is to minimize customer churn, then the information garnered from the data should reflect that goal. In this example, the number of customers unsubscribed from a service is just a number. More meaningful information is created by contextualizing the data – such as discovering who the customers are, how long they have been subscribers, which services did they unsubscribe from, and do they subscribe to other services. This information can be used for contacting customers to understand customer motivation, improve engagement, remediate any issues, and identify other opportunities at the individual customer level.

Further business insights can be derived by pairing the number of unsubscribed customers with supporting data, such as survey results. For example, it might be possible to determine whether the unsubscription occurred in conjunction with other seemingly isolated events (such as a server outage or a competitor’s promotion). As this example shows, by combining different information and correlating it, Information as a Service can help consumers of information (business end users as well as customers/clients) derive actionable insights efficiently.

Figure 2 illustrates how data can be transformed into business insights, using a combination of enrichment, analytics, and presentation.

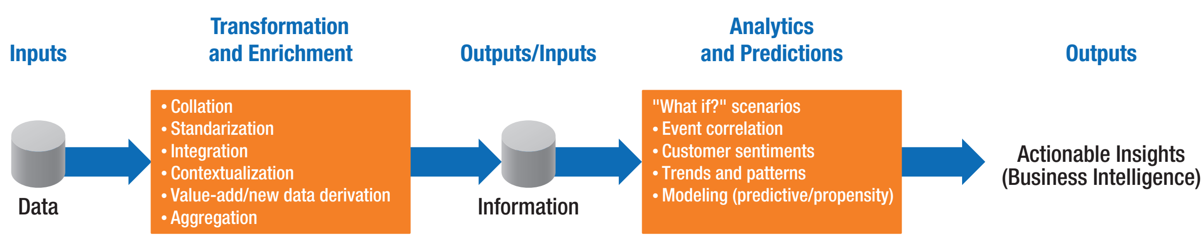


Figure 2. Transforming Data into Information and Actionable Insights

For Information as a Service to be sustainable and effective, it should have the ability to provide standardized and secure methods to create, manage, exchange, and extract meaningful information from all available data in the right format at the right time to meet the business goals. It is important to have a clear picture of what data exists and where it exists in the enterprise (see the discussions of "Enterprise Data Model" and "Catalog for Data Sources, Master Data Management, and Information Platforms" later in this document).

The transformation of data to information relies heavily on efficient integration of a variety of data, contextualization, enrichment, and aggregation. Presenting the information through appropriate visualization creates effective and impactful output. For example dashboards can illustrate the relationships between different parameters, using a wide variety of chart and graph types. Color-coding between graphs and charts can help the consumer of the information easily see these relationships.

Choosing the appropriate metrics, then setting thresholds and targets for those metrics, can help measure whether the information being created is the right information at the right time, for the right people. Consistent, reliable automation built into the system is an important technological component. Exploratory environments and new technologies can help organizations determine what the right combination of data and context creates the most useful information, and to refine techniques.

The Role of Cloud in Information as a Service

Although not required for implementing Information as a Service, cloud computing and data storage in the cloud present an opportunity to substantially transform information delivery. Cloud computing brings new capabilities to data and information management; therefore, technologies, tools, and components may not be housed within an organization's own data center. The cloud service provider can perform the orchestration function to gather data from multiple sources, both internal and external to the organization, and provide Information as a Service in an effective manner.

The distribution of services and solution components can be understood by examining the four primary cloud deployment models:

- **Private cloud.** A proprietary network or data center that supplies hosted services to an organization over an intranet (behind a firewall). It may be hosted on-site or outsourced, and managed by the company's IT team or a third party. Private clouds enable IT managers to install and maintain applications centrally rather than on each person's PC.
- **Public cloud.** The infrastructure and services are made available to the general public over the Internet, usually by an organization selling cloud services. Public clouds generally have intuitive front ends so that anyone can use the services without needing technical skills.
- **Community cloud.** A cloud environment that is shared by several organizations with common concerns, such as mission objectives, security requirements, policy, and compliance considerations.
- **Hybrid cloud.** A composition of two or more clouds (private, public, or community) that remain distinct entities but are bound together by standardized or proprietary technology that enables data and application portability.

Before establishing Information as a Service, organizations should carefully consider the role cloud services will play in their overall solution, and if needed, should carefully evaluate which cloud deployment model or combination of models would best fit their needs.

OVERALL INFORMATION ARCHITECTURE

Information as a Service requires a novel information delivery approach with new technology capabilities to achieve the following goals:

- Maximize the value of data for timely insights for decisions and actions—thereby enabling benefits such as increased revenue, productivity, and customer satisfaction.
- Drive down the cost of data and information processing and management.
- Reduce the operational and reputational risks caused by data quality issues.

Establishing the following foundational capabilities will help counteract the complexity illustrated in Figure 1, simplifying the information architecture and enabling efficient delivery of Information as a Service:

- **Information architecture framework.** In a highly complex data landscape this is required to define a target-state architecture and progressively simplify the current state in a structured manner. It also helps develop principles, standards, and design patterns for specific solutions. Finally, such a framework helps to define clear roles and accountabilities, a governance model, and solution process for end-to-end information delivery.
- **Enterprise data model.** In an era of data explosion, organizations should have a contextual view of the data that is critical to their business. An enterprise data model can help to consistently transform data into information, establish data standards, and facilitate data governance. An enterprise data model can include internal and external data, and a wide range of types and formats. Examples include customer data, operations data, and organizational data such as organizational charts and statements of policies and procedures.
- **Catalog for data sources, master data management, and information platforms.** This catalog provides a clear picture of what data exists in which systems and where it is located (internal or external). This catalog is required to help with solutions for information delivery and services, determining where the data should be sourced from, and where it should be integrated.

These foundation capabilities will help to progressively achieve simplified information architecture, such as the example illustrated in Figure 3, to efficiently support Information as a Service. (For real-world information architecture scenarios, see Appendix B: Sample Information as a Service Scenarios.)

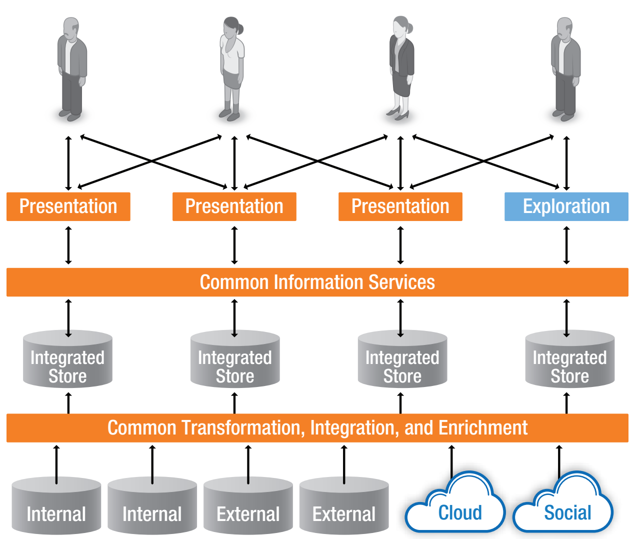


Figure 3. Simplifying the Information Architecture Using Information as a Service

Information Architecture Framework

In a highly complex "boundary-less information ecosystem," which includes multiple data and information systems that are internal to organizations, external, and in clouds, having an information architecture framework is a powerful tool to help simplify the overall information ecosystem in order to support effective delivery of Information as a Service. Figure 4 is a sample information architecture framework. The architecture simplification can be achieved by

- Mapping the existing systems, tools, and technology of the current-state information architecture (including what's external and in one or more clouds) to the various component layers of the framework. This mapping, based on the functionality or usage of a particular system or tool, helps the organization identify data redundancy and duplication, such as the existence of multiple customer master systems or multiple DWs. The mapping also can create opportunities for the convergence and decommissioning of systems, leading to architecture simplification.
- Identifying gaps in the current-state information architecture that need to be addressed for effective information delivery. Examples of such gaps could include data quality management tools, data security techniques, and metadata repositories.
- Defining an appropriate target-state information architecture that is simple and effective for the organization and building a roadmap in a structured manner, with clear business sponsors, IT/service provider accountabilities, and governance for delivering the target-state architecture.

