Model Driven Solutions

Where Business Meets Technology

A division of Data Access Technologies, Inc.
8605 Westwood Center Drive, Suite 505
Vienna, VA 22182

GSA ENTERPRISE ARCHITECTURE

BPA No. GST0006ACA2064-A

Task Order No. GST0007AC4144

GSA EA DATA ARCHITECTURE

WHITEPAPER AND INFORMATION MODEL

31 January 2008

Table of Contents

List of Tables

List of Figures

1. INTRODUCTION

1.1. Purpose

The purpose of this document is to provide a high-level description of GSA's target Data Architecture and Information Model to establish a common understanding of the agency's strategic direction towards information and knowledge management.

The GSA IT Strategic Plan FY 2007 - FY 2011 defines the strategic vision for using IT to meet GSA's business needs. A key to understanding GSA's business environment, processes and objectives is to understand mission critical information. Consequently, the higher quality the GSA information is, the better the view of GSA's mission and operational environment. Quality information also serves as an enabler for cooperation, interoperability and strategic decision making. Strategic Focus Area 1 of the GSA IT Strategic Plan is improving "timeliness and accuracy of data". In order to achieve this objective, GSA requires a well defined and disciplined processes for information capture, information stewardship and quality management.

As GSA seeks to enhance its services and understand the expectations of its customers, suppliers, and employees, GSA must be able to seamlessly integrate its strategy, people, processes, and technology internally, and with its external partners. The GSA Enterprise Architecture (EA) program has a key project, Integrated IT Portfolio Management (IPM), that joins and harmonizes internal policies with Capital Planning Investment and Control (CPIC), Systems Development Life Cycle (SDLC), and other processes and activities related to the inception, elaboration, construction, and management of IT initiatives across GSA. The One-GSA EA program includes information and data architectures as core deliverables, linking information with processes, services and data.

GSA Business Challenges

GSA is facing numerous challenges to its core businesses; especially its IT Services business. As a result, the agency has instituted a series of measures, including discretionary spending cuts, personnel cost reductions, and a hiring freeze to meet its legal obligation to 'break even'. With the continuing proliferation of Government Wide Acquisition Contracts (GWACs) across Government, GSA faces increasingly stiffer competition for its acquisition services from a growing Government competitor base.

The capability to supply acquisition service across the government – services based on a cost-effective, efficient and timely infrastructure – are the goals of OSERA.

In addition to the marketplace, GSA faces a number of internal challenges that have impact on this current task order:

- GSA continues to undergo a major transformation with the establishment of the Federal Acquisition Service (FAS). This transformation is a key component of the One GSA vision, but as with any major change, it adds to the uncertainty and discomfort within the organization.

- GSA has numerous processes related to its IT asset base (SDLC, PMP, EA, CPIC, etc.) that require constant management and update. Although related, these are not yet available to appropriate personnel in an easily accessible, configurable, and updateable environment.
- OMB and GAO are continually (and rightly) "raising the bar" in their assessment of Agencies Enterprise Architecture and overall IT maturity. GSA must keep pace with meeting the higher bar.
- The GSA OCIO has numerous stakeholders who want to be able to "view" the IT information asset base in ways appropriate for their business goals. Currently, each IT information asset has only a single "view".
- Much of GSA's data assets are currently siloed. GSA could achieve better use and management of these assets if they were appropriately available across the agency.
- Architectural information is frequently developed in forms appropriate to architects, but not business stakeholders. Information gets "trapped" into a single format that doesn't work for everyone.
- Currently the GSA EA "Bricks" are not linked directly to the OMB Federal Transition Framework nor the FEA Technical Reference Model
- While GSA has embraced the FEA, EA is not currently providing sufficient value to the enterprise.
- The GSA EA, itself, is not visible agency wide in a usable, accessible environment

To deal with these challenges, challenges that were clearly identified in the Integrated Portfolio Management project (IPM) is the need for better management of information in support of enterprise knowledge, decision making and execution.

The GSA Architectural Team has developed an approach with these challenges in mind.

GSA Knowledge Management Challenges

GSA is typical of most government organizations in that it has acquired and accumulated a great deal of information, architectures and plans that are not well organized, integrated or maintained. Much information, some of it costly, is lost, forgotten or simply not applied at the right time. Plans and architectures are done and redone due to changes in contractors, tools, methodologies, technologies and management.

The information that is retained is frequently not "linked" or consistent. The key drivers in an IT plan are not the same key drivers in a human capital plan, funding is not consistently applied to enterprise needs and architectures don't match reality. Decisions are made on inconclusive data after millions have been spent on analysis. <u>What information we have is not effectively transformed into knowledge.</u>

While the technologies and methods for managing operational data in DBMS systems are well established, the same level of maturity has not emerged for management and architectural information. This kind of information is less suited to the ridged structure of a DBMS, yet still needs some structure and management. It is this kind of information for which knowledge management is the right approach.

While the cost of this information loss is disturbing, even more troubling is that this is the kind of information that could help GSA achieve greater efficiency, transform to a more effective enterprise, improve its value to citizens and other agencies and achieve a more mature enterprise. OISP is the start of an initiative to address the core issues with knowledge management to help facilitate a more effective GSA. There is, of course, also process and culture changes required – some of which have been identified in the IPM project. The Internet based technologies of the OOSEA Portal can help facilitate these cultural and process changes with easy access to and management of knowledge.

Vision of ubiquitous knowledge

Consider that there will be a knowledge management platform available to everyone in GSA and, as appropriate, to external stakeholders and contractors. This platform has all the information that people need to do their job and plan for the future. This platform is not limited to a few structured databases, but allows virtually any kind of information to be used, analyzed and managed by anyone with the authority to do so. It is the easy way to find out what you need to know, what others have produced and the trusted place to store, manage and integrate the vast body of information people work with today. This platform leverages and is part of the internet.

This platform is also able to work with any kind of information in a variety of tools and formats. For example, the information in a spread sheet that is part of a human capital study can be used in a business case that is directly related to the systems architecture of an application being funded and then implemented. There is a vocabulary of common terms, accepted across GSA, that help more clearly define business goals and metrics – and these terms are used consistently across various information artifacts.

There is a direct link and continuity from business goals to enterprise architectures to systems architectures, acquisition and implementations. While different tools and diagrams are used for different stakeholders, the underlying information is consistent, traceable and "linked". Analysis can be done that crosses the "walls" between business and technical information.

