U.S. Environmental Protection Agency

INTO ACTION: Enterprise Architecture Status Report 2004

September 13, 2004

This page is intentionally left blank.

Table of Contents

List of Tables and Figures

Attachments

CIO's Foreword

"Into action." That's my take on what's happened with the architecture this year.

That action is translating into better services for citizens and better internal performance. Whether it's the better quality of data coming in over the Exchange Network, or more efficient processing of grants through links to Grants.gov, the target architecture is improving the way we do business at EPA.

As this report goes to press, we're working on the initial build-out of our cross-Agency Central Services infrastructure—building our first Central Services data warehouse and setting up the EPA Portal. Over the summer the Program Management Office selected the necessary tools, built development platforms, and started the first phases of a comprehensive, architecture-driven upgrade to the Office of Air's ambient air quality systems. We're upgrading network communications with our air quality partners to include up-front data quality checks, linking EPA's ambient air quality systems to our geospatial analysis tools, and are about to launch a national air quality notification system. For the first time, citizens will be able to request automated e-mails to alert them of upcoming air quality problems in their home metropolitan areas.

Meanwhile, Administrator Mike Leavitt has signed EPA's Interim Enterprise Architecture (EA) Policy and the EA Team is developing a comprehensive architecture governance system to implement that policy. Agency work groups are ironing out the details for formal change management of the architecture, from baseline to target to sequencing plan. They will create an operational architecture governance system that formally integrates architecture into the capital investment process, the system life cycle process, and the IT security process.

In addition, our target Research and Science Architecture is complete, creating a long-term design for high performance supercomputing grids, improving the quality of science data management, and implementing an Environmental Science Portal and Science Subnet that will provide a quantum leap forward in science collaboration with our state and university science partners.

As you browse through this status report, these and a host of other initiatives underscore how much the architecture has already done to improve our performance, and how many people have worked together to make it happen. As I look back on the last four years, our achievements are gratifying to say the least. The Exchange Network, the Enterprise Architecture program, the Environmental Indicators program—all are part of a business and information technology modernization effort that has literally taken us into a new millennium. I'm proud to have been a part of it all, and thank all of my colleagues and staff for their creative work and tireless efforts.

KIMBERLY T. NELSON

Assistant Administrator and
Chief Information Officer

Office of Environmental Information
U.S. Environmental Protection Agency

September 2004

This page is intentionally left blank.

Highlights

Moving the Architecture Program Forward

- The U.S. Environmental Protection Agency (EPA) completed the target Research and Science Architecture (RSA), setting out six focus areas for improving the management of scientific information Agency-wide and taking action to build new systems to support scientific collaboration within EPA and government-wide. Focus areas include science information management, quality management, scientific and high-performance computing, geospatial analysis, collaboration, and standard technology solutions.
- The Office of Solid Waste and Emergency Response (OSWER) is leading an Agency-wide Cleanup Architecture to unify information management support for site and area cleanup.
- The Office of Prevention, Pesticides, and Toxics Substances (OPPTS) is developing a business process modeling initiative to support OPPTS architecture development and lay the groundwork for improving OPPTS business processes.
- The Administrative Systems Architecture (ASA) is positioned to support the federal Financial Management, Grants Management, and Human Capital Management Lines of Business initiatives.
- EPA is upgrading the enterprise architecture (EA) governance program to make the architecture actionable. The Administrator has signed the Interim EPA EA Policy.

Building the Target Architecture

- EPA has identified the key technologies necessary to build out the EPA Portal and to construct the Data Integration Platform (DIP). The DIP will manage metadata centrally and populate data warehouses and data marts.
- The Central Services build-out began as a collaboration between the Office of Environmental Information and the Office of Air and Radiation. Projects include:
  - Development of the EPA Portal Air Analysts' Portal, providing access to the Air Quality Subsystem (AQS) and other EPA and Office of Air and Radiation (OAR) systems;
  - Construction of the Data Integration Platform and the first AQS Data Mart;
  - Implementation of two-way data flows over the Exchange Network through the Central Data Exchange (CDX);
  - Creation of an air quality notification system (EnviroFlash) for selected metro areas;
  - Implementation of new geospatial portal tools to support the Office of Air.
- The Exchange Network has added five new data flows over the past year and has 13 state nodes in operation. Twice as many files were submitted through CDX in the first half of 2004 than in all of 2003.
- The first phase of the RSA Science Subnet is providing convenient exchanges of high volume data across EPA's science community. The Environmental Science Portal will provide the science community with collaborative tools and enhanced access to high performance computing.
- EPA is linking several of its internal administrative systems to government-wide systems.

This page is intentionally left blank.

Introduction

This Status Report describes the third year of development of the Environmental Protection Agency (EPA) enterprise architecture (EA). It follows the December 2002 submission to the Office of Management and Budget of the *EPA Target Enterprise Architecture,* which presented the full scope of EPA's enterprise architecture program, and the September 2003 *Enterprise Architecture Status Report 2003*, which described EPA's progress during the initial year of the program's implementation. This report also builds on the completion of EPA's *2004 Sequencing Plan: Steps toward the EPA Target Architecture.*

Like last year's Status Report, this document provides an update on EA activities during the past year and outlines EA activities for the coming year. It also includes the full current version of the EPA EA in electronic form. The accompanying CD-ROM contains an executable copy of the EPA Architecture Repository and Tool (ART), which is also available to staff throughout EPA via the Agency's intranet.

All capital planning and investment control (CPIC) proposals are documented in EPA's e-CPIC system, and information technology capital investment proposals, including all investments listed on Form 53, are mapped to the EA as shown in the ART model provided on this year's CD-ROM. EA and CPIC processes and teams are even more tightly aligned this year than last. The architecture staff has reviewed all CPIC proposals in depth for compliance with the Federal Enterprise Architecture (FEA) and has provided training and guidance on EA issues to the preparers.

* * *

On June 29, 2004, EPA achieved a "Green" rating from the Office of Management and Budget (OMB) in both status and progress for the Agency's implementation of the E-Government initiatives on the President's Management Agenda (PMA) Scorecard. Green is the highest level an agency can achieve and signifies that EPA has met all the standards for success in the E-Gov category, including EA. The rating reflects significant improvements in how the Agency manages grants, payroll, and travel, as well as continued success in meeting the operational requirements of capital planning and investment, earned value management, enterprise architecture, security, e-rulemaking, information access, and records management.

The EA Sequencing Plan, signed by the Chief Information Officer (CIO) in August 2004, divides EA design and implementation into three phases of maturity. This year the Agency entered the second of these phases.

In Phase I, which ended on September 30, 2003, EPA convened architecture design groups and appointed a Chief Architect. The Agency laid the foundations for its architecture by launching the National Environmental Information Exchange Network (NEIEN), building the Central Data Exchange (CDX, EPA's node on that network), and creating the data registries that form the core of current data integration work (the Facilities Registry System and the Substances Registry System). The EA Team, cross-Agency workgroups, and the leads of the EA architectural

"domains"[1] developed baseline and target architectures and deployed the program's architectural tool (ART). The Office of the Chief Financial Officer (OCFO) led the effort to revise the Agency's Strategic Plan, which led to the update of the baseline and target architectures and the beginning of the integration of EA with the Agency budget and capital planning processes. Finally, EPA created a Program Management Office (PMO) to oversee construction of the target Central Services infrastructure to begin the integration of EPA's mission-support information systems.[2]

Phase II started with the current fiscal year and is the subject of this Status Report. A major initiative within this phase is the upgrading and integration of EA governance and planning processes to make the architecture actionable. The Administrator has signed an Interim EA Policy supported by a framework for developing EA procedures. Working within that framework, the Chief Architect and the Enterprise Architecture Coodination Committee (EACC) are completing a detailed set of architecture procedures and standards that will govern all aspects of architecture design and execution. To complement this effort the Office of Environmental Information (OEI), OCFO, and the Office of Administration and Resources Management (OARM) are working together to integrate the five major EPA planning processes: strategic planning, budget planning, human capital planning, enterprise architecture, and the capital investment planning process. In addition, the first version of the target Research and Science Architecture (RSA) was completed, rounding out the architectures of the three major EA domains.

Phase II also involved construction of the key elements of the target EA. The focus of the PMO this year has been to build out and validate major elements of the Central Services. The PMO is building the first version of the EPA Portal with initial supporting identity management and access control. It is also constructing—for the Air Office's Air Quality Subsystem (AQS)—the first EHPA warehouse. By the end of calendar 2005, the new Central Services infrastructure will be complete, with at least AQS data flowing through all its parts. In addition to the construction project of the PMO, the Agency is linking several of its internal systems with government-wide systems to support federal e-Gov and Lines of Business initiatives.

Phase III will start at the conclusion of the Phase II build-out projects, scheduled for the end of calendar year 2004. This phase will complete the EA governance system and implement its core elements, launch a number of other initiatives that will evolve the architecture program toward the higher levels of maturity as measured by OMB, and begin migration of additional EPA systems to the Central Services.

[1] The EA architectural domains include the Administrative Systems Architecture (ASA), the Research and Science Architecture (RSA), and the Environment and Health Protection Architecture (EHPA).
[2] The Central Services infrastructure includes the new EPA Portal, with its supporting Identity Management and Access Control System, as well as a suite of business intelligence tools for accessing EPA's data warehousing environment. Data is received by the CDX from industry and EPA's national environmental Exchange Network partners. The Central Services also include the System of Registries, which controls the Agency's metadata architecture and provides central Data Registries of shared information, such as on regulated facilities. Central Services also support the Agency's data warehousing strategy.