The components in Figure 4 can be grouped into four logical categories:

- **Core Data and Information Components** (dark blue). This category represents the basic components that information delivery relies on in an end-to-end information architecture. It includes **Data Sources and Operations** that captures or generates data; **Integrated Information Platforms** that integrate, enrich, and transform data into information; and **Master Data Management** systems that maintain data consistently across the information architecture ecosystem.
- **Data Integration and Distribution Components** (grey). This includes **Data Movement and Integration** tools, technologies, and services that enable data to move between the Core Data and Information component systems, and the **Data Acquisition and Distribution** component, which consists of intermediary or staging databases that augment the data consolidation and distribution.
- **Presentation and Information Delivery Components** (light blue). This category consists of the service layers of the framework. They include **Information Services** that create various business insights from the information in the Information Platforms and present them in consumer-friendly views, and the **Data and Information Access** component that functions as a two-way interactive channel to capture data and also publish information views and insights to business and customer end users, through appropriate digital and mobile applications.
- **Control and Support Components** (orange). This category consists of components that control the data content and movement regulations in the rest of the components of the framework and enables data quality, security, consistency, service delivery assurance, etc., via Enterprise capabilities, protecting the Info Asset across the organization.

Refer to Appendix A: Information Architecture Framework Component Details for detailed description of the various components of the information architecture framework illustrated in Figure 4.

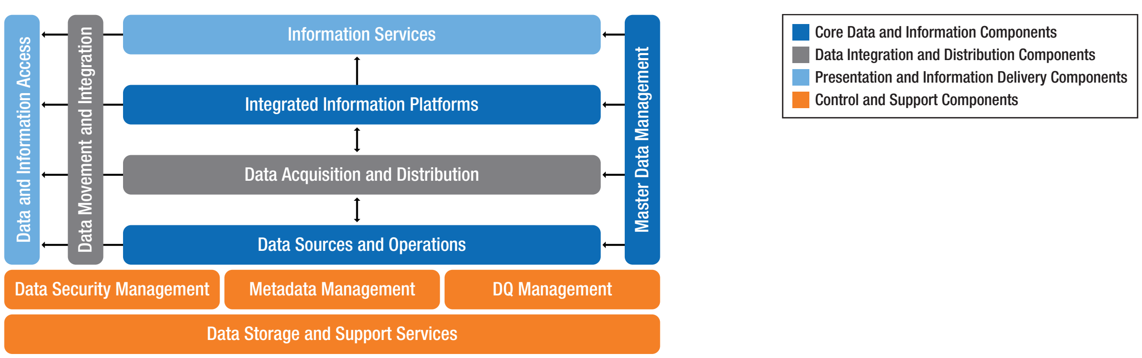


Figure 4. A Sample Information Architecture Framework

Enterprise Data Model

An enterprise data model defines the logical grouping of data and shows data interrelationships—thereby defining the data that are critical to an organization. Typically, an enterprise data model groups data into high level data domains (such as Customer), subject areas within each data domain (such as Customer Identity), and specific entities within each subject area (Customer Name). In general, some of the key data domains for organizations include Customer, Product, Events and Activities, Organization, Channel, Finance, and Risk.

Having an enterprise data model is a foundational step for enabling the following:

- **Consistent transformation of data into information.** The model defines what data is required to create a specific set of information or business rules.
- **Database design.** The model shows the logical data entity relationships, which are required to design physical data stores (such as DWs).
- **Data standards.** The model defines appropriate quality and security expectations. For example, it identifies sensitive data that requires masking or encryption; which data can be stored in a public cloud; and which data requires reconciliation controls for accuracy.
- **Data governance.** The model helps facilitate definition of data ownership and organizational priority for data that is critical for business operations or compliance and regulatory reporting. This enables an organization to focus governance efforts on the most important data and to define who should govern it—making data governance practical in a complex data landscape.

The enterprise data model can be developed in-house, adapted from industry models,[6] or developed by external service providers or consultants. Several enterprise data modeling tools exist with portal functionality to assist in developing models.

Catalog for Data Sources, Master Data Management, and Information Platforms

A data catalog contains information about what data is stored in which systems and managed by which owner (internal, external, or cloud provider). Having such a central catalog for data sources, master/reference data management assets, and information platforms can help to achieve the following goals:

- Increased efficiency in identifying systems and key stakeholders and service providers in scope for a specific information delivery service.
- Effective central management and governance of the design of information delivery solutions in the context of overall information architecture. This helps ensure consistency and enable change impact and dependency management, while supporting decentralized, scalable, and agile solution delivery.
- Reduced redundancy, duplication, and inconsistency of systems (at a more detailed level than Info Architecture Framework) by further rationalizing, converging, consolidating, and decommissioning assets.

A data catalog can be developed by mapping the information assets (operational data source systems, information platforms, assets in the cloud, and so on) in an IT asset registry against data subject areas and entities defined in the enterprise data model. For example, the catalog may state that "System X" is the authoritative source for customer identity data, while "System Y" is the reference data management system for organizational hierarchy data.

The data catalog can be a sophisticated IT management tool or as basic as a Microsoft Excel* spreadsheet to start with. It can also use a catalog that can be accessed at runtime, similar to the database catalog tables provided in many of the relational database management systems (RDBMS). Regardless of implementation, having a catalog of data with an appropriate change management process is vital to delivering Information as a Service.

---

[6] Examples include the Teradata's Financial Services Data Model (FSLDM) and the IBM Information Framework (IFW).

DATA MANAGEMENT

Data management is an important component of Information as a Service, because data management enables standardized and consistent access to data. By applying a standard set of transformations to the various sources of data and then enabling applications to access the data using open standards, service requestors can access the data consistently regardless of vendor or system.

Instead of building a custom distribution system for each business application, Information as a Service relies on an infrastructure that enables integration of back-end data sources and front-end information consumers. This infrastructure provides the ability to gather, process, and distribute content under a wide variety of new business models.

Figure 5 illustrates the five steps associated with the data supply chain, all of which are necessary to transform data to meaningful information and actionable insights (as discussed in earlier sections) that can add value to an organization. The data supply chain begins with data sourcing and collection, progresses through different stages of data standardization and management, and ends with information delivery.

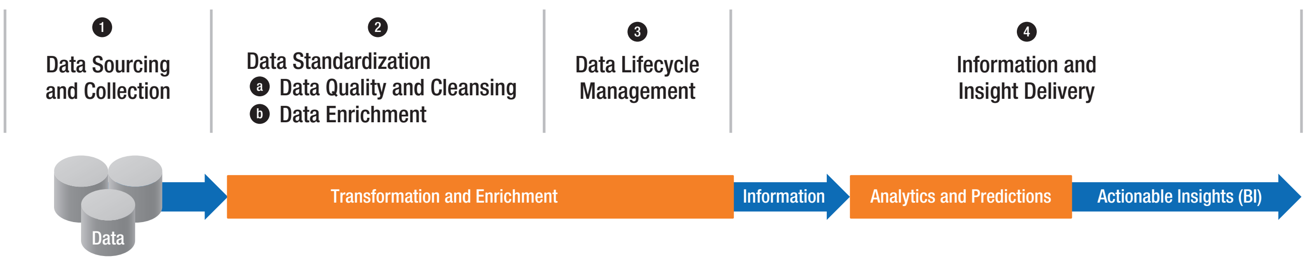


Figure 5. Data Supply Chain

The following list describes each data supply chain step in detail.