As information and assets are developed within GSA and externally there is a culture of retention, extension and reuse. New information assets are stored in the repository and categorized according to their purpose, content and context. This allows new information to be found and reused easily, as well as managed reliably. New versions are saved and tracked so that the change over time can be better understood.

With this ability to manage, analyze and repurpose information GSA could be more effective at achieving its goals, of better serving its customers and of being a more mature enterprise. Without a handle on its knowledge resources it is hard to imagine much improvement at all.

This kind of environment is not science fiction, it is possible now, and OSERA is part of the initiative to provide it. OSERA envisions two classes of information and will have different capabilities for each. Virtually any information artifact that can be rendered as a "file" will be able to stored, managed, versioned and located. Each such asset will be categorized and contextualized such that it is easy to browse or query for the information you need. Once found it can deliver that information to any browser or desktop and keep track of any changes to it so that teams can work together effectively, even distributed teams.

A deeper level of integration and analysis can be provided for "structured" information, where the structure of that information has been modeled in and adapted to the knowledge base. This

integrated and structured information will initially include information assets such as business cases, PMP, information models, business processes and business collaboration models. This same information will be integrated with and directly support the Federal Enterprise Architecture (PRM, BRM, SRM, TRM, DRM) and the federal transition framework (FTF) – this makes the FEA an integral part of the GSA architectural framework, not an afterthought.

Significant architectural focus areas such as data, process and services modeling will be a key aspect of the initial "integrated" capabilities. Besides the basic capabilities there will be contextual information in the form of policies and white papers on how GSA does architecture.

Various design tools will be able to "check in" and "check out" information in the format that is appropriate for that tool. The information in the repository will be able to be restructured in the format of specific tools and standards so that knowledge can be shared between organizations that use different tools, there is no more locking to a tool suite and information can be easily reused from project to project.

When stakeholders are viewing or even entering information they are doing so from their perspective, with a view that is appropriate for their needs, roles and authority. The underlying platform is responsible for repurposing information instead of making each user responsible for digging information out of alien formats and complex reports or diagrams.

The process of integrating all of these forms of information is a long one. What OISP does is lay the foundation with a platform and architecture that is sufficiently flexible to deal with this diversity of information, based on established standards and technologies known as the semantic web. OISP then starts the process of integrating information with some of the current information structures most crucial to GSA. As this capability evolves over time, and it becomes more integrated into the fabric of the enterprise, it will increase in value by being used as the repository for more and more information which is incrementally more integrated, linked and accessible.

What is currently being put in place is an operational prototype for this knowledge platform as well as integration of key assets such as the FEA, OneGSA and Business cases supporting better decision making. Since the core of OSERA will be open source, this opens the door for collaborative efforts with other agencies and for a pervasive intellectual capital platform across the government.

1.2. Document Organization

This document is divided into the following sections:

- Section 1 of this document provides background information, including the GSA's goals and vision for information management, and a description of the approach used to define the target EA Data Architecture.
- Section 2 of this document is about the GSA's Information Model, a key component of the target EA Data Architecture. This section provides a description of how it is extended to include Federal Enterprise Architecture (FEA) Data Reference Model (DRM) 2.0 standardization areas.

- Section 3 of this document provides a description of the features of the OSERA Enterprise Knowledge Base (OsEra-EKB); the second key component of the target EA Data Architecture.
- Section 4 of this document discusses EA Data Architecture Governance; which is the third key component.

The appendices contain additional detail.

1.3. Background and Overview of the Target Data Architecture

The GSA faces a number of business challenges related to information, data and knowledge management:

- **The GSA Data Assets**: Like any large organization the GSA develops, maintains and uses a large and diverse body of data that is crucial to the effectiveness of the enterprise. This data suffers from a variety of issues, common to data that has "grown" over decades:
    - **Data is siloed**: Most data managed by GSA is "trapped" within the confines of a particular application, data center or organization. Where other stakeholders need this information, it is hard to get access to and use.
    - **Data is redundant**: Due, partially, to being siloed, data is redundantly gathered and managed – often leading to inconsistencies and errors. Data is trapped within a particular application or technology and is not easily repurposed for the needs of stakeholders. The redundancy also increases costs – costs to manage the redundant sources, costs to synchronize the sources and costs of dealing with errors.
    - **Data is inconsistently represented**: The meaning and definition of data elements and structures is inconsistent – it is neither integrated, centralized or federated.
    - **Data quality is an issue**: Many data sources have quality concerns without a clear way to resolve them (See the data quality guidelines).
    - **Data is not information**: The purpose of data is to inform. There are concerns with the meaning and scope of data as well as the form of its presentation and analysis such that it can be used to inform stakeholders.

- **GSA Architectures, Plans and Governance**: GSA has multiple efforts to address data concerns in the form of architectures, plans, processes, policies and governance. These efforts also have problems which have reduced the effectiveness of dealing with the above, bulleted, issues.
    - There is no agency wide body, process or resource that has the desire and authority to "solve the data problem". While it may be in the best interests of GSA, there is no incentive for the owners of information to solve agency-wide problems.
    - The GSA has acquired and accumulated a great deal of information, architectures and plans that are not well organized, integrated or maintained. To a large extent,

the GSA's information archtiectrues are siloed. There are instances where information, some of it costly, is lost, forgotten or simply not available or applied at the right time; plans and architectures are duplicated due to changes in contractors, tools, methodologies, technologies and management.

- Inconsistencies in GSA's planning, governance and architectural information exist and the linkages that help to provide additional meaning to information are missing in most cases. Decisions are made on inconclusive information after a significant amount of funding is spent on analysis; and the information available is not effectively transformed into knowledge.
- Architectural information is often represented in a manner that is only appropriate for architects, not business stakeholders. Frequently, information is limited to a single format that is not suitable to everyone or transferable. Consumers expect information assets and resources to be enablers not obstacles to accomplishing their tasks and delivering services.
- The general emphasis in government for increased accountability requires that managers have IT governance mechanisms in place. Compliance with legislative and government mandates continues to place a heavy burden on the GSA's ability to effectively measure the success of EA and investment programs.

1.4. Processes and Governance

Infrastructural Support

The GSA's Open Source E-Government Reference Architecture (OSERA) Infrastructure Services Platform (OISP) addresses some of these challenges by providing for the ability to manage strategic information assets and architectures. One of the key aspects of this capability is defined by the target EA Data Architecture: a vision for a semantic based, model driven knowledge platform leveraging internet technologies.