1. Moving the Architecture Program Forward

1.1 EPA's EA governance system is extended

As the EPA EA program has matured, so too has its need for governance. In the early stages of the architecture, emphasis was on developing the main artifacts of the system: the baseline and target architectures. Phase II moved the architecture from analysis to action. The PMO started construction of the Central Services infrastructure and programs started integrating their systems with it. Programs began development of component architectures to expand the Agency EA along functional business lines. With so much activity along so many fronts, program offices, regions, and laboratories need to understand in detail what is expected of them. EA governance must expand accordingly.

Effective governance requires a full suite of documents and commitments from all levels of the Agency, from a high-level understanding of general goals, roles, and responsibilities, to detailed procedures and technical standards. The initial governance system described in the 2002 *EPA Target Enterprise Architecture* has therefore been elaborated in FY 2004, including the signing by the Administrator of EPA's Interim Enterprise Architecture Policy (September 2004) and creation of a framework for developing procedures for maintaining the Agency's core architecture products.

The EA policy structure, shown in Figure 1, shows EPA's overall model for EA governance. It is a four-tier sequence with policy at the apex, supported by operational procedures, technical standards, and guidelines.

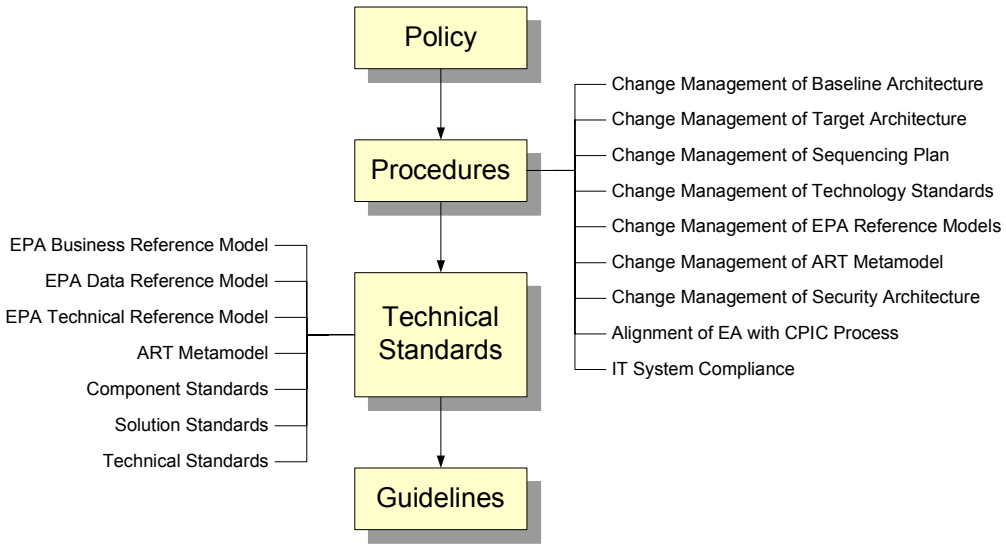

Figure 1: The Enterprise Architecture Policy Structure

Enhanced governance of the EA program will translate into performance efficiencies throughout the Agency. As the system life cycle and capital investment processes tie in more tightly with EA, information technology and information management systems will be subject to more detailed and timely performance review. Systematic, result-oriented portfolio management will become possible at all levels, not just for major investments.

1.1.1 The EA Interim Policy and Framework for EA Procedures Development provide general structure and direction to the EA program

The Interim EA Policy, provided in Attachment A, and the supporting Framework for EA Procedures Development, provided in Attachment B set the general requirements for Agency-wide utilization of EA. They position the Agency to create and maintain the baseline architecture, target architecture, and sequencing plan. They also link EA closely to the system development process by requiring all new or modified information technology investments to develop approved "solution architectures" that ensure adherence to the EA.

The policy has been submitted to the Directives Clearance Process for formal review throughout the Agency.

1.1.2 EA procedures will form a continuum with existing IT and program management processes

Procedures developed under the Framework for EA Procedures Development will set general requirements for Agency-wide utilization of EA. The EA procedures will define in detail how individual products (including the baseline and target architectures and the sequencing plan) are to be managed. Developed as a holistic set of related activities and processes, they will also specify the intersections between EA and existing information technology policies, such as system life cycle, security, and compliance. The procedures will include detail on annual schedules, individual and committee roles, and templates for reviews and approvals.

As the EA program matures further, EA procedures will be adjusted and expanded to reflect incremental process improvements and incorporate new functions and responsibilities. They will form a continuum with the Agency's major planning and management processes: strategic planning; budget planning, formulation, and execution; human capital planning; grants management; and contracts and acquisitions management.

1.1.3 EA tools are enhanced and will integrate with related systems

Last year EPA was the first agency to submit its architecture electronically through the ART tool. This year's electronic submission embodies a considerable number of improvements in the design and implementation of ART, with the EA Team continuing to improve the model's usability and interface. ART won an Innovator Award from Application Development Trends Magazine in April 2004.

ART is only one of a family of tools and systems used by the information technology planning and system management communities. EPA is therefore taking a system-wide approach as it makes further enhancements to ART and related tools, tailoring each tool to its user groups.

The first group—Information Technology/Information Management managers—will use the ART model in concert with other tools. These currently include e-CPIC, for management of Form 300s; the Registry of EPA Applications and Databases (READ); the Automated Security Self-Evaluation and Remediation Tracking (ASSERT) tool for annual Federal Information Security Management Act (FISMA) reporting; and the Federal Enterprise Architecture Management System (FEAMS) for cross-federal EA information. Over the coming year, the EA Team will

work with managers of the other information technology and information management-oriented models to discuss data interoperability and standards, as well as distribution of functional capabilities.

The second group—system developers—will begin to consult ART to ensure architectural compliance, but will also use the READ system for reference information on applications, the IT Roadmap for technology standards, the Environmental Data Registry (EDR) for data standards, the upcoming Developers' Portal, FEAMS, and Core.gov (the federal registry of system service components).

1.2 The architecture has been enhanced at multiple levels over the past year

The EA Team has enhanced EPA's EA tools and methods at multiple levels of the architecture pyramid over the course of FY 2004, addressing issues at the business, data, applications, and technology layers of the architecture.

1.2.1 Business Architecture

The EPA business architecture has been expanded to include a number of new architectural relationships documented in the Agency budget process. During FY 2004, the EA Team updated the EPA Business Reference Model (BRM) to fully reflect additional OMB A-11 guidance on the use of the BRM Mode of Delivery layer and to include new business functions. Additional changes were also made to the EPA Mode of Delivery layer to better reflect the needs of the target Research and Science Architecture. EPA is supporting the Department of Health and Human Services (HHS) in the development of the Federal Health Architecture (FHA) Line of Business Initiative to identify opportunities for business collaboration and information sharing on Public Health Surveillance and Food Safety Protection lines of business.

1.2.2 Data Architecture

During FY 2003, the EA Team developed a Strategic Information Model that describes the major data types necessary to implement the EPA mission as described in the 2003 – 2008 EPA Strategic Plan. In FY 2004, the Team further developed the Strategic Information Model, laying the groundwork for an integrated Agency data model that captures the major data entities necessary to implement the Framework of Business Warehouses within the Central Services. The relationships among these classes of data represent the high-level logical model of the EPA Strategic Plan developed in the Strategic Information Model. Work will continue to complete this high-level model and to integrate it with parallel efforts, such as XML schema development for the Exchange Network, the Environmental Indicators program, and the Core Reference Model.[3]

EPA's investment in the Core Reference Model and the XML exchange schema of the Exchange Network will contribute in major ways to implementation of the FEA Data Reference Model (DRM). Improvements in data quality and the efficiencies of data exchange are at the core of the Agency's mission, whether it be in delivering better public information to citizens, or protecting the confidentiality of business information for industry partners. Even though the DRM is in draft form, the EPA EA program is putting high priority on operationalizing its concepts in the coming year.

[3] The Core Reference Model is a collaborative effort to develop a high-level depiction of the major environmental data groupings and their relationships to support the National Environmental Information Exchange Network (NEIEN).

1.2.3 Applications Architecture

Refinement of the target applications architecture continued in FY 2004. Over the past several months, the technical design of the Central Services' Framework for Business Area Warehouses (FBW)—the operational core of the Central Services—has been more fully defined. It is here that data for analysis and decision-making will be integrated and made available for use through the Portal. The FBW now incorporates the concept of a Data Integration Platform (DIP)[4] as the mechanism for populating data warehouses and data marts, controlling the creation and flow of metadata, and ensuring flexible and responsive data availability. This bus architecture[5] will be simpler, less expensive, and more quickly implemented than the approach envisioned one year ago. The EPA target applications architecture is now even more efficiently oriented than before to support solutions-driven, component-based systems. The current proposed implementation of the target applications architecture is shown on Figure 2.

The EPA Portal is shown at the top of the diagram (green). It includes the single sign-on layer, identity management, and access control layers as before. Individual custom sub-portals are shown in darker green. Below the Portal layer, the diagram shows the flow of information across the applications architecture from left to right. At left is shown the exchange of data with industry and governmental partners through the Exchange Network to CDX. Individual systems (operational data stores) receive data from CDX, as do the data registries (contained within the System of Registries).[6] This allows for a level of interoperability that greatly improves the Agency's capacity for decision-making—a core goal of EPA's Strategic Plan.

To the right of the System of Registries is the new DIP (yellow), which provides the data integration bus for transferring data from source systems (including the Exchange Network) to warehouses and marts. Within this bus is the central metadata repository, which integrates metadata from the System of Registries with metadata generated in the process of extracting, transforming, and loading data from data sources into the Framework of Business Warehouses' various data warehouses and marts. Included within the DIP are data profiling and source analysis tools, administrative tools, and data cleansing tools.

To extract data from the data warehouse/data mart environment, the applications architecture relies on a suite of business intelligence tools shown at the right of the data mart layer. These include a range of commercial off-the-shelf (COTS) and custom products, from integrated business intelligence suites, to GIS tools, statistical packages, and customized reporting tools ("enterprise applications") built for specific program uses, but available at the enterprise level.