1. **Data sourcing and collection.** This step involves acquisition of data from appropriate sources and its ingestion. Due to the emergence of a "boundary-less information ecosystem," data resides within many systems and repositories and is becoming extremely challenging to integrate in real-time. End users want
   - Increased data accessibility and agility for information delivery
   - Improved data quality, security, and data-change impact controls to reduce regulatory, operational, and reputational risks
   - Reduced cost of delivery, support, and maintenance of multiple data interfaces

   Information as a Service uses concepts such as centralized data hub and staging databases to facilitate the data collection and consolidation, simplify the augmentation of data distribution, and make the data complexity transparent to end users. Even if data is stored in multiple locations across the enterprise, it is easily available for consumption, leading to lower costs and reduced time to access and obtain value from data.
2. **Data standardization.** This step involves the process of harmonization of data, enriching it, and cleansing based on business needs. Two sub-steps exist at this stage—data quality and cleansing, and data enrichment.
   a. **Data quality and cleansing** refers to identifying incomplete, incorrect, inaccurate, or irrelevant parts of the data and then replacing, modifying, or deleting dirty data to reach a higher level of data integrity and consistency. The quality issues handled during this step may have been originally caused by user entry errors, by corruption in transmission or storage, or by different data dictionary definitions of similar entities in different data stores. In some cases, this step attempts to fill gaps and add missing values by extrapolating data and other data augmentation techniques. After the data is audited and cleansed, it will be more accurate, valid, complete, and consistent. High-quality data is a critical factor in enabling Information as a Service, as well as a key differentiator that increases efficiency, enhances customer service, and drives profitability.
   b. **Data enrichment** (see "Transforming Data into Information and Deriving Business Insights" earlier) is a value-adding process that supplements, refines, and improves the existing data by adding calculated data elements and external data from multiple sources to an existing data set to enhance the richness and completeness of that data. Data enrichment is often applied to customer records to facilitate demographic and psychographic marketing analysis, and to increase the usability of unstructured data when context or meaning is missing or unclear (e.g., tagging and enriching to address ambiguities and facilitate search and discovery). Data enrichment enables the transformation from data into information and maximizes the potential business value and data re-usability.

3. **Data lifecycle management.** This stage is actually a series of data management activities that form a comprehensive approach to managing an organization's data, involving procedures and practices as well as applications. The effective management of corporate data has grown in importance as businesses are subject to an increasing number of compliance regulations. Therefore, organizations require the development and execution of architectures, policies, practices, and procedures to manage the information lifecycle in an effective manner. Companies that borrow orchestration patterns from cloud computing and apply them to data lifecycle management, will create a more nimble environment able to absorb and flex with the ever-increasing rate of change.

   The enterprise data model described earlier, combined with data modeling techniques and methodologies, helps store data in a standard, consistent, and predictable manner—enabling organizations to manage data as an asset. The conceptual model is translated into a logical data model which is then transformed to a physical data model. The data model is also further extended to be used in real-time messaging services. Forrester has identified the practice of canonical modeling as a best practice for Information as a Service and SOA.[7]

   Greater availability of data increases the risk of data loss and misuse of data, which must be addressed for Information as a Service to be a viable option for large enterprises. Data loss protection procedures, redundancy and referential integrity management, data privacy and confidentiality management, and corporate policies can help ensure appropriate data preservation and access over prolonged periods. Approaches include dedicated backup processes, mirroring, availability zones, replications, and disaster recovery plans. Data encryption and compression, both in storage and in transit, are prevalent techniques to gain enhanced data protection and efficiency.

4. **Information and insight delivery.** This step covers how data is accessed, explored, and visualized. Information as a Service should be able to deliver information in a form that is convenient to the consumer. No single approach suits all environments, but the delivery strategy must be consistent. Information as a Service is often delivered through a SOA infrastructure. Delivery forms can vary and include XML, JavaScript object notation (JSON), SQL, graphics, interactive analytics cubes, formatted text files, RESTful (Representational State Transfer) service or other proprietary technology.

Tools and Technology

No single tool or vendor solution available today addresses all the layers, components, and capabilities of an Information as a Service ecosystem. A particular organization's philosophy regarding technology choice will determine the types and number of tools required to implement an Information as a Service solution. For example, if an organization prefers large vendor-packaged solutions, fewer separate tools and technologies will be required. Conversely, organizations that prefer to avoid vendor lock-in, often choosing open source solutions over vendor-packaged solutions, will need to implement and integrate a greater number of tools and technologies.

Rather than detailing one of these approaches or the other, Table 2 provides a sample of technologies for each of the components listed in Tables A1 – A3 in Appendix A: Information Architecture Framework Component Details. This list is not meant to be a comprehensive pick-list for assembling an Information as a Service ecosystem. Rather, it is included to demonstrate the scope and complexity of an Information as a Service ecosystem, and to provide a starting point for exploration of the types of technologies that an organization must either master, or identify service providers to address for them.

[7] Forrester defines a canonical model as "a model of the semantics and structure of information that adheres to a set of rules agreed upon within a defined context for communicating among a set of applications or parties." The goal of the canonical model is to provide a dictionary of reusable common objects and definitions at an enterprise or business-domain level to enhance system interoperability. For further discussion, see http://blogs.forrester.com/mike_gilpin/11-05-27-canonical_information_modeling_a_best_practice_for_soa.

Table 2. Sample tools and technologies that can help implement an Information as a Service architecture framework

Core Data and Information Components		
Component	**Purpose**	**Technology Discussion**
Data Sources and Operations	Generates and processes data for operational and core front-end functions.	This component may include traditional relational database management systems (RDBMS) such as Microsoft SQL Server*, Oracle*, Netezza*, DB2*, and MySQL*. It may also include NoSQL databases such as HBase*, Mongo*, Cassandra*, Riak*, or CouchDB*. If unstructured data is a source, then technologies such as Apache Hadoop*, BigTable*, or one of the hybrid databases such as Greenplum* or AsterData* may be useful. If an organization leverages cloud services, data may be sourced directly from cloud storage or through a cloud service, such as Software as a Service (SaaS).
Integrated Information Platforms	Integrates, stores, and enriches data from multiple sources and distributes data for a variety of front-end business usages.	This component may include traditional databases providing operational data store (ODS) functionality or acting as a data warehouse (DW). These may include Microsoft SQL Server, Oracle, Netezza, DB2, and MySQL. If high performance is desired, organizations can use one of the in-memory database solutions such as Hana* or Oracle X10*. Other applicable technologies include the following: • Hybrid databases such as Greenplum or AsterData • NoSQL databases such as HBase, Mongo, Cassandra, Riak, or CouchDB • NewSQL and Big Data technologies such as Hadoop and BigTable • Cloud storage from cloud providers such as Google, Amazon, and Rackspace
Master Data Management	Holds key dimensional data.	Many of the RDBMS vendors provide master data management solutions. These include companies such as Oracle, IBM, SAS, Informatica, TIBCO, and Information Builders. Similar to other components of the Information as a Service ecosystem, Master Data Management is being provided by more companies through cloud models (for example, Informatica Cloud*, SnapLogic*, Dell Boomi, and Cognizant*).
Data Integration Components		
Component	**Purpose**	**Technology Discussion**
Data Movement and Integration	Represents tools, techniques, capabilities, and services that move and integrate data between the subsystems in the Core Data and Information components and connects them to the Presentation and Information Delivery Components for distribution.	This component will include either physical or logical integration technologies. Physical integration is accomplished through the use of extract, transform, and load (ETL) tools such as those provided by Talend and Informatica, as well as IBM Infosphere DataStage*; or through cloud-based ETL companies such as SnapLogic. Many of the traditional ETL vendors, such as Talend and Informatica, are now providing cloud-based versions of their products. These tools address batch data integration, managed file transfer and change data capture (CDC). Most also include service interfaces for accessing data integration using simple object access protocol (SOAP) or RESTful (Representational State Transfer) APIs. Logical integration is typically done through one of the data federation technologies such as Composite*, Denodo*, or Informatica.
Data Acquisition and Distribution	Acts as an intermediary layer that extracts and replicates raw data from sources, then standardizes, consolidates, and distributes the data to the Integrated Information Platforms component.	Two key technical capabilities are involved in this component. • **Intermediary storage.** Intermediary storage is often achieved by using a managed file system (file system with file management software), RDBMS technologies. Alternatively, it can be achieved through Big Data platforms such as Hadoop or BigTable. • **Distribution or movement of data.** Minimizing the number of physical technologies can be accomplished by leveraging the ETL tools from the Data Movement and Integration component—tools such as Talend, Infosphere DataStage, or Informatica. Many of the ETL platforms also provide data profiling and data cleansing add-ins.