Target Data Architectures

The target EA Data Architecture centers on the development of the OSERA Enterprise Knowledge Base (OsEra-EKB). The OsEra-EKB is an enterprise-wide knowledge repository that contains, manages and provides access to the GSA's existing strategic information assets.[1] This OsEra-EKB is intended to enable the GSA to become a more effective organization, make decisions on the basis of knowledge, align with administration policy and advance significantly in both Office of Management and Budget (OMB) and Government Accountability Office (GA) maturity assessments.

[1]This repository will be populated with many of the existing strategic information assets including the One GSA EA, the OMB FEA Reference Models, the OMB Federal Transition Framework, the GSA ITAPC "Bricks", etc. Additional capability will be provided to include other strategic information asset types such as the PMP and the GSA Executive Business Case within and accessible through the repository.

Another component of the target EA Data Architecture is the GSA Information Model; which represents the GSA's vision for semantic-level models that extend beyond structural descriptions of information. The GSA's Information Model combines the original One GSA models with new information from the Financial Management Enterprise Architecture (FMEA) segment, Financial Management Enterprise Architecture Continuation (FMEA-C) segment, Contract Writing System segment and Integrated Portfolio Management initiative.

The OsEra-EKB and GSA Information Model are critical components needed to address the issues with information and knowledge management in order to help facilitate a more effective agency. However, achieving enterprise maturity requires the use of the capabilities available from the OsEra-EKB and GSA Information Model as well as a transformation of the culture and processes of the agency in concert with leadership as is described in Integrated Portfolio Management (IPM) project and strategic assessment. The third key component is the stewardship role to support the governance of the EA Data Architecture. It is important for the GSA's existing governance bodies to have collaborative groups that will actively come together to discuss and review how the GSA's critical information assets are being created and used. The Information Model and OsEra-EKB are components that will assist the GSA with facilitating stewardship and governance by providing easy access to and management of information and knowledge.

The combination of the Information Model, OsEra-EKB platform and supporting governance structure results in the GSA's vision for its EA Data Architecture. This vision is intended to enhance knowledge throughout the GSA and achieve the following goals:

- **Support collaboration, sharing and reuse**: By enabling and supporting a culture of shared information assets, services, and processes across GSA.
- **Enable strategic thinking and information driven decision making**: By providing a flexible, model driven tool to manage information assets and developing an integrated information model that aligns to GSA's business architecture.
- **Improve the value and quality of information**: By facilitating agency level comprehension of information and providing the capability to communicate within GSA and/or with partnering agencies about information with a clear understanding of its meaning and confidence in its quality.
- **Reducing duplicative efforts**: Redundancy in information and processes in GSA is not only time consuming and expensive, it is a source of inconsistent and sometimes contradictory information.

By providing a means to explore the opportunities for the integration of business processes and information, GSA will become more adept at consistent planning and execution of strategic initiatives.

1.5. Approach

The target EA Data Architecture is expressed as a set of shared metadata concepts using Web Ontology Language (OWL) and Resource Description Framework (RDF) in the ontology of architecture. These concepts will be synthesized from the current GSA data architecture, as expressed in previously developed EA artifacts, as well as the Federal Enterprise Architecture (FEA) Data Reference Model (DRM) 2.0. These concepts will be made available for use via the OsEra-EKB.

1.5.1. Model Integration

The target EA Data Architecture combines the original One GSA models with new information from the Financial Management Enterprise Architecture (FMEA) segment, Financial Management Enterprise Architecture Continuation (FMEA-C) segment, Contract Writing System segment and Information Technology Portfolio Management initiative. This integrated set of models utilizes both Enterprise Distributed Object Computing (EDOC) and Unified Modeling Language (UML) to represent the GSA's enterprise and segment architectures. A two-way mapping is defined and implemented from EDOC to instances of these shared concepts in the OsEra-EKB; and a one-way mapping is defined and implemented from these shared concepts to the FEA DRM 2.0.

1.5.2. Aligning to the Federal Enterprise Architecture (FEA) Data Reference Model (DRM) 2.0

A unifying data architecture is essential in providing a foundation for consistent information. The FEA DRM 2.0 establishes a common data model for the purposes of streamlining information exchange processes within the Federal government and between government and external stakeholders.. To that end, GSA, as per the agency's EA goals, is committed to adhere to the standards and guidelines expressed in the FEA DRM 2.0.

1.5.3. Transforming the Enterprise-Wide Repository to the OsEra-EKB

Many of the processes and artifacts to support the GSA's IT vision have already been defined as part of the Integrated Portfolio Management (IPM) project. The target EA Data Architecture enhances the utility of these essential products by providing automated support via the OsEra-EKB's change and configuration management capabilities, shared concepts, mapping components, views and user interfaces (see Section 3 and APPENDIX A:).

2. THE GSA'S INFORMATION MODEL AS THE DRM

The FEA DRM 2.0 is a framework in which the primary purpose is to enable information sharing and reuse across the federal government and promote uniform data management practices. The DRM provides a standard means by which information may be described, categorized, and shared. These standardization areas are expressed in the DRM Abstract Model and in GSA's Information Model as described in the following sections and detailed in the GSA's DRM Report.[2]

2.1. Data Description

The Data Description standardization area provides a means to uniformly capture the semantic and syntactic structure of information. The GSA's Information Model is an EA Data Architecture artifact that clearly describes the meaning of common information through the use of metadata and shared concepts. The GSA's Information Model shall be used to ensure that the agency's information assets have consistent and complete representations so as to make them comprehensible and usable by any interested, authorized party within the agency and beyond.

2.2. Data Context

The Data Context standardization area establishes an approach to the categorization of information assets using taxonomies and other descriptive information. Data Context also enables discovery of information, and provides linkages to the other GSA reference models (e.g. Business Reference Model). The Information Model is integrated with SOA concepts as expressed in EDOC and associated UML profiles. For this reason, the Information Model is part of the overall architectural approach used by the GSA to contextualize its information assets. The other part of the approach involves process analysis and collaborations among domain experts to define the data layer of GSA's EA based on the business architecture.

2.3. Data Sharing

The Data Sharing standardization area describes the access and exchange of information, where access consists of recurring requests (e.g. a query of an information asset), and exchange consists of fixed, recurring information exchanges between parties. Data sharing is enabled by the OsEra-EKB capabilities described in Section 3 of this document. The GSA's Information Model describes the business concepts and terms that are individual data elements that are aggregated into message taxonomies. These "messages", or "business transactions", can be transformed directly into XML schemas, and their instances get attached to the services information exchanges, as defined by the collaborative role interactions in the business architecture.

[2]The DRM Report is a work product of the GSA Enterprise Architecture Data Architecture project. It will be accessible via www.osera.gov.