[4] A Data Integration Platform is a middleware strategy for accessing, integrating, and delivering data from any source, to any source. A DIP can utilize varying technologies

[5] Bus architecture refers to a structural configuration for interconnecting data warehouses, distinguished from the hub and spoke configuration customary in traditional central data warehousing environments.

[6] The registries include general enterprise information, such as the Environmental Information Management System (EIMS), which identifies all data sets available within the EA, and READ, which identifies all EPA applications. The registries also include registries that tie to specific program information, such as the Facilities Registry System (FRS), which identifies all sites and facilities of interests to EPA, and the Substances Registry System (SRS), which describes all substances of interest to the Agency—specific chemicals, waste streams, and biologicals.

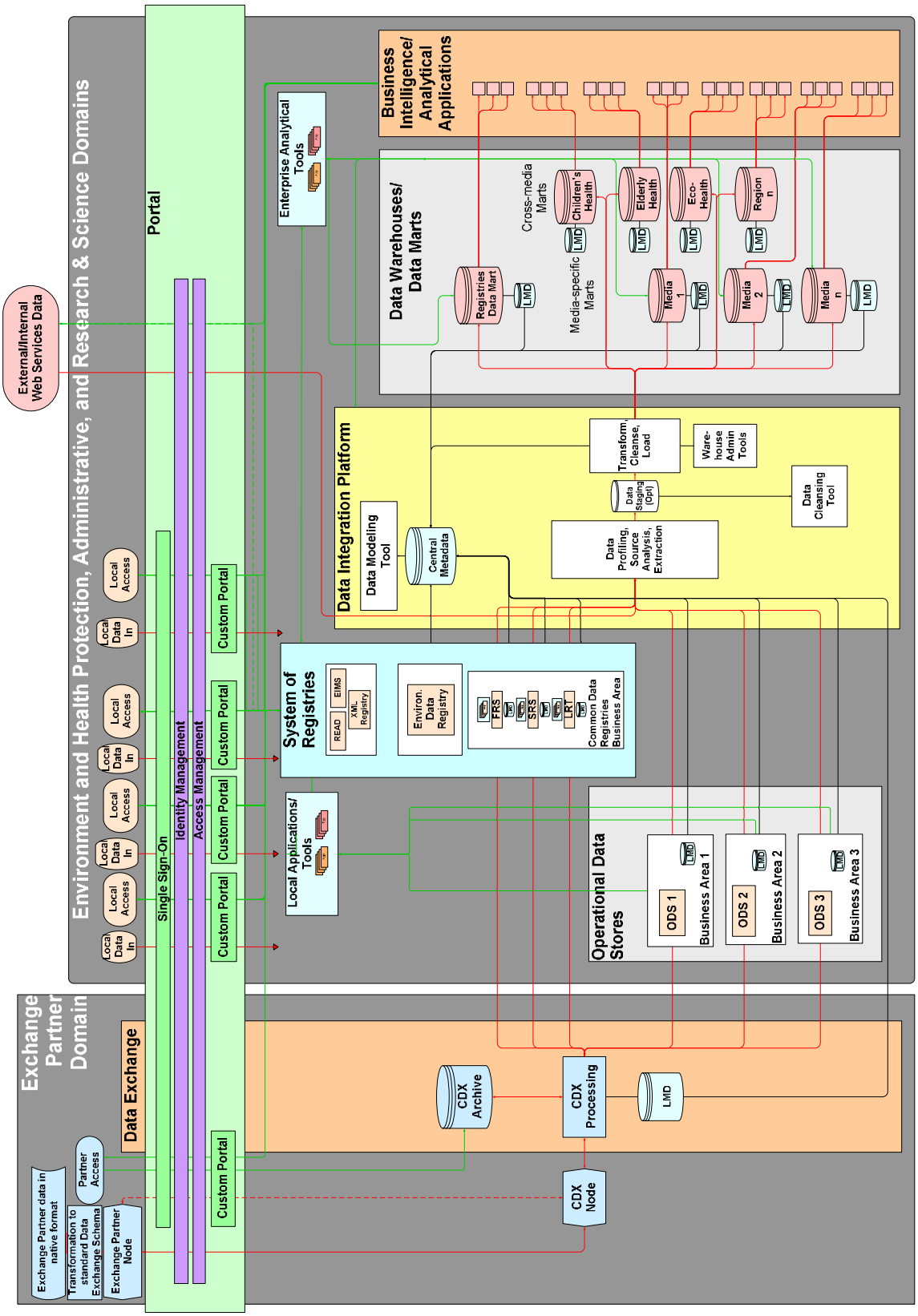

Figure 2: Applications Layer of EPA Target Architecture

An important aspect of the refined applications architecture is that it allows systems to migrate to the target one component at a time.[7]

1.2.4 Technology Architecture

In July 2004, the Agency named a Technology and Security Architect to work with the Chief Architect to coordinate development of these two vital sectors of the EA. The Technology and Security Architect will direct the Technology Architecture Work Group (TAWG) and work closely with the Technical Information Security Staff (TISS) to link the EA more closely with the ongoing security program.

Prior to the naming of this new position, the TAWG has focused on updating standards within the IT Roadmap—EPA's database of technology standards. The TAWG is also developing a process for updating the Technology Architecture Sequencing Plan (TASP) and defining the relationship between the TASP to the integrated Research Agenda and Standards process.

Development of Standards (the IT Roadmap and Standards Profile)

Between February and June 2004, the TAWG developed a list of technology types, reflecting the state of the art of industrial best practices, to be added to the IT Roadmap.[8] It proposed 26 such additions to the QIC Technology Subcommittee (QTS) in June, which it approved. Technology types are grouped into a limited set of categories, which are reflected by the TRM service areas. For 2004, most of the new technology types were in the Data and Technology Management service areas.

1.2.5 Security Architecture

In December 2003, the FEA PMO released the FEA Security Profile to federal government agencies for evaluation and comment. The Security Profile is a system that complements the FEA Reference Models and integrates security planning in alignment with them. All agencies will be required to align their security architectures to the FEA Reference Models by applying the concepts and procedures of the FEA Security Profile to their enterprise architectures after the Security Profile is released.

EPA is conducting a case study analysis to establish a repeatable methodology for applying the FEA Security Profile to the EPA EA. This work is being conducted on behalf of the CIO Council's Architecture and Infrastructure Committee (AIC); recommendations will be forwarded to the AIC for possible use across federal departments and agencies.

As part of this evaluation, EPA is now in the process of re-baselining its security architecture to accommodate the Security Profile and current applicable security standards published by the National Institute of Standards and Technology. OEI met with information security and technical experts representing Regions 4, 8, 9 and 10. Through these meetings, OEI is gathering key data on the regions' technology infrastructure, security controls and business practices used to protect the Agency's information technology resources. Regional information technology managers are also being asked to identify significant issues influencing their current operating environment and emerging technologies that require consideration in the Agency's proposed target architectures.

---

[7] Under the defined applications architecture, migration of systems can be incremental, allowing systems to move forward independently of each other, rather than in a series of "waves", in which systems would be queued in groups to migrate to the target, as originally proposed in the *Enterprise Architecture Status Report 2003*.

[8] http://basin.rtpnc.epa.gov/ntsd/ITRoadmap.nsf

1.3 Component architectures advanced the needs of business priorities

Work on a number of components of the Agency EA advanced over the past year. The most important addition was the completion of the target RSA (see Attachment C). In addition, the Administrative Systems Architecture (ASA) pursued several cross-cutting initiatives, the Office of Solid Waste and Emergency Response developed an Agency Cleanup Architecture, and the Office of Prevention, Pesticides and Toxics did valuable work on advancing methodologies for business-level architectures.

1.3.1 The target Research and Science Architecture serves the unique priorities of EPA's science community

The target RSA[9] is built around accommodating the technical needs of working scientists, as well as the access and dissemination needs of consumers and producers of scientific information, inside and outside the Agency. To manage the magnitude and complexity of the RSA, the target architecture framework was broken down into six focus areas that work in concert (Figure 3).

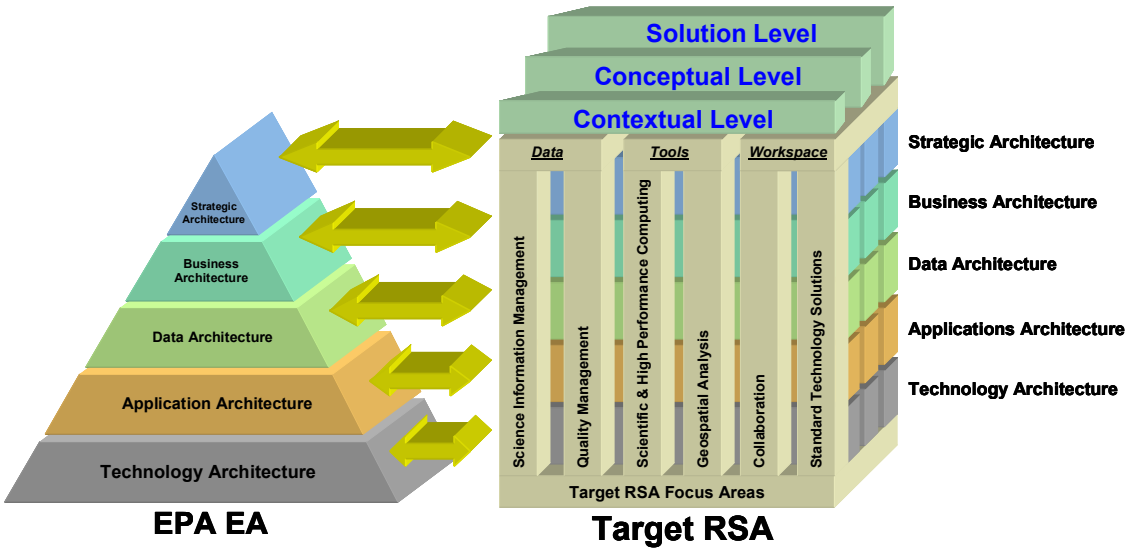

Figure 3: The RSA Framework

These are Science Information Management, Quality Management, Scientific and High-Performance Computing, Geospatial Analysis, Collaboration, and Standard Technology Solutions. These areas of interest are shown in detail on Table 1, which also summarizes each layer of the target RSA in relation to solutions that have been deployed or are in development to date.