Presentation and Information Delivery Components		
Component	**Purpose**	**Technology Discussion**
Information Services	Extracts integrated data from the Integrated Information Platforms and Master Data Management components and develops reports, analytic cubes, data marts, and business intelligence (BI) applications.	This component includes many of the traditional BI and reporting technologies. These technologies include both open-source platforms such as Jaspersoft* and Pentaho* as well as vendor-provided solutions such as Microstrategy*, Cognos*, Business Objects*, Roambi*, and Tableau*. These applications may be deployed within the data center or accessed through cloud-based models. Other vendors, such as Apptio, provide domain-specific SaaS information services. Many of these technologies bundle data access with capabilities for a common semantic view of the data. Some Information as a Service implementations will present a common semantic view of the data through relational database views or through applications service interfaces such as SOAP, or RESTful, or JavaScript object notation (JSON) APIs.
Data and Information Access	Represents the digital interaction layer that includes portals, websites, and digital/mobile applications.	Multiple application groups comprise this component. Portals can be provided through technologies such as Microsoft SharePoint*, one of the Oracle Web portals, Hippo*, eXo Platform*, and many others. Some companies will use web content publishing or website generation technologies as well as a significant number of mobile applications that have entered the scene in recent years. Other companies will chose to move some or all of these components to the cloud.

In addition to the technologies discussed in Table 2, an Information as a Service ecosystem will require additional technologies:

- **Data security.** Many of the RDBMS, extract, transform, and load (ETL) tools, along with BI platforms, and other technologies previously listed include security capabilities. When choosing tools or technologies, it is critical that organizations also evaluate the tool's security paradigms to determine if they are compatible with the organization's security requirements. Security must span the entire Information as a Service ecosystem; otherwise, an organization risks exposure of one of its most critical assets, data.
- **Data catalog.** Another component critical to the Information as a Service ecosystem is the data catalog function. Traditionally, cataloging of data has been done through a data modeling tool that produces a published data dictionary. This is supported at runtime by the system tables and data catalog within relational database technologies. However, with the popularity of unstructured data and schema-less databases of recent years, a different implementation of data catalog is developing, such as Apache HCatalog*. The data catalog expands on the data model to provide an interactive catalog of elements that can be interrogated at runtime. Typically, utilization of a tool such as HCatalog or a custom implementation of the catalog using a relational database, and NoSQL technology or a file store is required to achieve the full potential within an Information as a Service ecosystem. However, by whatever means the catalog is implemented, it is critical that it provides the ability for end users or service developers to seek and identify available data, understand that data, and act upon it.

The breadth and complexity of an Information as a Service ecosystem can be daunting. Considering the preceding list of tools and technologies, an organization must have a broad range of technical skills or sufficient relationships with solution providers to support a full Information as a Service implementation. A variant of this approach is to take advantage of services offered in the cloud-based model. In either case, the ODCA recommends that organizations start with a carefully planned information architecture framework, refer to ODCA Usage Models and white papers during implementation, and use ODCA RFP requirements when selecting vendors.

GOVERNANCE

Information as a Service can be thought of as an information broker – it implies a two-way relationship between data sources and information consumers. Therefore, governance for Information as a Service requires a proactive focus on both ends of this relationship. General Information as a Service governance capabilities are as follows:

- **Data ownership.** Because Information as a Service may have various data sources and information consumers, it must maintain data ownership to help to define who owns the intellectual property of the applicable data and information, and who is responsible for the applicable data and information. In other words, it is essential for Information as a Service to control who has the ability to access, create, modify, package, or remove data, as well as who can assign these access privileges to others.
- **Data change management.** To avoid errors and inconsistency in data, as well as the disappearance of data, it is necessary to prevent data from being changed improperly. It is critical to establish appropriate Data Stewardship and solution design review processes in the System Development Lifecycle to ensure new solutions don't adversely affect the quality or flow of the data that could impact downstream and upstream business processes consuming it, breach information policies, and regulatory compliance, etc. Also, it is important to be able to trace data changes for audit purposes or fault recovery. Information as a Service must have the following abilities:
  - Control of data change privileges (this includes authentication and authorization)
  - Control of data revision and a log of revision history
  - Tracking of data status information (this includes Deletion Status: Active/Deleted and Approval Status: Awaiting Approval/Approved/Rejected)
  - Enforcement of data destruction policy and procedure
- **Data protection.** Information as a Service should include the following data protection parameters:
  - Prevent improper changes to data
  - Prevent data from being retrieved by consumers without permission
  - Ability to restore data when disaster occurs

  Besides the data change management considerations discussed previously, Information as a Service must also provide mechanisms for data access permission and restriction, data access logging, data encryption and key management, data backup and recovery.
- **Legal Compliance.** One of the most important governance capabilities Information as a Service must provide is legal compliance. As mentioned previously, Information as a Service has to maintain two-way relationships between data sources and information consumers; if data sources or information consumers are from different countries or geographic regions, Information as a Service should have a mechanism to meet legal compliance requirements for each relevant country or region. In addition, if information is used across industries, Information as a Service must include the ability to meet all relevant industry compliance requirements.

SERVICE DELIVERY ARCHITECTURE

While it is possible to implement Information as a Service without taking advantage of cloud computing, the cloud is a natural fit for delivering such a service. Therefore, this section discusses service delivery architecture in the context of the cloud.

Within a cloud environment, everything—data, information, platform, infrastructure, and software, just to name a few—is described as being delivered "as a Service"; this section describes what that means. However, this document does not dictate what can be delivered as a service and how; rather, it is up to the organization or cloud service provider to determine how to fulfil the requirements discussed in this section.

The description of the service delivery architecture must cover two basic aspects—the service structure, which defines the contents of the service, and the organization and process structure, which defines how the demand/supply model will be constructed. Table 3 provides an overview of each of these aspects of the service delivery architecture; they are discussed more completely in the following two subsections.

Table 3. Service Delivery Architecture Overview

Service Structure	Organization and Process Structure
How the services will be defined and what service-level elements will be delivered. This aspect should include the following items: • The framework for describing the catalog of services in terms of verifiable service attributes[8] • Definitions and implementation for key concepts such as service catalogs, classes, attributes, and resource groups, as well as commitments, reports, and triggers • A means by which the service attributes can be described in a standard way • Identification of potential standard service levels	How the demand/supply model will be constructed. This aspect should include the following items: • Identification of the actors involved, and their objectives and requirements. Types of actor include Demand, Supply, and Broker. • Processes required to establish a service for any customer and to manage its delivery • Organization structure, including the demand/supply model • Tool requirements for supporting the organization and process structure

Service Structure

Depending on an organization's choice of cloud deployment models, the organization or a third party may be responsible for certain components of the Information as a Service ecosystem. In a hybrid situation, there exists a mixture of internal, external, and various forms of DaaS and SaaS from other organizations.