2.4. Key Benefits of the Information Model

The GSA's Information model represents the semantic structure that is common to multiple information resources across the agency. The key benefits of the Information Model include:

- The Information Model is expressive and provides for the accurate, easy interpretation of information based on a business context.
- Semantic features of the Information Model enables enhanced automated discovery and usage of information due to the enhanced meaning that is provided for information. The result is information consumers feel confident and empowered to pose intelligent, ad hoc queries for decision support.
- Due to the platform-independent nature of the Information Model, individual domains can leverage components of the Information Model for development of applications that are specific to their domain or business unit.
- The Information Model supports the creation of common entities by providing a means to compare and relate information assets across GSA through a common, well-defined structure.

3. THE OSERA ENTERPRISE KNOWLEDGE BASE (OsEra-EKB)

The OsEra-EKB provides the ability to search and retrieve information assets for any given concept; administer information, and create articles about information assets. Through semantic metadata capture, information assets will be categorized and contextualized to ease search and retrieval. This affords information suppliers and consumers the capability to easily determine what information is available, the subject of the information, and how to obtain the information. Additionally, when information components are visible, managed, and accessible within the GSA community, information suppliers and consumers will be more inclined to share and reuse common information rather than re-inventing it.

3.1. Utilities and Features

The OsEra-EKB is an operational, open source prototype using ontologies, semantic web standards and technologies to integrate, transform and repurpose GSA information. The knowledge base supports various kinds of information in a variety of tools and formats. The OsEra-EKB provides a means for defining and shared concepts that are used for the categorization of information assets. The OsEra-EKB platform includes a set of utilities that leverage shared concepts and provide the following "ontology model driven" capabilities:

1) **User Interface:** The user interface consists of a simple web based forms interface, allowing information to be entered, categorized and related.[3]
2) **Query:** The query interface provides a user friendly layer over SPARQL to query the knowledge base, locating information for browsing, editing or analysis. Query will also be able to export into standard XML files for further processing with widely available tools.
3) **Upload/Download:** The upload/download utility will accept data in any format to provide configuration management and categorization of that data in the repository. Data in supported artifact formats (e.g., EDOC, DRM, BPMN) will be able to be mapped to the knowledge repository Articles directly. Artifacts linked to the knowledge repository will automatically update that repository when checked in and reflect any changes to the repository when checked out. This simple "check in/check out" paradigm for linked artifacts presents a very simple interface to the leading-edge capability underneath that maps between the data and file formats using shared concepts.
4) **Mapping Facility**: The mapping facility will implement the generic infrastructure for mapping between ontologies and for import/export of external artifacts in XML Metadata Interchange (XMI). Note that it is not the intent of this task to define new languages, ontologies or methods for mapping but to provide a framework where multiple ways to define or implement mappings may be used together in support of the OsEra-EKB. The mapping facility will be component oriented and will map import export components to source and target requirements.

[3]Virtually any information asset that can be rendered as a "file" will be stored, managed, versioned and located. The metadata about the information asset will be delivered to any browser or desktop.

Once shared concepts are defined for a language an "adapter" will be developed to map between the XMI representations of each artifact to instances of the shared concept ontology. These adapters will use a "change" approach such that the entire artifact is not mapped each time it is checked into the repository. Only changes to the artifacts will be used. The change based approach serves to keep the external artifacts intact and not require that every model element in the external element need be mapped to an instance of a shared concept (since many concepts are not shared). A change-based integration utilities for diff/merge will be created for XMI and RDF. These resulting changes will be the medium of exchange between shared concepts and artifacts. This will also allow a change in an artifact to propagate to instances of shared concepts which will then propagate to other external artifacts (which are linked to those shared concepts).

3.2. Key Benefits of the OsEra-EKB

The OsEra-EKB will increase the GSA's core capabilities, support the development enterprise-level information services and mission critical applications and serve as a basis for information modeling standards. Some of the key benefits include:

- Complete and current information will be captured and stored for use by the entire agency.
- Supports an information sharing environment because information is made generally available and accessible to any interested party through a simple, user-friendly interface; and provides a platform to migrate internal agency information assets from internal resources to cross-agency information assets where appropriate. [4]
- Provides efficient discovery features so that information consumers can obtain information, formulate answers to business questions and exploit knowledge for better business decisions.[5]
- Reduces the need for labor intensive collection and reconciliation activities to satisfy requests for information.
- Increases stakeholder collaboration by providing a central store for information assets; enhances the ability to implement standards and best practices by providing insight into existing resources.

[4]Sensitive information will be handled according to GSA security standards and policies.
[5]For example, the information in a spread sheet that is part of a human capital study can be used for a business case that is directly related to the systems architecture of an application being funded and then implemented.

4. GOVERNANCE OF THE DATA ARCHITECTURE

Governance is a critical aspect of successful EA. It provides a link between business and IT to facilitate the development of a common way of understanding shared business information. Governance enables agreement with architecture policies and standards set forth to benefit the GSA. As the EA continues to mature and evolve, the governance structure can enable consistent application of EA standards; and the development and use of information assets and/or resources.

Governance of the EA Data Architecture involves people, processes and technologies for sharing, recording and using information. It encourages the understanding and management of information from both business and technical perspectives, and it advocates the importance of information as a valuable resource, allowing GSA to consume and share information confidently to satisfy business needs and regulatory requirements. Other benefits include:

- Increased standardization of information models and Data Architecture practices;
- Increased ability to integrate information assets, eliminate redundancies, while preserving necessary information assets;
- The ability to extract knowledge offering high levels of visibility and alignment with IT strategic objectives; and
- The ability to adapt and evolve information models as GSA's key business processes change.

Governance of the GSA EA Data Architecture shall be a component of the GSA's EA governance structure. The intent is to leverage the sponsorship and organization of the EA Program Management Office (EAPMO) that is already in place and adopt existing management policies and practices in coordinating stakeholders and stewards to create an actionable management plan for the EA Data Architecture.[6] Working groups consisting of stewards, architects and subject matter experts (SMEs) from business and IT shall convene on an as needed basis to provide authority over standards and develop solutions for EA Data Architecture related issues.

Additionally, the OsEra-EKB shall be leveraged as a mechanism to encourage a strategy for architecture governance and provide support for governance activities. For example, the OsEra-EKB can be used as a tool for identifying stewards and their relationships to the information that they steward. The knowledge repository can aid in coordinating key governance activities by storing plans, procedures and critical correspondence; and locating stewards so that the appropriate individuals are engaged at the appropriate time during the course of an initiative. This idea is based on the existing knowledge management and transfer capabilities that the OsEra-EKB has to support the Information Model and other EA and IPM artifacts that are relative to governance.