[9] In September 2002, the *EPA Baseline Enterprise Architecture* included the baseline Research and Development Architecture (RDA), which was developed primarily by EPA's Office of Research and Development (ORD). In FY 2004, the RDA was expanded beyond ORD to include all science activities across the Agency. It was accordingly renamed the Research and Science Architecture (RSA).

BOLD ITEMS are Sample RSA Solutions To Date In Development or Deployed	**Data**		**Tools**		**Workspace**	
	Science Information Management	**Quality Management**	**Scientific & High Performance Computing**	**Geospatial Analysis**	**Collaboration**	**Standard Technology Solutions**
Strategic Focus	**Easy Access and Use of Scientific Data, Information, & Tools**	**Defensible Science through support of EPA Quality System**	**Maximum leverage of computational tools in support of EPA science**	**Easy Access and Enterprise Licensing of Geospatial Data & Tools**	**Maximal Collaboration Among EPA & With Trusted Partners**	**Best Technology ID and Use for EPA Science on a Uniform Basis**
RSA Progress Measures[20]	•Accessible Science Records •Metadata Integrity •Defined Data Admin Processes	•Quality Doc/Meta Data Pedigrees •QM Workflows Supported	•"In Silico" Science Support[4] •Access to Tools •Effective Use of Computing	•Available & Accessible Geo Data and Tools[18] •Metadata Integrity •Established Geo Governance	•Colleague & Partner Sharing •Data & Tools Availability Behind Firewall[17]	•Best Emerging Tech Applied •Apply Uniform Set of Standards
Business Architecture	•Enviro-Science •Enviro-Engineering •Monitoring Model Methods •Assessments	•QM Procedures Inside EPA •QM Procedures Outside EPA	•Develop Models & Applications •Provide Resources and Services	•Logistics •Engineering •Modeling •Monitoring •Assessments	•Metadata Analysis •Experiment •Analyze & Discuss •Publish	•Discover •Innovate •Standardize •Communicate •Coordinate
Data Architecture	•**EIMS** •**Geospatial Data Warehouses** •Science Info Repositories[1] •Derived Info[2]	•**EIMS** •Metadata Quality Pedigrees •Document Quality Pedigrees	•Data Wizards •Very Large Scientific Data Sets •Complex Schema	•Metadata (**EIMS\GDI**)[13] •Vector & Raster[12] •Georeferenced Data •Model Derived •Tool Derived (**ESRI**)[9]	•**Data Grid/EIMS** •Streaming Data •Government Data •Academic Data •Private Data	•IT Roadmap •Technologies under review •Managed Discussion DB
Applications Architecture	•**EIMS component applications[15]** •**Representative RSA systems[16]** •**IRIS** •**GIS**	•Doc, Metadata & Workflow Mgt. •Field & Lab Info Mgmt Systems	•Tool Wizards •**Simulation**[5] •Risk Analysis[6] •Statistical[7] •Info Mining[8]	•**GIS/Visualization**[9] •Geostatistical •Geo-enabled Models and Tools	•**Science Portal** •Workgroup Tools •Web Services •Application & Data Sharing	•**EIMS** •Moderated Discussion Group •Literature search tools
Infrastructure Technology Architecture	•**Data Grid** •**Documentum**[11] •**EIMS** •**ESRI – ArcIMS**[9] •LIMS	•**Documentum** •EPA Standard Technology •**Electronic Notebook**[3]	•**Sci Off of Future**[19] •**"NESC[2]"** •**Grid Computing** •**Hi-End Computing** •**Compute Clusters**[10]	•Imaging Technology •High-Speed Net •GPS Technology •HPC Technology •Wireless/mobile	•**Sci. FTP Server** •**Science Subnet** •Enhanced Subnet •High-Speed Net •AV Conferencing	•**RSA Web site**[14] •Standard Utilities •Specialized COTS (e.g., Notes app.)
Security Technology Architecture	•EPA ID Security Management •EPA Role-Based and Policy-Based Access Management •NTSD (EPA's National Technology Services Division) Models •Information Assurance with Releasability Data Tagging					

Table 1: Overview of Target Research and Science Architecture

TABLE FOOTNOTES:

1. These include Scientific Data Management Systems, Geographic Information Systems, Document Management Systems, Metadata Management Systems, Data Modeling Tools.
2. These Derived Information Collections include include Portals, Warehouses, Information from Partners.
3. This is specifically the Field & Lab Electronic Notebook technology.
4. This is a third leg of science combined with the two traditional ones of Theoretical and Experimental and is also often referred to as Computational. "In silico" Science is brought to bear in the research, science and decision-making activities of EPA where the problems are too large or too small to resolve by ordinary empirical means.
5. These include air quality forecasting, air dispersion models (CMAQ), and meteorological models (MM5).
6. These include TRIM and Total Human Exposure Risk database and Advanced Simulation Environment (THERdbASE).
7. These include commercial products such as SAS applications, ArcView, StatLIb, and RS/1, and EPA-developed tools such as applications for Dunnett's Procedure, Probit Analysis, Trimmed Spearman-Karber Method, and Linear Interpolation Method.
8. These include data mining, information pattern recognition, and knowledge management applications.
9. The predominant GIS applications at EPA rely on the ESRI suite of tools, including ArcView, ArcGIS, and ArcIMS. However, there is also a large collection of other visualization tools available for specialized purposes including Eiffel, AVS, Houdini, Shade Tree, Pixar Renderman, MAYA, and others.
10. These clustered computing platforms are mostly comprised of highly connected, rack-mounted systems that serve as small-scale high-end computing platforms.
11. This is one of EPA's standard applications intended to provide an agency-wide cross-cutting technical foundation for document management.
12. In general, Geospatial Data means vector or raster information that identifies, depicts, or describes the geographic locations, boundaries, or characteristics of inhabitants and natural or constructed features on the Earth. Please refer to the "Geospatial Analysis Focus Area" discussion for a more detailed definition.
13. Involved with Geo OneStop too.
14. Includes STS catalog.
15. EIMS related systems include:

    a. partners using the EIMS database to support a topical web presence, for example the Science Inventory,
    b. partners using EIMS to database information for organizational web-sites, for example NCEA,
    c. partners using EIMS for certain types of information aligned with their own database, for example HEDS,
    d. partners who do not have an individual web presence, but who use EIMS to store information in collections.
    A full list of EIMS partners can be found at www.epa.gov/eims
16. FY06 Exhibit 53 Systems: IRIS, EMAP, EMAP Surface Waters, ECOTOX, HEDS, Modeling (Health & Ecological Research), Modeling (Compliance & Stewardship Research), OMIS, EIMS, ORD Websites (Research Results), ORD Websites (Grants & Fellowships).
17. Establish firewall guidelines to support effective collaboration while managing risk.
18. Key Geospatial data sets have been identified and during FY2005, the science community will set priorities and write business cases.
19. Scientific Office of the Future will soon be known as Scientific Group Workspaces.
20. The RSA Progress Measures are the first step toward developing a complete set of RSA Performance Measures.

TABLE ACRONYMS:

ArcIMS – ESRI solution for delivering dynamic maps and GIS data and services via the Web.
CREM – Council on Regulatory Environmental Modeling knowledge base
ECOTOX – ECOTOXicology database
EIMS – Environmental Information Management System
EMAP – Environmental Monitoring and Assessment Program
ESRI – Environmental Systems Research Institute
GIS – Geographic Information Systems
HEDS – Human Exposure Database System

Scientific Information Management

Under the target RSA, scientific information will be managed efficiently, transparently, and holistically. The RSA community must cope with a geometrically-expanding proliferation of data, gathered and developed by collaborating scientists within the Agency and across the country. Different scientists and institutions have different working styles, needs, and approaches to managing data. But regulatory-based science demands transparent and reliable documentation and access. These different approaches must be integrated into a single architecture.

The target solution lies in the strategic use of modern Science Information Management (SIM) tools, some of which are already in use to a certain extent.[10] Effective integration of these tools with emerging new applications that improve the management of scientific data and geospatial information will lead to more consistent and thorough metadata and the establishment of more repeatable data management procedures across the research and science community. The RSA SIM initiative addresses the information management problem holistically, forming a foundation for achieving the other RSA objectives discussed below.

Quality Management

Under the target RSA, the quality of scientific data—the foundation of science—will be more efficiently controlled and documented. Improving data quality is not a new problem, but it is central to the RSA: regulatory-oriented science activities must be planned, managed, and documented appropriately to provide products of known quality. This issue relates to Scientific Information Management, but it focuses less on systems for data management than on explicit RSA support for the EPA Quality Program's policies, procedures, and standards.

Scientific and High Performance Computing

The target RSA will make high-performance computing available for many more uses than it is today. Many of EPA's concerns are amenable to so-called *in silico* approaches—those that rely on environmental simulations and computational modeling. Air dispersion analysis, climate change, waste load allocation in water, subterranean pollutant migration and many other problems cannot be adequately addressed other than through large-scale computer modeling. *In silico* science is especially adaptable to problems that are either too large or too small in scale to study on the bench. Furthermore, a host of analyses that used to be done on the bench can now be done on computers, often more appropriately and cheaply. The goal of the RSA is to make the most of costly high-end computing resources, efficiently allocating access to high-end machines and networks and assigning tasks appropriately to different classes of computing.