The service contents include the following facilities:

- The ability to capture and store large volumes of raw data, possibly on a structured federated basis, with its associated metadata
- The analysis and extraction of this data from those sources, to produce meaningful information
- Use of the reporting and analytics layer to deliver information in a format useful to the end user

These facilities should be delivered in a cohesive end-to-end manner, rather than as disjointed components.

Also, the cloud service provider must commit to agreed-upon quality levels for each of these facilities and the services that enable them. These levels can be described using the following terminology:

- **Inputs.** At the lowest service level, the service is simply to "do something." For example, the service will react when a particular event or series of events happen.
- **Outputs.** At a slightly higher quality level, the service goes beyond merely "doing something" to "delivering something." For example, the service will deliver data at a specified availability level.
- **Outcomes.** At the highest quality level, the service is to "achieve something." For example, the service will provide a certain level of user satisfaction.

These quality levels are cumulative; that is, a cloud service provider needs to deliver inputs in order to be able to deliver outputs, and cannot deliver outcomes without both inputs and outputs.

The cloud service provider(s) should use an SLA to commit to the levels at which services will be provided. For instance, the SLA would state the cloud service provider's committed levels of availability and turn-around time.

[8] Refer to the ODCA Service Catalog Usage Model for more information.

Organization and Process Structure

A unified client interface should be implemented from the following two perspectives:

- Service administrator, who can commission new services, monitor running services, or report on previous use of services
- End users, who access the services

The processes involved in delivery of any cloud service are illustrated in Figure 6.

The processes shown in Figure 6 form the foundation for any Information as a Service architecture, which is layered on top of them to deliver an added-value service.

Staffing Considerations

Capable staff are expected to be available to fulfil a number of aspects regarding these processes:

- **Operation.** Staff must monitor and manage the entire environment, so that it continuously delivers the necessary service levels.
- **Data curation.** Staff must know the following: what data is available; where it is located; its origin, reliability, and currency; who owns any intellectual property associated with the data; the security of the data; and how the data relates to other available data.
- **Analysis.** Staff must provide advice and guidance as to how the relevant information can be obtained from the data.

Requirements when Multiple Service Providers Are Used

Fulfilment of the complete Information as a Service environment will typically involve the collation and aggregation of services from multiple service providers. Therefore, it is required that single points of contact can be provided to customers, to assist in incident management and problem-solving. This prevents end users from being subjected to "ticket bouncing" between several suppliers. Also, transparent charging and billing should be in place, so that customers can receive a single invoice, making it clear which charge was incurred where, with a simple means of payment.

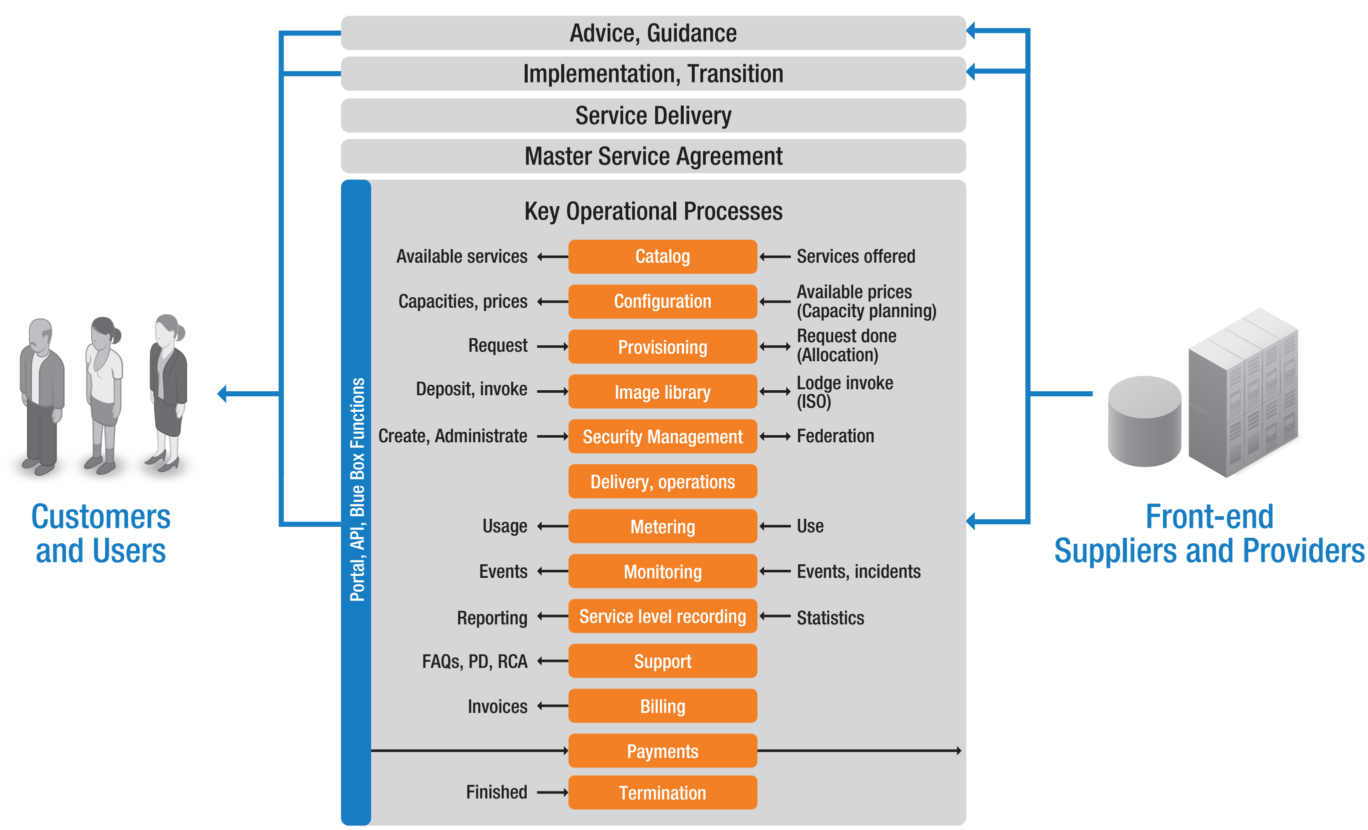


Figure 6. Cloud Service Delivery Processes

PROGRESSION TO INFORMATION AS A SERVICE

Because the majority of companies already have existing information and storage platforms, they will naturally use these existing resources as they begin their implementation of an Information as a Service ecosystem. Therefore, it is helpful to have a tool to more closely examine how to move from a legacy environment to the ideal target state. The purpose of this Master Usage Model is not to define in detail what organizations or service providers deliver in terms of precise technical features. However, implementation-level details are often required in order to understand the nuances of new technology models – particularly something as expansive and multi-layered as an Information as a Service ecosystem.

Table 4 provides a list of characteristics critical to the concept of Information as a Service and the progression that solution providers can take to move their products and services to this model. The characteristics are derived from the sample Information as a Service scenarios in Appendix B: Sample Information as a Service Scenarios of this document and are aligned with the content of other ODCA Usage Models[9].

For this Master Usage Model, it is not intended that all of the characteristics in a given column of Table 4 must be supported as a group. In practice, a given organization's or information service provider's solution will combine different service levels for different elements. Also, not all characteristics need to progress at the same pace; organizations can examine their specific needs and formulate decisions about which characteristics are most critical and are therefore worth significant investment.