[6]Executive-level sponsorship is assumed through the EAPMO and EA governance body, therefore it is not discussed in this document as an explicit role in EA Data Architecture governance and stewardship.

APPENDIX A: OsEra-EKB COMPONENTS (LAYERS)

Figure A-1 – OsEra-EKB Structure

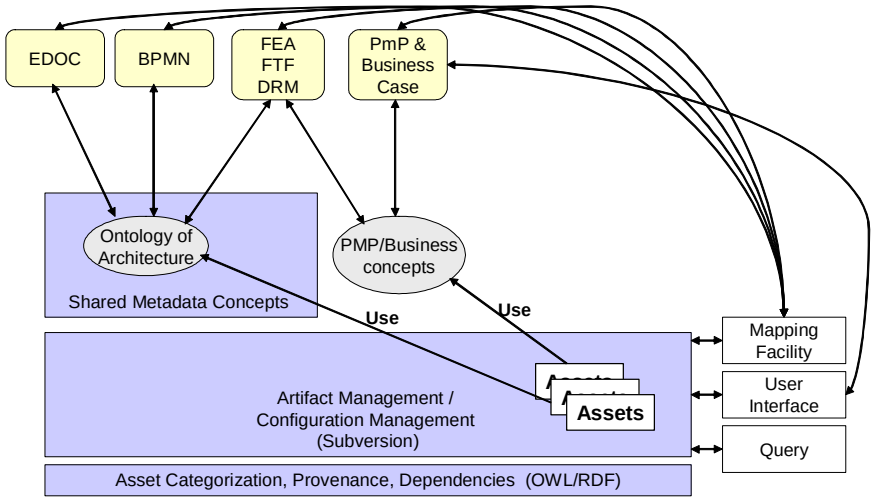

Figure A-1 – OsEra-EKB Structure illustrates the layered structure of the OsEra-EKB. The OsEra-EKB shall be layered as shown and described below:

- **Artifact Management Layer**: (Artifact Management/Configuration Management in Figure A-1) Artifact Management is the foundation of the OsEra-EKB. It is based on the Subversion (SVN) (see: http://en.wikipedia.org/wiki/Subversion_(software) and http://subversion.tigris.org/ for further details) configuration management system, which provides a versioned, distributed, federated and reliable data storage for information assets or artifacts. The OsEra-EKB assigns a URL to each artifact, making the artifact a web resource that is available to any user that has permission to access the artifact.
- **Information Asset Ontology Layer**: (Asset Categorization, Provenance, Dependencies in Figure A-1) The next layer, Information Asset Ontology, augments SVN metadata with a OWL-RDF ontology that categorizes information assets, records their dependencies, their context, their origin and the syntax in which they are expressed. [7] This ontology for managing information assets is a component of the shared concept system and integrates concepts from existing standards (e.g. Simple Knowledge Organization System and Reusable Asset Specification).

  The OsEra-EKB places some conventions and restrictions on the physical directory structure of the repository to allow for flexibility in naming and categorizing information assets in a variety of ways for different purposes. The "concepts" used to categorize information can be user defined and evolve over time and be utilized in different contexts with different terms.

[7] RDF – Resource Description Framework (http://www.w3.org/RDF/)

The Information Asset Ontology is dynamically synchronized with the SVN repository allowing any change to be recognized and the results processed as updates to the RDF store. Correspondingly, changes to the RDF store can result in changes to the assets or the asset metadata in SVN.[8]

- **Ontology of Architecture Layer**: (Shared Metadata Concepts in Figure A-1) Shared metadata concepts are a distinguished set of concept hubs that normalize the shared concepts relative to architecture as are found in the FEA, Federal Transition Framework (FTF), DRM, EDOC, Business Process Modeling Notation (BPMN) and current GSA architectures. Shared metadata concepts are concepts used to define other concepts and are usually found in modeling languages and ontology languages. Additionally, shared metadata concepts are the basis for the integration of information in different tools, standards and methodologies.

  The Ontology of Architecture is comprised of definitions of those shared concepts that are found in two (2) or more of the reference languages. The set of architectural shared concepts are synthesized from FEA, FTF, DRM, EDOC and BPMN to augment the "semantic core" work already done as part of OSERA. Adapters and transforms are used to project these shared concepts onto external artifacts in the native syntax of FEA, FTF, DRM, EDOC or BPMN.[9]

[8] As information is managed in the repository, the RDF store for the OWL ontology will be available to provide SPARQL queries to locate and analyze information in the OsEra-EKB. This query function is provided through a user friendly web interface to users.

[9] This will then make the GSA architectural assets and the FEA a part of the same conceptual framework.

APPENDIX B: STEWARDSHIP ROLES IN A FEDERATED DATA ENVIRONMENT

Any individual or organization within GSA involved in stewardship has certain responsibilities with respect to the information integral to their business, regardless of the technology that is used to store and manage it. These responsibilities include plans and procedures for:

- Identifying and publicizing clear, concise and valid business definitions for information, considering both business and IT stakeholder perspectives.
- Identifying required business validation rules to be applied when capturing and processing data.
- Taking a proactive approach to improve the accessibility, integrity, timeliness, accuracy and completeness of data; assessing and resolving data quality issues.
- Facilitating sharing and usage of the data while addressing access and security issues that may pertain to certain data.
- Ensuring that existing and future data structures are developed according to established architecture practices and standards; and are flexible enough to meet the needs of the various types of consumers.
- Using appropriate existing communications channels and methods to engage business and IT stakeholders.

GSA's information environment is geared towards a federated structure where information is categorized as distinct, common and core. For this reason, a federated information stewardship structure where there is a corresponding steward for each category of information is recommended so that the appropriate individual or group is utilized at the right time. Table B-1 provides a high-level description of this concept.[10]

Table B-1 – Stewardship for Distinct, Common and Core Information

Category	Level of Stewardship
Distinct - This term applies to business information elements that are exclusive to a business unit.	Operational-level stewards; information suppliers and consumers that create, define and read information.
Common - This term applies to business information elements that are available to at least two business units, but not available to all units.	Multiple domain stewardship with coordination.
Core - This term applies to elements that are available to all business units; corresponding to GSA-wide business elements.	Enterprise-level stewardship represented by a council or working group.