EPA will participate in such special purpose Grids as the Biomedical Information Research Network (BIRN), the United Nations Environmental Programme (UNEP), or the Access Grid Network (AGN). Under the RSA, as the power of individual platforms increases, and as computing Grid services and data Grid services become more sophisticated (including the increasing ability of scientific instrumentation of all types to talk to computers), the scope and practicality of *in silico* science will increase dramatically. The cost efficiencies of this approach are obvious; the practical impacts on day-to-day scientific operations will be transformative.

Geospatial Analysis

The RSA envisions the complete integration of geospatial information into the analysis and display of environmental information through the use of better tools and high speed networks. Although geospatial data, analysis, and visualization are a cross-cutting component of the entire

[10] The Environmental Information Management System (EIMS) is already being adapted to use within the Central Services. Documentum™ has been selected as the basis for E-Content Management.

EPA Enterprise Architecture, they are of particular concern to the RSA. Environmental research as a field is inherently oriented toward geospatial information, and as reliance on modern geospatial tools increases, existing *ad hoc* approaches must give way to more systematic development and quality control of data sets, and enterprise-wide rather than project-specific acquisition of data and tools. Ready and efficient access to this array of GIS tools and their associated data is being pulled together through Web service portlets in the Environmental Science Portal (under development), serviced by exceptionally high-speed networks (i.e., increased bandwidth) for rapid exchange of the extremely large data sets that GIS generates.

Collaboration

Under the target RSA, scientist across the nation will be able to collaborate more efficiently and conveniently through electronic means. Collaboration is the norm in science, but it is made more difficult by necessary security requirements and the large number of dispersed research and science projects that EPA conducts simultaneously. The RSA target addresses collaboration through a number of solutions, including the Environmental Science Portal, the Scientific Office of the Future, and the Science Subnet.

The Environmental Science Portal will be the scientific community's central point of access to environmental information and EPA's computing resources. It will be developed in collaboration with the PMO as a sub-element of the Agency's general Portal, becoming the central point of access to grid computing systems, data sets, and analytical tools. It will serve the entire RSA community, including researchers, scientists, laboratory staff, managers, support contractors, international partners and stakeholders, states, tribes, private industry, academia, and external contractors and consultants.

The Scientific Office of the Future is an initiative to upgrade the standard hardware and software configurations available to EPA scientists. These systems will support small-scale collaboration efforts at the desktop, mid-size collaboration in which multiple scientists gather to conduct team science with remote partners, and ultimately allow shared access to large-scale and high-end collaborative facilities containing the latest technologies for dynamic, multi-site team research and science activities.

The Science Subnet ties the Office of the Future to other systems. It will be a secure network environment that moves the Agency toward creation of virtual "Collaboratories". The Science Subnet will provide scientists with freer access to tools, data, and each other while protecting secure information systems at their current level.

Standard Technology Solutions

The target RSA will make it easier for scientists to get access to the technical tools they need. Developing a Standard Technology Solutions (STS) catalog will enable the RSA community to efficiently find technology solutions that satisfy their requirements, or to determine when new solution must be sought. It will also help to identify partial solutions that may be leveraged through modification or integration to avoid redundant investments or duplication of effort. This approach maximizes the potential for reuse and collaboration, and directly supports the goals of solution-oriented architecture.

Next steps for the RSA

The target RSA will be expanded through Joint Architecture Development (JAD) sessions that will complete the details of the above focus areas. Upcoming priorities include the completion of the RSA Business Reference Model (BRM) and initiation of an RSA Data Reference Model (DRM). More information on the target RSA can be found in Attachment C.

1.3.2 The baseline ASA is being expanded and the ASA Executive Steering Committee is leading several cross-cutting initiatives

The ASA Work Group has focused on several initiatives during FY 2004—expanding the ASA baseline architecture, aligning the Agency's strategic planning processes, analyzing and standardizing administrative data, and supporting cross-cutting federal lines of business.

Expanding the ASA Baseline

The first comprehensive version of the ASA was developed in 2002. The initial version of the ASA was developed in collaboration with partners across the Agency and focused on EA as it relates to headquarters administrative offices and lines of business. The baseline ASA is being expanded to incorporate regional and program office administrative functions, applications, and associated data. This complete, Agency-wide baseline will serve as the foundation for a revised target ASA.

Strategic Process Alignment

In FY 2004, OCFO, in partnership with OEI and OARM, launched the Strategic Process Alignment initiative, which is aimed at documenting the as-is state of EPA's strategic business processes to better understand their interconnections and identify opportunities for better alignment with each other and with the capital investment planning and Enterprise Architecture processes. These processes involved include strategic planning; budget planning, formulation, and execution; human capital planning and management; financial management; grants management; and contract and acquisitions management. This alignment will increase the utility of the enterprise architecture for use by the Agency's strategic and management planning processes.

ASA Data Initiative

In FY 2004, the ASA Work Group started a data analysis and standardization effort that will directly improve the operational effectiveness of the administrative organizations at EPA. This initiative focuses on two primary administrative data management areas: standardizing data associated with as-is systems that will be consolidated and modernized in the ASA target architecture, and the analysis and documentation of data associated with core administrative work products delivered to external customers. These two areas will directly contribute to near-term Agency goals, ensure compliance with OMB directives, and improve both data quality and customer satisfaction.

Status of the Financial Management, Grants Management, and Human Resources Lines of Business initiatives

EPA is involved in three cross-federal lines-of-business initiatives that are designed to increase the efficiency and effectiveness of service delivery and support the goals of a number of E-Gov initiatives, such as E-Payroll, E-Travel, E-Learning, and others. EPA is participating in line-of-business architecture development for financial management, human resource management, and grants management. EPA anticipates modifying its existing target architecture to accommodate any requirements arising out of the implementation of these three cross-agency projects. For instance, as the Agency has already aligned its payroll system to interface with Defense Finance and Accounting Service (DFAS) payroll processing and has connected its grant management system with Grants.gov to be the EPA intake vehicle.

1.3.3 The Environmental and Health Protection domain moved forward with business component architectures

Two EPA programs moved forward individually to develop architectures that answer specific business function needs. OSWER's Cleanup Architecture[11] focuses primarily on a component-based applications solution. The OPPTS process modeling initiative develops this office's business architecture across all levels of its internal programs, providing a basis for streamlining and coordinating business functions across disparate OPPTS offices.

Cleanup Architecture

As the agency with primary responsibility for environmental cleanup activities, EPA has led the effort to consolidate environmental remediation and cleanup business practices. In response to this business initiative, EPA's OSWER initiated a component architecture for the Clean Up Pollution function identified in the EPA BRM. In the EPA BRM, the Clean Up Pollution function includes Site and Area Cleanup. Although the concepts of this architecture can apply to other EPA offices and functions, the first iteration of the Cleanup Target Architecture (CTA) focuses only on OSWER activities involved in the Cleanup Pollution function and regions that implement and support OSWER programs.

The CTA outlines a vision for better integrating the information and systems supporting OSWER's cleanup function. The architecture is business-driven, based on EPA's strategic vision and priorities, and is fully compliant with, and integrated into, the broader EPA target architecture.

The CTA describes a component-based, service oriented, and distributed environment for applications supporting EPA's Clean Up Pollution function. Application functionality is encapsulated as a collection of distributed services that communicate with each other to fulfill end-user requests. Application services are built of modular and reusable software components that are distributed across the network.

The distributed applications environment of the CTA allows applications to use shared infrastructure components and services established in the EPA target architecture. It also allows EPA to create the necessary components and services unique to the Clean Up Pollution function. This represents a significant shift from EPA's current environment, in which applications supporting cleanup have been built in an application-centric fashion, with individual applications supporting each cleanup business process separately. Figure 4 presents the logical view of the CTA vision.

[11] This maps to the FEA BRM Environmental Remediation line of business.

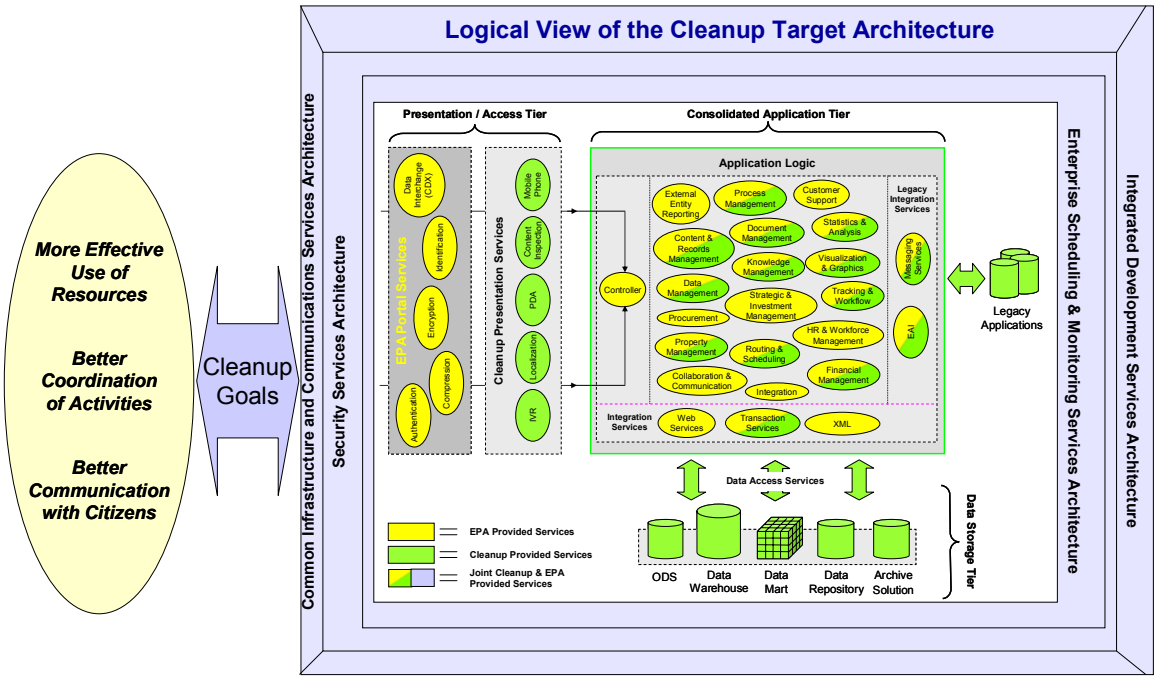

Figure 4: Logical View of the Cleanup Target Architecture

The CTA will improve the alignment of OSWER information technology investments with EPA's cleanup business needs.