[9] www.opendatacenteralliance.org/ourwork/usagemodels

Table 4. Progression of Characteristics for Information as a Service

	Legacy Environment – Not Information as a Service	Approaching Information as a Service	Typical Information as a Service	Market-Leading Information as a Service
General	Solution is open and standards-based	Solution is open and standards-based	Solution is open and standards-based	Solution is open and standards-based
Data Sources and Operations	Simple manual user interface, self-service	Some real-time interfaces to data acquisition services	Real-time interface to a full range of data acquisition services, cross-provider interface support	Managed provision, including maintenance of alternative sources
	Capabilities and features support a minimum of ITIL processes	Capabilities and features that support one or more ITIL processes	Capabilities and features that support ITIL change, incident, and configuration management	Capabilities and features that dynamically support a wide array of ITIL processes
	Solution is comprised of multiple services, each with its own user interface	Multiple service user interfaces are fronted by a single portal	All user interaction with services is done through a single user interface (single pane of glass) – management console may be different	All user interaction and all management of sub-services are integrated into a single user interface
	All management functionality exposed through RESTful API		All management and all user functionality exposed through RESTful APIs	Native capabilities for extending management and user functionality RESTful APIs
	Operational metrics exposed through user interface	Base capabilities to define SLOs and SLAs, document metrics, track metrics, and generate SLO and SLA variance report[10]	Advanced capabilities to define SLOs and SLAs, document metrics, track metrics, and generate SLO and SLA variance report	Includes prebuilt and user selectable SLA models and capabilities that provide dynamic capture of metrics, ability to extend existing and define new metrics, and SLA variance reporting
Integrated Information Platforms	Simple manual user interface, self-service			Fully-managed provision, including from multiple independent sources
	Short-term data storage (Hub/staging)	Features to support dynamic data activities, such as data discovery, innovations and experiments, and model testing	Broader features to support dynamic data activities, such as data discovery, innovations and experiments, and model testing	Dynamic management of storage to support preceding features
	Support for integration with the top-tier social media platforms		Capabilities to integrate with products and features of social media platforms	
Master Data Management	Simple manual user interface for top two or three master data management subject areas[11]	Expanded master data sets, increased integration of master data into data, and information access services	Advanced metadata manipulation capabilities	Full catalog of catalogs; capabilities for real-time match/merge
Data Movement and Integration	No built-in data movement or integration, ability to integrate off-the-shelf ETL/ELT tools without significant integration effort	Built-in support for data movement and integration	Support for real-time messaging between components, dynamic orchestration	Same as for Typical Information as a Service
	Batch and real-time integration			
	RDBMS, NoSQL, and Apache Hadoop* as sources and targets			
Data Acquisition and Distribution	Simple manual user interface, self-service	Micro-batch capabilities, introduction of simple acquisition services	Real-time interface to a full range of data acquisition services, cross-provider interface support	Managed acquisition and distribution services, with ILM
	Source from file, directory, URL, or other network addressable storage location	Same as for Not Information as a Service		
	Data hub capability	Additional data management controls within the data hub		

	Legacy Environment – Not Information as a Service	Approaching Information as a Service	Typical Information as a Service	Market-Leading Information as a Service
Information Services	Simple manual user interface, self-service		Real-time interface to a full range of information services Cross-provider interface support	End-user experience monitored, significantly increased service penalties for failure or degradation
	Capabilities to provide a common semantic view of the data	Broader semantic capabilities	Integration with third-party semantic tools	
Data and Information Access	Database and API access to all data within the ecosystem		Logical or physical federated query capabilities across data sources	
	Mobile access to information reporting and presentation	Mobile content publishing	Management capabilities exposed through mobile applications	APIs and SDK to support mobile application development using Information as a Service ecosystem resources
	Support for search, reporting, business intelligence, and analytics	Visualization capabilities	Natural language processing capabilities	
	Ability to publish data to web-based portal	Interactive portal capabilities, such as comments, blogs, and data mark-up		
Data Quality Management	Baseline data quality capabilities	Addition of data correction and more advanced data quality capabilities	Full-featured data quality capabilities including profiling and data augmentation	Advanced data quality features including options for market/ domain-specific data augmentation
Data Security Management	As per ODCA Provider Assurance Usage Model[12]			
Metadata Management	Simple manual user interface, self-service metadata management	Addition of metadata services, operating independently of the metadata itself	Integrated use of metadata to drive orchestration decisions	A rich repertoire of metadata supported for use to drive orchestration, distribution, and analytics
Data Storage and Support Services	Simple manual user interface, self-service capabilities for resource consumption	Limited rules-driven resource consumption capabilities		Dynamic resource allocation based on resource, cost, and performance optimization models
	Integrated data loss and business continuity capabilities		Geo-distribution and dynamic load balancing capabilities	Self-healing capabilities
	Implementation of a basic catalog capability	Catalog integration with master data management capabilities		Dynamic workload based on catalog values
Architecture Governance	Information output to support the information governance process	Capabilities to record and track data elements that require governance	Governance defined through meta models	Full cooperative information lifecycle management
				Capabilities to affect data integration and orchestration based on governance meta models
Security	As per ODCA Provider Assurance Usage Model[12]			

ELT = extract, load, and transform; **ITIL** - Information Technology Infrastructure Library; **SLA** - service-level agreement; **SLO** - service-level objective; **ILM** - information lifecycle management; **REST** - Representational State Transfer

[10] A service-level objective (SLO) is used to define the availability and performance goals for an application.

[11] Master data includes the core entities (also called domains) that describe the ongoing activities in an organization. Examples include Parties (business partners) in the roles as contacts, prospects, customers, suppliers, members, citizens, and any other parties with whom the organization interacts; Products and Assets that the organization buys, produces, has, and sells; Places (locations) such as production plants, stores and warehouses, delivery points, and visit destinations; and Periods (calendar) for passed, current, and future activities.

[12] See "Open Data Center Alliance[SM] Usage: Provider Security Assurance" at www.opendatacenteralliance.org/library.

RFP REQUIREMENTS

The ODCA brings consumers and providers of data center technologies together to advance capabilities in the market. These groups work together to address challenges in today's solutions, and outline a future state for the industry to work toward. One technique utilized in defining this future state is the compilation of request for proposal (RFP) and request for information (RFI) requirements. Solution providers can use these requirements as input into their product roadmaps. For companies wishing to acquire some or all components of their Information as a Service ecosystem, these requirements can be used as-is, reducing the time required in the procurement stage.

The Information as a Service ecosystem is complex and multifaceted. Each tier of the architecture contains multiple RFP requirements. Rather than detailing all of these requirements in this document, the ODCA Data Services Working Group has outlined key requirements through a general description. Specific, individual RFP statements can be found through the ODCA Proposal Engine Assistant Tool (PEAT)[13].

First and foremost, it is critical that all solutions relating to Information as a Service include the ODCA principle requirement – that the service is open and standards-based.

Other general requirements include the ability for solutions to receive, transmit, and persist structured and unstructured data. These capabilities must be implemented in a manner that provides logical separation between consumers and providers of the data. Solutions participating in the Information as a Service ecosystem must support a wide range of operating systems, must support management compliant to the Information Technology Infrastructure Library (ITIL) operations framework, and should provide a single management interface for interaction with all components within the ecosystem.

Components of the Information as a Service ecosystem must provide short- and long-term data storage and the ability to move data between persistence stores, BI applications, analytical tools, and other consumers of information. Movement may be done with a traditional ETL tool, message bus, or through application services – drawing on cloud orchestration patterns. All data and functionality should be exposed through RESTful APIs, as well as through management, social, BI, and reporting user interfaces. Capabilities for standardizing, cleansing, and augmenting data must exist, as well as metadata and master data management functionality. The ability to contextualize information by utilizing a common semantic view of data is critical. Tagging and other attribution of data and information elements is also required.