Enterprise Level Stewardship

[10]Stewards should not be difficult to find. It is a matter of identifying those who are already filling the data governance responsibilities as part of their current job function at GSA. For example, they may be an information architect , business analyst, program manager, information consumer and/or information supplier.

Data Architecture related decisions about information that will have an impact across GSA (e.g. core components), need to be made by a crosscut of individuals that are in the position to represent their business units in the decision making process. Strategic decision making about enterprise-level Data Architecture also requires that the individuals making the decisions are knowledgeable about information assets and are equipped with sufficient documentation to help them make the right decisions. The concept of a working group or data council satisfies this requirement for a assembly of individuals who are responsible for the strategic decisions being made around the definition, production and usage of enterprise-level information.

In terms of key roles and responsibilities, the data council or working group:

- Possesses an interest in Data Architecture governance due to an interest in improving GSA information management practices.
- Understands the definition of Data Architecture governance and the benefit of governance to GSA.
- Develops and executes a plan for Data Architecture governance and works to support the coordination of information stewardship activities.
- Participates in the approval of data policies, management plans, and governance tools.
- Ensures that Data Architecture practices are agreed to, improved and continuously promulgated to business units.
- Make strategic-level decisions in a timely manner given the appropriate knowledge to make that decision.
- Manages participation in discussions to remain informed of Data Architecture governance activities.
- Identifies and coordinates critical data governance roles for core information (e.g. domain stewards and coordinators).

Domain Stewardship

Domain stewards fall within a business unit and are usually determined by the logical position or natural relationship to the domain of information. For example, the logical choice for a domain steward of financial data is the Chief Financial Officer (CFO) role. Domain Stewards may also be designated or volunteer to play the role. Despite the method by which domain stewards are identified, they should be knowledgeable about their respective domain of information and work to facilitate acceptable standards of data within their domain and facilitate acceptable resolution of issues pertaining to the information in their domain.

In terms of key roles and responsibilities, the Domain Steward:

- Supports consensus building around how information should be represented and the development of standards for information within their domain.
- Leads initiatives to resolve cross business-unit issues pertaining to information in their domain.

- Ensures that there is adequate documentation collected and shared among all stakeholders on how information within their domain is being defined, produced and used.
- Identifies operational-level stewards of their respective business unit and acting as a liaison to ensure that information standards and policies are communicated to operational-level stewards.

Operational -Level Stewardship

Operational-level stewards ensure that information definition, production and usage activities adhere to architecture policies and standards. There are three (3) categories of stewards at the operational-level.[11]

- **Information Definition Stewards**: Information Definition Stewards play a critical role in defining and documenting the information that is required to operate within their business unit and to maintain the values and consistency of this information. They are also responsible for assisting with identifying and leveraging existing data resources for reuse.
- **Information Production Stewards**: Information Production Stewards are responsible for ensuring that all information is created or passed into systems in accordance with the definitions set forth by Information Definition Stewards in a timely manner.
- **Information Usage Steward:** Information Usage Stewards are responsible for understanding information and using the information for its intended purposes.

Preparing for Stewardship

The first priority to implementing stewardship is establishing focus and the value proposition. It is important to understand how a stewardship structure for the EA Data Architecture would contribute to the GSA's EA and IT strategic objectives while maintaining a low level of complexity and cost for such a structure. To establish a focus and value proposition for stewardship, the EAPMO should provide support by facilitate discussions with stakeholders about what drives the need for stewardship (e.g. the EA Target Data Architecture, knowledge repository, IT management initiatives, etc.).

The second priority is to prepare a roadmap with milestones to communicate to stakeholders the plan for how the vision for a stewardship component for the EA Data Architecture will be achieved. This roadmap shall also include key roles and responsibilities of individuals and/or groups within the existing EA governance body during planning. Once the roadmap is understood and agreed to by stakeholders, the EAPMO and can consider design activities for the EA Data Architecture governance component.

[11]The responsibility to accept, propagate and comply with enterprise data standards is the primary responsibility of all data stewards.

APPENDIX C: GLOSSARY OF RELEVANT TERMS

This section provides a glossary of terms that are used throughout this document.

Table C-1 – Relevant Terms

Term	Description
Architecture	Representation of the structure of a system or community that describes the constituents of the system and how they interact with each other such that the goals and responsibility of the system or community are met.
Best Practice	A group of tasks that optimizes the efficiency or effectiveness of the business discipline or process to which it contributes. Best practices are generally adaptable and replicable across similar organizations or enterprises - and sometimes across different functions or industries.
Business Objective	Objectives state what is to be achieved, and the results and activities required to measure progress towards reaching the desired state.
Business Process	A business process is one aspect of a business model intended to specify the services, participants, interactions, resources and course of activities required to realize business value. A business process is a set of linked activities that create value by transforming an input into a more valuable output. Both input and output can be artifacts and/or information and the transformation can be performed by human actors, organizations, machines, or both. A business process can be decomposed into activities that may be atomic (E.G. “Delete file”) or utilize sub-processes (E.G. “Build Ship”), which contribute to achieving the goal of the super-process. The analysis of business processes typically includes the mapping of processes and sub-processes down to activity level. A business process may specify how processes are currently executed or may specify a future-sate process intended to improve business value and/or reduce costs.
Business Process Modeling Notation (BPMN)	A graphical notation that depicts the steps in a business process, as defined by the Object Management Group (OMB). BPMN depicts the end to end flow of a business process. The notation is designed to coordinate the sequence of processes and the messages that flow between different process participants in a related set of activities.
Business Rules	The assertions that define information from a business perspective.
Common	Elements at this level are available to at least two business units but not to every one.
Core	Elements at this level are available to every business area.
Concept	What something “means” in a given context.

Term	Description
Information Quality Improvement	The process of improving information quality to the level desired to support the enterprise information demand.
Data Reference Model (DRM)	The Data Reference Model (DRM) describes, at an aggregate level, the data and information supporting government program and business line operations. This model enables agencies to describe the types of interaction and exchanges occurring between the Federal government and citizens. The DRM categorizes government information into greater levels of detail. It also establishes a classification for Federal data and identifies duplicative data resources. A common data model will streamline information exchange processes within the Federal government and between government and external stakeholders. The DRM provides a standard means by which data may be described, categorized, and shared. These are reflected within each of the DRM's three standardization areas: – Data Description: Provides a means to uniformly describe data, thereby supporting its discovery and sharing – Data Context: Facilitates discovery of data through an approach to the categorization of data according to taxonomies; additionally, enables the definition of authoritative data assets within a community of interest (COI) Data Sharing: Supports the access and exchange of data where access consists of ad-hoc requests (such as a query of a data asset), and exchange consists of fixed, re-occurring transactions between parties.
Diff/Merge	A tool or function that assists with comparing, merging, and synchronizing text or source code.
Domain	A scope of information definition. A domain defines a collection of information generally recognized as appropriate to a field of study, a business process or function, or mission.
Distinct	These elements are exclusive to a business unit.
DRM Abstract Model	The DRM abstract model depicts the major concepts from each DRM standardization area and the relationships between them.