The next step to realize the goals of the CTA is the Cleanup Target Implementation Plan (CTIP). The intention of the CTIP is to include both the analysis of baseline applications and the recommended schedule of activities required to consolidate and integrate the applications in support of consolidated cleanup goals.

The CTIP will define how the consolidation effort will migrate from the baseline to the target. Using the strategy, vision, and principles outlined by the FEA Service Reference Model, the CTIP will serve as a bridge between the current and future enterprise architectures. The SRM service domains provide a foundation, independent of business functions, "to support the reuse of applications, application capabilities, components, and business services"[12], and will serve as a guide in the development of the implementation plan.

Pesticides and Toxics Business Process Modeling

The goal of OPPTS's business process modeling is to document existing processes and discover opportunities for streamlining, collaboration, elimination of redundancies, and more efficient utilization of data and applications. This effort supports EPA's Manage Toxic Substances and Ensure Safe Use of Pesticides business functions.

Work across OPPTS to develop business process models began in 2002 with seed funding from OEI EA program. Under the initial effort, an analytic team studied one major business process area in each of the three suboffices of OPPTS. For the Office of Pollution Prevention and Toxics (OPPT), the team documented the business activities associated with almost 30 business processes within OPPT including the new chemicals program, the existing chemicals program

[12] FEA PMO https://www.feapmo.gov/feaSrm2.asp.

and efforts on national program chemicals. For the Office of Pesticide Programs (OPP), the team modeled the registration of new pesticides. Finally, for the Office of Science Coordination and Policy (OSCP), the team modeled business processes associated with the endocrine disruptor screening program.

Recent OPPTS work has focused on documenting the Federal Register Staff's baseline business architecture. Figure 5 shows how documents for OPPTS are prepared for publication in the Federal Register. The baseline is detailed. The color-coding in Figure 5, for instance, identifies the activities of individual staff members. OPPTS develops this level of detail to define individual workload and create the ability to perform future analyses of such issues as the risks of staff turnover, comparative human capital allocation, and business process reengineering.

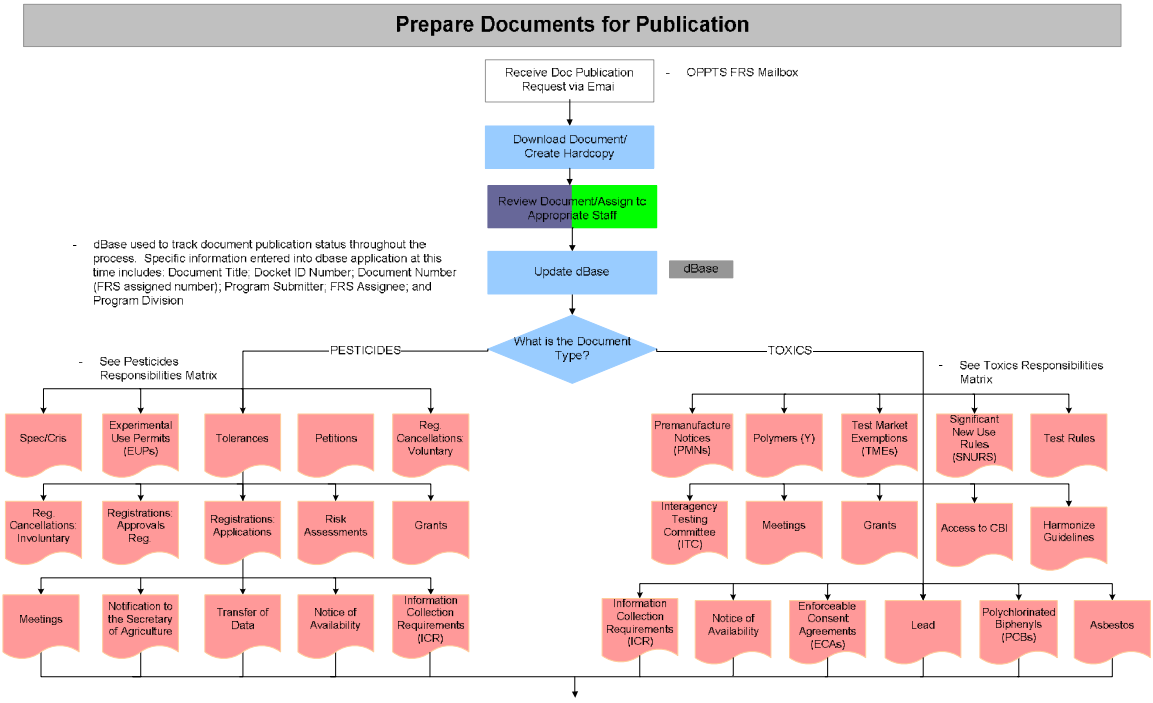

Figure 5: Example of Business Process Decomposition for the Office of Pesticides: "Prepare Documents for Publication"

The final step of this initiative is to transfer the detail developed in the detailed business process diagrams to the OPPTS implementation of ART.

The EA Team will use these efforts as a reference for creating Agency-wide Component Standards for business process modeling and, eventually, institutionalizing business process streamlining and optimizing methodologies. In addition, OPPTS's use of the ART model shows how the ART tool supports knowledge management: new employees will be able to use ART as a reference for how individual business processes work in detail.

In addition to the EA efforts conducted by OPPTS, OPPT has engaged in its own EA efforts, which focus primarily on documenting major lines of business and inventorying all applications and data holdings of the office. Analysis of the information collected has led to numerous business process improvement initiatives, including streamlining the way Toxic Substances Control Act (TSCA) information is collected from Industry. OPPT has developed a prototype

XML-based[13] reporting application that will be circulated on CD-ROM to corporations that report TSCA Inventory Update Rule (IUR) data. The data will arrive in XML format, consistent with Agency data standards and with the emerging OPPT data model. The goal of these initiatives is to document existing processes and discover opportunities for streamlining, collaboration, elimination of redundancies, and more efficient utilization of data and applications. The effort directly implements the EA principle that programs must revise and improve program processes prior to developing IT solutions to support them.

[13] XML is eXtensible Markup Language, a general standard for exchanging data over the Internet.

2. Building the Target Architecture

Construction of the target architecture began in a concerted way in FY 2004, both by OEI's PMO, as well as by programs elsewhere in the Agency acting in collaboration with the EA effort.[14]

Phase II build-out of the target architecture proceeded on two fronts: (1) upgrading the native technical capabilities of individual elements of the Central Services under the PMO's coordination, and (2) building out specific projects that validate how these elements will work together to support program needs (see Figure 6). Some of these projects were carried out jointly by the PMO and the Office of Air; others are underway independently by the Office of Research and Development to build out elements of the RSA, and by OCFO to build out sections of the ASA.

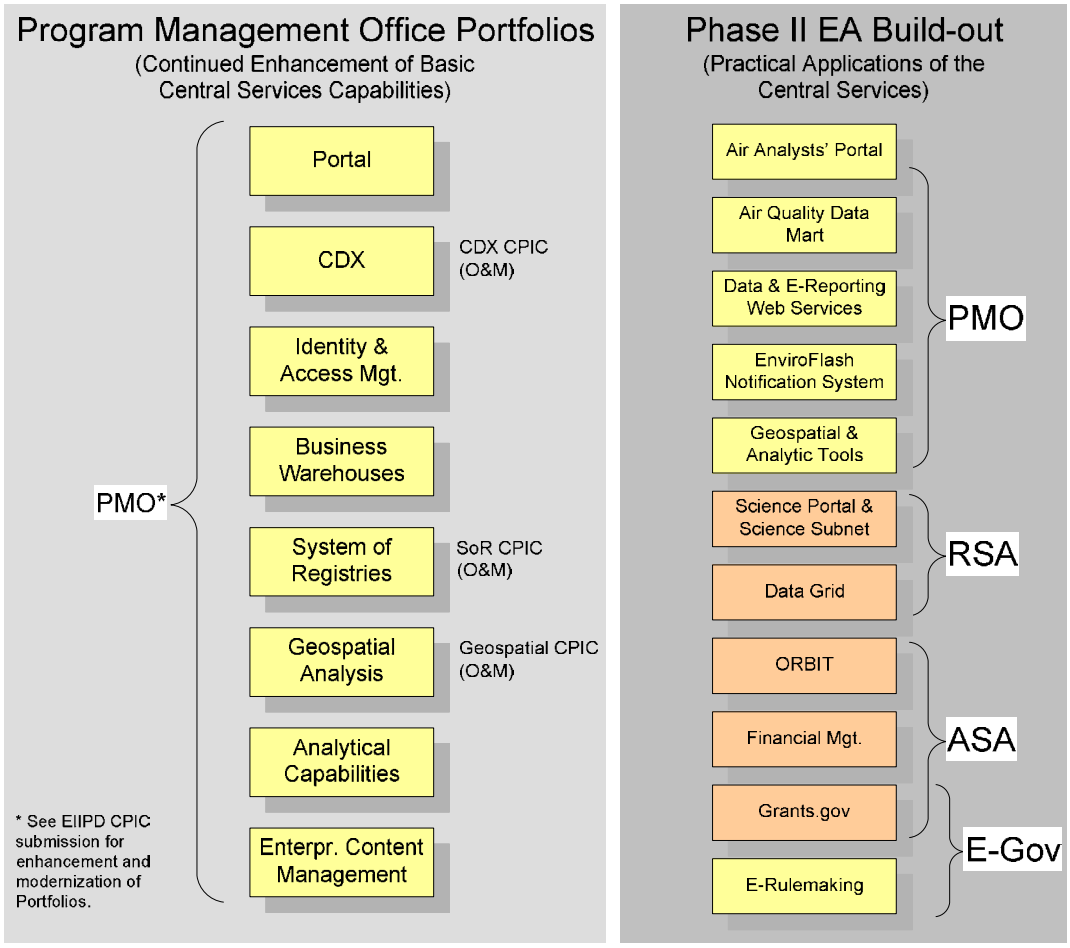

Figure 6: The PMO Central Services and FY 2004 Build-out Projects

[14] The Sequencing Plan summarizes the activities and programs that together have built EPA's EA over the past several years. Readers may refer to it for more information on the EA build-out prior to FY 2004.