Access to information must be provided through database query and APIs, as well as through search, reporting, BI, and analytic interfaces. Information should be accessible from within the Information as a Service ecosystem, from across other systems within the company's data center, and from external clouds or across the public Internet. These accessibility requirements necessitate a comprehensive and strict security model, with particular attention to the regulatory and compliance requirements of the geography from which the data is originated, stored, and consumed.

Most foundational to the requirements for Information as a Service ecosystem is a strong architectural base from which components can be applied in a consistent and coherent manner, with a governance model for ongoing change management and technology refresh.

SUMMARY OF INDUSTRY ACTIONS REQUIRED

This document describes an Information as a Service ecosystem and provides a summary of requirements for the solution provider community to consider when delivering entire Information as a Service solutions or services that function within the ecosystem. The ODCA Data Services Working Group provides this document as a means of moving toward the goal of fostering collaboration between Information as a Service solution providers and large enterprise consumers of information services. In the interest of accomplishing the mission of the ODCA, we hope to motivate these two groups to work together to define the open specifications, formal or de facto standards, and common intellectual-property-free solution designs—all of which are required if we are to realize an open, interoperable, and thriving market of information-driven businesses and information-driven services.

The following actions are required by the combined solution provider/consumer community:

- Broad adoption of a common and consistently used reference architecture for Information as a Service.
- Investment by solution providers in delivering products that meet the requirements of Information as a Service as outlined by the large enterprise consumers participating in the Open Data Center Alliance.
- Further maturation of Big Data, NoSQL, and NewSQL technologies to include open and interoperable interfaces as well as support for metadata, master data management, data quality, and other standard data management features that the BI and data warehousing industries have long supported.
- Adoption of the data supply chain metaphor and implementation of the components and services required for enterprises to take advantage of it.
- Continued focus on data security, cloud security, and mobile security.

[13] www.opendatacenteralliance.org/ourwork/proposalengineassistant

APPENDIX A: INFORMATION ARCHITECTURE FRAMEWORK COMPONENT DETAILS

Table A1. Core Data and Information Components

Component	Description	Data Type and Hosting Requirements
Data Sources and Operations	This component represents the operational transaction/data sourcing layer which generates and processes data for operational and core front-end functions. This component includes transaction systems, product/account/services origination systems, and customer origination systems. It does not integrate data outside its operational processing needs. The data history stored in this component is minimal or sufficient to run the operational function—this is to maintain high performance. Reporting is limited to out-of-the-box application-level reporting. This component supplies source data to associated components including the Information Access component.	This component can contain structured or unstructured data and can be hosted internally or in the cloud.
Integrated Information Platforms	This component integrates, stores, and enriches data from multiple sources and distributes data for a variety of front-end business usages, such as data warehouses (DWs), operational data stores (ODSs), and rules engines. This component can contain short- or long-term data history for trend analysis. The subsystems of this component are considered enterprise assets, and can use near real-time or batch-oriented processing. This component can serve as a short-term data store for dynamic data activities, such as data discovery, innovations and experiments, and model testing. This component can integrate data from multiple sources.	This component can contain structured or unstructured data and can be hosted internally or in the cloud.
Master Data Management	This component holds key dimensional data (low volatility and non-transactional data, such as customer and product data) that is of critical enterprise value. It acts as the centrally managed and synchronized authoritative source in situations where the same data is generated by multiple applications. This single view of data is enabled using data match, data merge, and rules of survivorship. Also, this component can contain multiple hierarchical views of commonly used data for centralized distribution. This component publishes data to all other Core Data and Information components for processing, integration, and presentation. This helps enforce data consistency and data architecture simplicity.	This component applies mostly for structured data and ideally should be managed internally within the organization.

Table A2. Data Integration Components

Component	Description	Data Type and Hosting Requirements
Data Movement and Integration	This component represents tools, techniques, capabilities, and services that move and integrate data between the subsystems in the Core Data and Information components and connects them to the Presentation and Information Delivery Components for distribution. This component includes a real-time messaging and application integration service bus, batch data integration, managed file transfer, and change data capture (CDC). Batch data integration can be accomplished using extract, transform, and load (ETL) and/or extract, load, and transform (ELT).	This component can apply to structured and unstructured data and the service can be provided by internal staff or third-party service providers.
Data Acquisition and Distribution	The data in this component is not integrated, transformed, or enriched. This component acts as an intermediary layer that extracts and replicates raw data from sources, then standardizes, consolidates, and distributes the data to the Integrated Information Platforms component in a consistent manner – functioning as a data staging and distribution hub. When properly implemented, this component enables consistency in data integration. In addition, this component helps avoid multiple point-to-point integrations that contribute to a complex data architecture. This component can also be used for data profiling and cleansing activities to support the Data Quality Management component. This component simply distributes the data to target systems as per the source (cleansing where appropriate). This differentiates the subsystems in this component from the Integrated Information Platforms component's subsystems.	This component can apply to structured and unstructured data and the service can be hosted and supported internally or externally.

Table A3. Presentation and Information Delivery Components

Component	Description	Relevant Open Data Center Alliance℠ (ODCA℠) Usage Models
Information Services	This component extracts integrated data from the Integrated Information Platforms and Master Data Management components and develops reports, analytic cubes, data marts, and business intelligence (BI) applications — thereby delivering information with value and insights to the business and to customers. This component is system-agnostic, with a common semantic view of data. This approach enables consistent, reusable, and sharable information delivery services such as analytics and reporting. This component includes modeling and discovery services that utilize the innovation/exploration environment of the Integrated Information Platforms component.	
Data and Information Access	This component represents the digital interaction layer that includes portals, websites, and digital/mobile applications. This layer publishes the information from all the data/info layers using capabilities from the Data Movement and Integration component, such as APIs, real-time/application integration, batch/data integration, and data cache. The subsystems in this tier can also act as the entry point where the business or customers enter data using portals and websites. The data in this component is then integrated with the other components. In such scenarios the Data and Information Access subsystems can be part of the Data Sources and Operations component, as well.	Particular attention should be paid to security. Several of the Open Data Center Alliance Secure Federation Usage Models can help in this area.

Table A4. Control and Support Components

Component	Description	Relevant Open Data Center Alliance[SM] (ODCA[SM]) Usage Models
Data Quality Management	The subsystems in this component deliver the data controls that are required in data stores and flows across the Core Data and Information components to help ensure accuracy, consistency, and completeness of data. This can help meet regulatory and business expectations concerning data quality, and can reduce operational risks relating to data quality issues. Subsystems include profiling tools, financial reconciliations, error controls, and file controls.	Data controls are detailed in the ODCA Data Security Framework. Access control, data encryption, security information, event management, and backup, archiving, and deletion are addressed. The ODCA Regulatory Framework may also the helpful.
Data Security Management	The controls in this component protect data from privacy and confidentiality issues, and help prevent fraud, loss, and corruption. Subsystems include data encryption, tokenization, and identity management.	Similar to the Data Quality Management component, the ODCA Data Security Framework and Regulatory Framework will both be helpful when implementing this component.
Metadata Management	This component manages information about the data and is critical for sustainable data management. It includes the following: • Business metadata such as definitions, a business glossary, descriptions, derivation rules, and conceptual or logical data models • Technical metadata such as physical data models, data structures, and data lineage • Operational metadata such as delivery time, exception and error handling rules, and thresholds	
Data Storage and Support Services	This component manages the disaster recovery of the data, capacity management, data retention (as per policy), and operational service-level agreements (SLAs). This component interacts with all other components of the Information as a Service architecture. Subsystems of this component include Infrastructure as a Services (IaaS), Platform as a Service (PaaS), and production support services.	IaaS has been explored in multiple ODCA Usage Models. The Compute Infrastructure as a Service Usage Model covers this area most thoroughly. Several other ODCA Usage Models, including some of the Secure Federation and Interoperability Usage Models may also be helpful.
Architecture Governance	This component provides overall governance of principles and standards across all other Information as a Service architecture components, to help ensure sustainable and scalable capabilities in each component. It also helps to improve the cohesiveness of the entire Information as a Service ecosystem.	The ODCA recommends that individuals responsible for this area of the overall Information as a Service ecosystem be well-versed in the principles and core concepts of the ODCA Usage Models.