Term	Description
Enterprise Distributed Object Computing (EDOC)	EDOC is the OMG standard for role based collaborative modeling in support of MDA and SOA. GSA has adopted EDOC as the basis for One-GSA and Model to Integrate. EDOC provides a standard model and notation for describing collaborations at both the business and technical level. At the level of the business model EDOC collaborations represent business roles and responsibilities that collaborate by providing and using services described as business protocols. Technical EDOC models describe enterprise components that realize the business interactions as service interfaces across the enterprise service bus. Both roles and components may be further specified based on the choreography of activities and interactions required to provide and use these services. Once the technical components are defined they are then the bases for acquiring these components (by wrapping legacy systems, purchasing components or implementing new components) that are then deployed on the enterprise service bus (ESB) and can be then tested for compliance in support of model based testing.
Enterprise Architecture (EA)	A process and the associated strategic information asset base to support enterprise objectives, which includes: – the mission; – information necessary to perform the mission; – the technologies necessary to perform the mission; – the transitional processes; and – the business model for supporting the transformation of an enterprise to meet its changing business objectives.
Extensible Markup Language	Extensible Markup Language is a non-proprietary subset of SGML (Standard Generalized Markup Language). It is focused on data structure and uses tags to specify the content of the data elements in a document.

Term	Description
Federal Enterprise Architecture (FEA) Reference Model	The FEA consists of a set of interrelated "reference models" designed to facilitate cross-agency analysis and the identification of duplicative investments, gaps and opportunities for collaboration within and across agencies. Collectively, the reference models comprise a framework for describing important elements of the FEA in a common and consistent way. Through the use of this common framework and vocabulary, IT portfolios can be better managed and leveraged across the federal government. This chapter introduces the purposes and structures of the five FEA reference models: Performance Reference Model (PRM); Business Reference Model (BRM); Service Component Reference Model (SRM); Technical Reference Model (TRM); Data Reference Model (DRM). [Source: OMB Federal Enterprise Architecture Program Management Office; FEA Consolidated Reference Model Document, version 2.1; December 2006; page 5.]
Federal Transition Framework (FTF)	The Federal Transition Framework (FTF) is a single information source for cross-agency information technology (IT) initiatives using a simple, familiar and organized structure. It contains government-wide IT policy objectives and cross-agency initiatives including: – OMB-sponsored initiatives, e.g., E-Gov and LoB initiatives – Government-wide initiatives, e.g., Internet Protocol Version 6 (IPV6), Homeland Security Presidential Directive 12 (HSPD 12)
Information	The term "information" means any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms.
Information Architecture	The framework for organizing the planning and implementation of information resources. The set of information, processes, and technologies that an enterprise has selected for the creation and operation of information systems.
Information Model	Shows the relationships or linkages between major areas of interest to the business creates a shared business vocabulary (e.g. semantic model), defining a community's agreement on important concepts and relationships between those concepts.[12]
Information Life Cycle	The stages through which information passes, typically characterized as creation or collection, processing, dissemination, use storage, and disposition.
Information Quality Improvement	The process of improving information quality to the level desired to support the enterprise information demand.

[12][Source: Treasury Board of Canada, Chief Information Officer Branch; Business Transformation Enablement Program Glossary ; online: http://www.tbs-sct.gc.ca/btep-pto/documents/2004/gloss/gloss03_e.asp ; published October 7, 2004. Accessed December 11, 2007.]

Term	Description
Information Technology (IT)	The term 'information technology', with respect to an executive agency means any equipment or interconnected system or subsystem of equipment, that is used in the automatic acquisition, storage, manipulation, management, movement, control, display, switching, interchange, transmission, or reception of data or information by the executive agency.
Initiative	A project or program used to implement a new capability or improve an existing capability. Initiatives need time and resource commitments and should be aligned with the organization's strategy.[13]
Integrity	Refers to the security of information from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.
Integrated Portfolio Management (IPM)	Integrated IT Portfolio Management (IPM) joins and harmonizes internal policies with Capital Planning Investment and Control (CPIC), Systems Development Life Cycle (SDLC), and other processes and activities related to the inception, elaboration, construction, and management of IT initiatives across GSA.
IT Strategic Plan	A document representing GSA's blueprint for utilizing information technologies to support the President's Management Agenda and the Agency's Strategic Plan. The Plan identifies both the mission achievement and the management improvement strategies that will support enterprise efforts to improve customer service, increase fee-for-service revenues, lead to shared services and capabilities, and ensure advancement towards a service oriented architectural environment that improves other enterprise business outcomes across GSA's value chains.
Measure	The criteria, metric or means to which a comparison is made with output.
Metadata	Semantic information associated with a given variable; includes business definitions of the data and clear, accurate descriptions of data types, potential values, original source system, data formats, and other characteristics. Metadata defines and describes business data. Examples of metadata include data element descriptions, data type descriptions, attribute/property descriptions, range/domain descriptions, and process/method descriptions.
Metric	A standard for measurement.
Mission	The stated purpose of an Organization. A Government Agency's mission must reflect its authorizing legislation. The GSA mission is: We help federal agencies better serve the public by offering, at best value, superior workplaces, expert solutions, acquisition services and management policies.

[13]Source: COLAB Collaborative Work Environment's Wiki http://colab.cim3.net/cgi-bin/wiki.pl?initiative ; Last accessed December 11, 2007.