The PMO coordinates eight individual "portfolios" of information management and information technology services: the Portal, the CDX, Identity and Access Management, the Framework for Business Area Warehouses, the System of Registries, Geospatial Analysis, Analytical Capabilities, and Enterprise Content Management. The operation and maintenance of the pre-existing PMO portfolios (CDX, the System of Registries, and Geospatial Analysis) are funded under their own CPIC proposals. New services and the development, enhancement, and modernization of CDX and Geospatial are funded by individual offices under the Environmental Information Integration and Portal Development (EIIPD) program CPIC.

2.1 Central Services technical capabilities are expanding

Work proceeded in FY 2004 to upgrade and enhance the technical capabilities of each of the eight PMO-coordinated portfolio services.

2.1.1 Portal Portfolio

In December 2003, EPA completed a operational capabilities demonstration (OCD) as a first step in developing the EPA Portal. The goals of the OCD were to identify potential portal user groups, define sample functionalities for various user groups, demonstrate the technical feasibility of a portal solution within the target EA, and obtain senior management approval of the portal project generally. Example mock-ups of portal functionality are included in Attachment E.

The PMO is currently evaluating three COTS packages for establishing the Agency-standard Portal Framework (the hardware and software infrastructure of the EPA Portal) and the various operating portals within the framework. The selected package will share back-end functionality among portals, enabling developers to deploy the same "portlet" applications from one to another. The Agency expects to select its preferred software package in September 2004.

2.1.2 Central Data Exchange Portfolio and Exchange Network Development

CDX is EPA's node on the Exchange Network, developed under the Information Management Work Group (IMWG). The achievements of CDX and the Exchange Network during FY 2004 included the addition of the five new production data flows shown on Table 2.

Phase II CDX Capabilities and Data Flows			
Production Data Flows	**Data Format**	**Source**	**Production Date**
Safe Drinking Water Information System (SDWIS)	Flat files	States	February 2004
Institutional Controls (ICTS)	Web forms	Regions	April 2004
National Emission Inventory	Web forms, flat files, XML files	States, Locals	April 2002 Node May 2004
RMP Web Reporting Center	Web forms	Industry	May 2004
Radionuclide National Emissions Standards for Hazardous Air Pollutant (NESHAPS)	Web forms, flat files	Industry	June 2004

Table 2: Enhancements to CDX Data Exchanges for FY 2004

The current nodes and exchanges supported by CDX are shown on the map in Figure 7. Node flows, the most highly automated form over exchange via CDX, are now in operation in 13 states. They include flows for the Facility Registry System (FRS), the National Emissions Inventory (NEI), e-Discharge Monitoring Reports (e-DMR), and Beaches. Nodes flows are in development for the Air Quality Subsystem (AQS), the Resource Conservation and Recovery Act Information System (RCRAInfo), and the Institutional Controls Tracking System (ICTS).

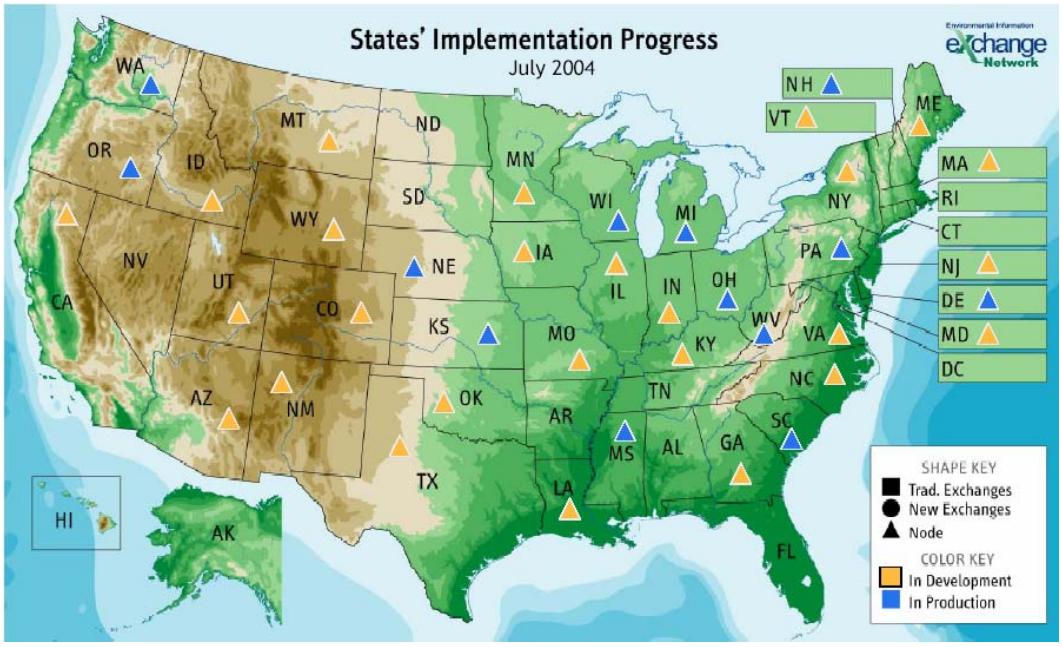

Figure 7: Operating Nodes of the Environmental Information Exchange Network, Current as of August 2004

At present, 49 states and nine tribes use CDX to file environmental reports across the major environmental "media" areas (air, water, waste, etc.). Fourteen data systems are in production, with ten others in design, development, or testing stages. Growth in the number of users and data flows has been significant over the past year: 8,000 new users were registered, bringing the total to over 30,000, and nearly twice as many files (26,000) were submitted to CDX in the first half of 2004 than in all of 2003 (14,000).

Growth in CDX usage translates into quantitative performance benefits for programs and users. For example:

- Submitters to the Office of Water's unregulated contaminant monitoring program have saved 30% of their data submission effort.
- The Toxic Release Inventory's processing time for CDX submissions is four minutes, as opposed to eight minutes for diskette and 35 minutes for paper.
- States filing Notices of Intent (eNOIs) for stormwater construction coverage have seen a 30 percent reduction in time required to fill out applications, and the error rate has dropped from 80 percent to virtually zero.

2.1.3 Identity and Access Management

EPA plans to implement an Enterprise Identity and Access Management infrastructure. The strategy for this infrastructure is based on an enterprise approach to directory services, which the

Central Services will leverage to provide user management and administration, authentication, and authorization services to a wide range of EPA applications. This Enterprise Identity and Access Management approach will result in improved security, significantly reduced user administration costs, simplified and accelerated application development, and enhanced user experience through self-service and reduced sign-on. The infrastructure will also be designed to support and integrate with the federal E-Authentication architecture.

2.1.4 Framework for Business Area Warehouses Portfolio

The PMO is now implementing the FBW through the use of a bus architecture that relies on a high-performance extraction, transformation, and load (ETL) tool to move data from sources (the Exchange Network and various program office data management systems) to data warehouses and data marts. This will enable the FBW to operate as the operational core of the Central Services, making Agency data available for analysis and decision-making through the various EPA Portals. EPA is in the process of conducting comparative evaluations to select an Agency standard. Once selected, EPA will develop the initial data mart, which is scheduled for completion within the first quarter of FY 2005. The ETL tool will be the center of the proposed DIP, which will cleanse, profile, extract, transform, and load data into the various data warehouses and marts of the FBW. It will also manage a central metadata repository that will tie together all warehouses and marts into an integrated EPA database environment.

2.1.5 System of Registries Portfolio

The System of Registries (SoR) portfolio includes two types of registries: Metadata Registries that manage data flows throughout the Agency, and Data Registries that provide uniform reference data, such as names of facilities and chemical substances, to individual program systems. Implementation of both sets of registries has advanced significantly in FY 2004.

The Metadata Registries include the Environmental Data Registry (EDR), which stores data element standards and definitions for systems across the Agency; the XML Registry, which maintains XML schema and XML schema components for use on the Exchange Network; the Environmental Information Management System (EIMS), which is evolving as the general Agency data set registry; and a yet-to-be designed data model registry to store individual data models of operating systems.

Substantial progress has been made this year on implementing the XML registry, which is now into its second year of development. The XML registry now contains 20 XML schema in various stages of working development. Individual data schema contained in the XML Registry are consistent with the Core Reference Model (CRM) developed under the direction of the Network Steering Board (NSB). The CRM (now at Version 1.20) is under continuing development; EPA is playing a leading role in expanding the CRM to become the definitive data standard for environmental information exchange. Schema contained in the XML registry are being used by program offices as the reference standard for modernizing their operational data flows.