APPENDIX B: SAMPLE INFORMATION AS A SERVICE SCENARIOS

The following two sample scenarios illustrate how Information as a Service can be implemented.

Sample Scenario 1 – Private Facility

Figure B1 illustrates an Information as a Service scenario where data comes from internal and external sources, including social media. It is distilled into information for internal and/or external customers. The key components of this scenario are as follows:

- **Data hub.** Acts as a data clearing house that provides a Data as a Service (DaaS) function – this hub acquires data from multiple internal and external sources and distributes it to target destinations.
- **Information systems.** Includes DWs and ODSs. Performs the core information processing function.
- **Information publishing.** Information is published from the information systems and the data hub for internal and external customer consumption using common semantic views and platform.

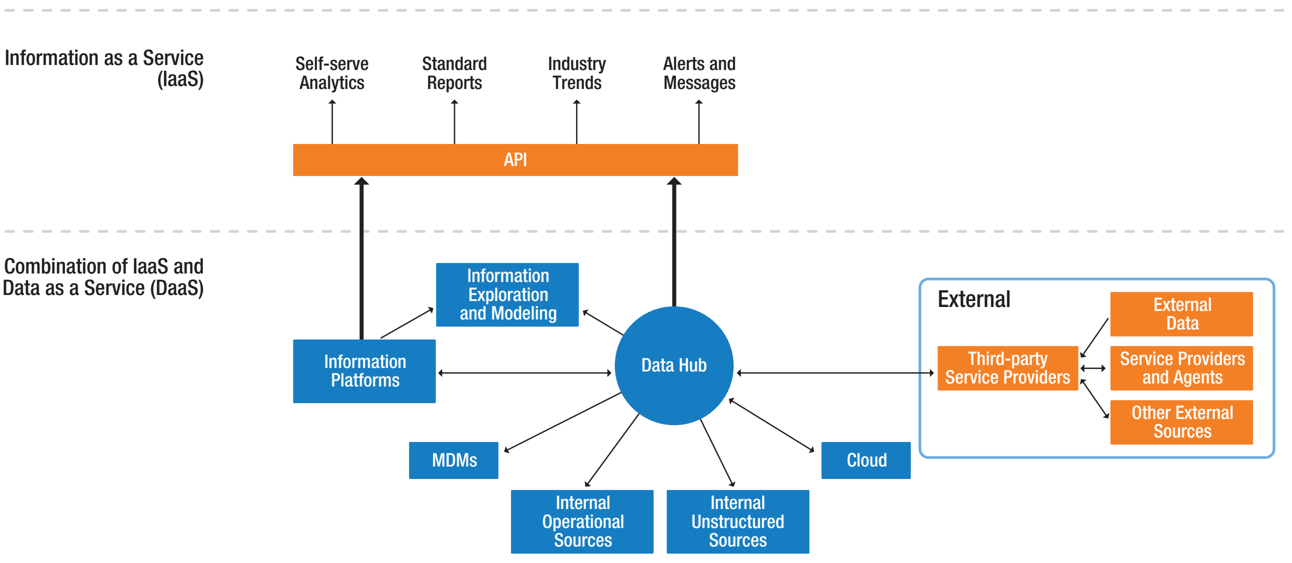
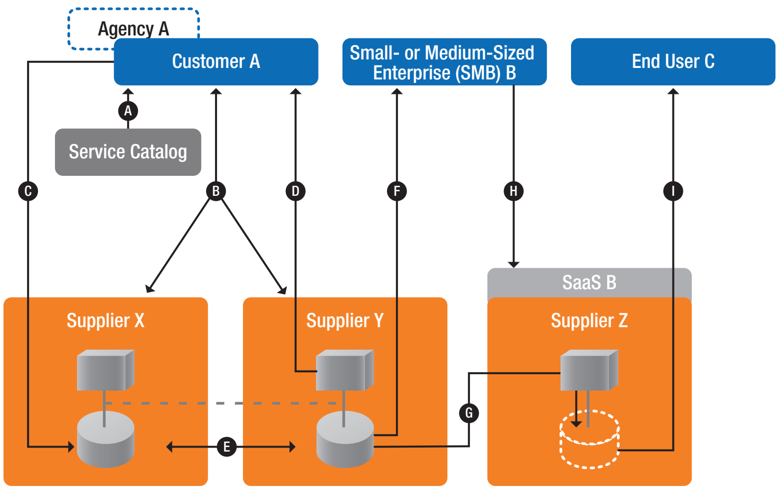


Figure B1. Sample Private Facility Information as a Service Scenario

Sample Scenario 2 – Helix Nebula

Figure B2 illustrates how Helix Nebula uses Information as an Architecture.[14] This infrastructure is designed to ultimately provide physical and organizational structures and assets needed for the IT-related operation of research institutions, enterprises, governments, and society. This pan-European partnership across academia and industry is working to establish a sustainable European cloud computing infrastructure, supported by industrial partners, which seeks to provide stable computing capacities and services that elastically meet demand.

In particular, as shown in Figure B2, Helix Nebula seeks to provide small- and medium-sized enterprises with access to large-scale data, such as from the European Space Agency—although the application of this Information as a Service scenario could be applicable to providing large enterprises with access to data as well.



- **A** Customer A determines what services are required, and available from the catalog.
- **B** Customer A then invokes supply arrangements with Suppliers X and Y.
- **C** Customer A stores data with Supplier X.
- **D** Customer A invokes processing capacity from Supplier Y.
- **E** Suppliers X and Y manage data availability and access.
- **F** Small- or medium-sized enterprise (SME) B determines the location of and accesses the required data, which maybe open data.
- **G** SME B processes the data, with Supplier Z, to produce useful information.
- **H** SME B establishes an SaaS-based information service with Supplier Z.
- **I** End user C invokes the information service.

Figure B2. Helix Nebula Information as a Service Scenario

Within Helix Nebula, the agencies are research organizations (CERN with their Large Hadron Collider and the European Space Agency with Earth observation satellites) that accumulate large quantities of data. End users may be organizations such as insurance companies that need information about earthquake or flood risk.

Within the Information as a Service scenario used by Helix Nebula, there are coordinated points of support contact between Suppliers X, Y, and Z, to provide necessary support facilities to all users. Also, SLAs exist to define agreed-upon levels and formats of service levels and reporting. Billing and payments are also incorporated. Billing occurs for data storage, processing, and information; the funding agency and end users provide payment to the suppliers for storage and processing, to Customer A for access to the data, and/or to SME B for information. At every point in the data supply chain, payments are made by the consumer to the supplier of data or information, and each participant in the supply chain gets paid for their role as supplier, processor, or pass-through service.

[14] Helix Nebula - the Science Cloud is designed to support the massive IT requirements of European scientists. The project aims to pave the way for the development and exploitation of a Cloud Computing Infrastructure, initially based on the needs of European IT-intense scientific research organizations, while also allowing the inclusion of other stakeholders' needs (governments, businesses and citizens). (www.helix-nebula.eu)

FURTHER READING

This section provides references to relevant material, both that are created by the ODCA and by other authors and organizations.

Secure Federation Usage Models

The following Secure Federation Usage Models are available from the ODCA website[15]:

- Data Security Framework
- Security Provider Assurance
- Security Monitoring
- Identity Management Interoperability
- Cloud-based Identity Governance and Auditing
- IaaS Privileged User Access
- Cloud-based Identity Provisioning
- Single Sign-on Authentication

[15] www.opendatacenteralliance.org/library