Term	Description
Objectivity	Involves a focus on ensuring that information is accurate, reliable and unbiased and that information products are presented in an accurate, clear, complete and unbiased manner
One GSA Enterprise Architecture	The One GSA Enterprise Architecture is GSA's Business Modernization Blueprint that provides a framework and methodology for modeling business processes, stakeholder groups, and the information flows between these groups, based on a Model-Driven Architecture (MDA) approach. The One GSA EA encompasses and links both the GSA business models (processes, services, roles responsibilities and goals) with the systems architecture (IT services, components, applications and systems) into a coherent and executable architecture to realize business objectives.
Ontology	An ontology is a controlled vocabulary that describes objects and the relations between them in a formal way, and has a grammar for using the vocabulary terms to express something meaningful within a specified domain of interest. [Source: members.optusnet.com.au/~webindexing/Webbook2Ed/glossary.htm. Note: text and URL displayed as part of Google results when searching for term "ontology definition" but URL not accessible on December 22, 2006.]
Open Source eGov Reference Architecture (OSERA)	GSA's Open Source eGovernment Reference Architecture (OSERA). OSERA's mission is to provide government and industry with a reference architecture and ready-to-run toolset and platform that enables business objectives to be realized more quickly and less expensively using an integrated approach to enterprise, business and technology architectures. OsEra is based on the principles of MDA, SOA and Model to Integrate based on open standards and open source implementations.
Performance Reference Model (PRM)	The PRM is a "reference model" or standardized framework to measure the performance of major IT investments and their contribution to program performance. The PRM has three main purposes: – Help produce enhanced performance information to improve strategic and daily decision-making; – Improve the alignment — and better articulate the contribution of — inputs to outputs and outcomes, thereby creating a clear "line of sight" to desired results; and – Identify performance improvement opportunities that span traditional organizational structures and boundaries
Procedure	A course of activities describing how to carry out a process.
Program	An organizational unit within GSA with responsibility for delivering on a clearly defined mission or service. Programs may include one or more Projects. Projects may fall under one or more Programs.

Term	Description
Project	Planned undertaking by within a program to implement a solution identified in a Segment Architecture Transition Plan. Projects may include one or more investments. Investments may fall under one ore more Projects
Reproducibility	Means that the information is capable of being substantially reproduced, subject to an acceptable degree of imprecision.
Resource Description Framework (RDF)	The Resource Description Framework (RDF) integrates a variety of applications from library catalogs and world-wide directories to syndication and aggregation of news, software, and content to personal collections of music, photos, and events using XML as an interchange syntax
Reusable Asset Specification (RAS)	The specification defines a standard way to package reusable software assets. http://www.omg.org/technology/documents/formal/ras.htm
Role	A role is a representation of a set of capabilities and responsibilities for work to be carried out within the context of one or more business processes. The role can then be assumed, or played, by one or more individuals, organizations or systems.
Service	Discrete unit of functionality offered by one party and then able to be consumed by another. At the business level the service is the actual delivery of business value and the specification of the interactions required to deliver that value. In a technical architecture a service defines the interface to a component realizing a role and returning either the business value directly or a document describing the business value.
Service or Staff Office	A Program Office within GSA responsible for coordinating nationwide programs and supporting Federal agencies and citizen-oriented organizations.
Service Reference Model (SRM)	The Service Component Reference Model (SRM) is a business and performance-driven, functional framework that classifies Service Components with respect to how they support business and/or performance objectives. The SRM is intended for use to support the discovery of government-wide business and application Service Components in IT investments and assets. The SRM is structured across horizontal and vertical service domains that, independent of the business functions, can provide a leverage-able foundation to support the reuse of applications, application capabilities, components, and business services.

Term	Description
Simple Knowledge Organization System (SKOS)	SKOS is an area of work developing specifications and standards to support the use of knowledge organisation systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web. http://www.w3.org/2004/02/skos/
SPARQL	SPARQL (pronounced "sparkle") is an RDF query language; its name is a recursive acronym that stands for SPARQL Protocol and RDF Query Language
Strategy	A broad course of action or methodology designed to achieve a Long-Term Outcome Goal.
System	Any organized set of constituent parts that behaves according to a unifying plan or principle. Organizations and communities as well as computer programs or hardware may all be considered “systems”. Architectures are frequently used to describe the current and/or desired state of systems with respect to a particular goal set. Every division or aggregation of real objects/entities into systems is arbitrary, therefore it is an abstraction.
System Development Lifecycle (SDLC)	System Development Life Cycle, or SDLC, is the process used to develop an information system, including requirements, validation, training, and user ownership through investigation, analysis, design, implementation and maintenance. SDLC is also known as information systems development or application development. An SDLC should result in a high quality system that meets or exceeds customer expectations, within time and cost estimates, works effectively and efficiently in the current and planned Information Technology infrastructure, and is cheap to maintain and cost-effective to enhance. SDLC is a systems approach to problem solving and is made up of several phases, each comprised of multiple steps.
Technical Reference Model (TRM)	The Technical Reference Model (TRM) provides a foundation to categorize the standards, specifications, and technologies to support the construction, delivery, and exchange of business and application components (Service Components) that may be used and leveraged in a Component-Based or Service-Oriented Architecture. The TRM unifies existing Agency TRMs and E-Gov guidance by providing a foundation to advance the re-use of technology and component services from a government-wide perspective.
Transparency	The inclusion of source references (citations) and possible clarifications for the information being disseminated.
Unified Modeling Language (UML)	The industry-standard language for specifying, visualizing, constructing, and documenting the artifacts of software systems. It simplifies the complex process of software design, creating a "blueprint" for construction.
Utility	The usefulness of the information to the intended users.

Term	Description
Web Ontology Language (OWL)	The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies, and is endorsed by the World Wide Web Consortium.

APPENDIX D: REFERENCES

1) Federal Enterprise Architecture (FEA) Data Reference Model - http://www.whitehouse.gov/OMB/egov/a-5-drm.html

2) GSA IT Strategic Plan 2009 – 2011 http://www.gsa.gov/gsa/cm_attachments/GSA_DOCUMENT/strategicplansept4_R2EH1F_0Z5RDZ-i34K-pR.ppt

3) Object Management Group (OMG) MetaObject Facility (MOF) - http://www.omg.org/technology/documents/formal/MOF_Core.htm

4) One GSA Enterprise Architecture, Version 3.0, Volume 1: Overview

5) One GSA Enterprise Architecture Policy (CIO 2110.1)[14]

6) Resource Description Framework (RDF) - http://www.w3.org/RDF/

7) Reusable Asset Specification (RAS) – (http://www.omg.org/technology/documents/formal/ras.htm)

8) Roadmap for Semantics in Netcentric Enterprise Architecture - http://osera.modeldriven.org/documents/Roadmap%20for%20Semantics%20in%20Netcentric%20Enterprise%20Architecture.pdf

9) Semantic Web - http://www.w3.org/2001/sw/

10) Simple Knowledge Organization Systems (SKOS) - http://www.w3.org/2004/02/skos/

11) Web Ontology Language (OWL) - http://www.w3.org/TR/owl-features/

12) Wikipedia - http://en.wikipedia.org/wiki/Wikipedia

[14]This order was only published electronically.