The EDR is also being expanded. It operates as the system of record for the Environmental Data Standards Committee (EDSC) and is the repository of EPA data element definitions. As the Central Services mature, data definitions within the EDR will be integrated and link directly to the CRM-driven XML Registry through the ETL tool and its central metadata repository.

2.1.6 Geospatial Analysis Portfolio

Significant progress was made in FY 2004 toward implementing the Agency-wide approach outlined in the EPA Geospatial Blueprint (June 2003). In March 2004, the Agency hired a Geospatial Information Officer (GIO), who is overseeing the implementation of the Geospatial

Program and the integration of geospatial components into the EPA Enterprise Architecture. Geospatial data and technology were highlighted in the target RSA, with the RSA approach fully integrated with that taken in the EHPA component of the EPA EA.

In FY 2004 the Agency also completed an update to the Locational Data Policy and the Locational Improvement Strategy. The updated policy will result in consistent approaches to, and documentation of, geocoding environmental data collected/generated during EPA funded business operations. Funds allocated to implement the Locational Improvement Strategy enable new geocoding efforts and provide enterprise location siting tools. Both will result in more accurate location attributes for facilities and other areas of interest and facilitate the integration of location-based approaches into everyday business. The Agency is in the process of upgrading its locational reference tables to accommodate polygons and lines, which is critical for the exchange of data between EPA and its partners on the Environmental Network exchange and for use with analytical tools. Work has also begun on geo-enabling the Central Data Exchange. The primary objective is to develop the schema and Web services necessary to exchange geospatial metadata and data among partners. Work is also underway with the PMO to incorporate geospatial analytical tools into the first version of the Portal, slated for completion by the end of 2004.

In FY 2004, the Agency also participated in the government-wide SmartBUY initiative to enter into an enterprise license agreement (ELA) with the Environmental Systems Research Institute, Inc. (ESRI) for GIS software, data services, and training. This ELA benefits the organization by centralizing purchasing and consolidating asset management. It will save the Agency hundreds of thousands of dollars over three years by substantially reducing license fees, freeing FTE for additional work, and providing users with better terms and conditions. The ESRI ELA makes GIS software, geospatial data services, and training available Agency-wide to programs. The ESRI ELA will help incorporate location-based approaches into EPA business by making a broader array of products available. It will also expedite the completion of the geospatial metadata architecture by making geospatial metadata management software available enterprise-wide which will expedite consistent Agency-wide geospatial metadata management.

The Agency also procured several other enterprise data sets that ensure Agency-wide consistency, increase quality of the data used, improve the products developed as well as the ability to share them Agency-wide. These data sets have the same cost saving benefits as the ESRI ELA. Beginning in FY 2005 these enterprise licenses assets will be managed centrally through the EPA Working Capital Fund as a new Geospatial Tools and Data Service. This service stabilizes planning and budgeting, enhances tracking of service use and service requirements, ensures standardized technical configurations of GIS across EPA, and reduces administrative costs by eliminating the need for separate funding processes.

In FY 2004, the Working Capital Fund Board approved a capital investment for two servers dedicated to Geospatial data, metadata and applications. This will enable the Agency to move forward toward completing its geospatial network, which will be an integrating component of both the EHPA and RSA. The Geospatial Network will be critical for linking and making accessible data in Agency-wide geospatial business area warehouses/data mart and implementing the geospatial metadata, data, and analytical services envisioned in the Geospatial Blueprint.

2.1.7 Analytic Capabilities Portfolio

The Analytic Capabilities portfolio has added the ability to display the locational information through the use of GIS tools. Analytic capabilities are also enhanced through the use of Business Objects™ to access budget information generated by OCFO. The Agency has procured enterprise

licenses of both Business Objects, dynamic choropleth mapping capabilities, and SAS™ that will enable portals to provide their analytical capabilities to users and partners.

2.1.8 Enterprise Content Management

As OEI and Agency staff began planning for a much needed electronic records and document management system, the concept of "enterprise content management" (ECM)—recognizing content as an organizational asset and managing it across the organization, irrespective of where it exists—began to emerge.[15] After reassessing the Agency's needs and requirements, it became clear that it was necessary to move beyond only document and records management: the entire Agency's electronic content in its various forms needs to be managed. Therefore, the initiative formerly known as the Electronic Records and Document Management System (ERDMS) is now the broader Enterprise Content Management System (ECMS) initiative.

After an extensive selection process, Documentum™ was identified as the Agency standard software for electronic document and records management. Given the multiple benefits derived from managing information on an enterprise level (reuse of information in multiple ways, multi-media approach to environmental protection, creation of an enterprise knowledge tool), OEI and other EPA offices interested in implementing a content management system are combining resources to launch ECMS. Initially, this will involve the collaborative efforts of OEI, Region 6, and OSWER.

Initial ECMS deployment, anticipated to be completed by December 2005, will represent the first time that the full ECMS system is deployed in a real business environment. The new advisory board will provide the governance for the enterprise implementation of ECMS. Region 6, OSWER, and OEI will be the initial partners in building out Phase I of ECMS. With one of the largest collections of Superfund data, Region 6 makes an ideal Phase I site because of its experience in managing large databases. Work will include creating a paperless office in Region 6 that can scale to the enterprise as a whole; integration of the Agency Superfund Document Management System; and identifying all security, privacy and records/document management issues necessary for enterprise-level implementation. Phase I will include building ECMS's hardware and software infrastructure as well as designing the repeatable processes, procedures, questionnaires, and templates necessary for use throughout the enterprise.

2.2 Central Services build-out projects provide practical products and services

The PMO's main task is to coordinate the design and building of the Central Services infrastructure. FY 2004 PMO activities have focused on implementing the EPA Portal and on validating and deploying the revised data warehousing strategy through construction of the DIP and an initial data mart. The PMO has worked in close collaboration with the Office of Air and Radiation (OAR) to prove out the components of the Central Services through a series of operational projects.

[15] At their most basic level, ECM tools and strategies allow for the management of an organization's unstructured information (images, documents, graphics, drawings, web pages, e-mail, etc.) wherever that information resides. ECM is about making information available across organizational boundaries, that is, placing information where it is most easily accessed by Agency employees. Document and records management applications fall under the umbrella of ECM and are accompanied by correspondence management, Web content management, knowledge management, customer or citizen relationship management, and learning content management.

2.2.1 Program Management Office Projects

The PMO is leading development of five related projects, each of which builds a portion of the Central Services infrastructure or demonstrates a technical capability enabled by the Central Services. These projects include the first iteration of the EPA Portal, an example data mart within the Framework for Business Area Warehouses, application of Web Services to improve the way in which air quality data is reported over the Exchange Network, the Geospatial Analytical Tools project, and the EnviroFlash Notification System.

The Air Analysts' Portal

The Agency's portal framework software will allow for multiple portals within a single server installation, with each portal sharing system-enabled "portlets" that support user-customized home pages for every registered portal user. It provides a "one-stop" point of access for exchanging data, integrates access to all data sources and applications, and provides dynamic and static reporting capabilities. Its interface is intuitive and easy to use, with underlying technology that meets E-Gov and eSignature requirements. The system provides secure access for the entire EPA community to conduct its business more efficiently and securely.

The Air Analysts' Portal will access data within the Air Quality Subsystem (AQS) Air Quality Data Mart (see below) and be the basis for a more general, program-oriented Air Portal. The initial version of the Air Analysts' Portal will be completed in the first quarter of FY 2005. It is a joint task between OEI and OAR, but will be the prototype portal development project for all other EPA portals, combining OAR-specific portlets with generic Agency portlets.

AQS Air Quality Data Mart

It is critically important to the success of EPA's data warehousing strategy to demonstrate the viability of data mart technology based on the use of the central DIP. This project will deploy an initial data mart populated by the DIP. The initial data mart will contain raw data from the AQS system, made conveniently available for the first time to analysts with access through the EPA Portal.

The AQS Data Mart is designed to meet the needs of a mature environmental program such as the air program, which manages large amounts of historical data. AQS was an appropriate system for the PMO to use as an initial project because air office business processes are well established and understood. Although the data volumes involved are large, the business requirements of the data mart were readily identifiable.

Air Quality e-Reporting and Data Submission Web Services

One of the ancillary issues of concern with the AQS system is users' difficulty in submitting ambient air quality data to AQS and obtaining basic reports.[16] Submission error rates are approximately 20 percent, and procedures for obtaining basic reports are cumbersome.

This project is deploying a Web services-based quality assurance (QA) system that will automatically test incoming data flows against reference data such as chemical and facility names. If submissions are referencing incorrect data, the submitting state will be notified by the system. This approach implements the concept of moving data QA as close to the source as possible, a basic priority of data stewardship. Previously, data errors did not surface until

[16] AQS manages data from ambient air quality monitoring stations across the nation.

submissions were processed at AQS. This project is also developing similar QA screens for data flows into the National Emission Inventory (NEI).[17]

In addition to improving the QA of incoming data streams, the project is also applying Web services to improving the AQS reporting process. A series of standard reports will be available to allow end users to extract real time data from the AQS system.

Enviroflash Notification System

Finding new ways to leverage EPA data is a priority of mutual priority of the EA and the President's Management Agenda E-Gov initiatives. This system expands the proposed use of CDX and other Central Services into capabilities that will allow it to actively "push" data out to clients—in this case private citizens.

The system allows private citizens to register one or more email addresses and receive air quality forecasts and notifications of declared air action days. Users will go to their home air board's Web site or the EPA AIRNow Web site to sign up for the service, which will provide information customized to their locality. Local administrators will be able to customize the look and feel of the e-mailed alerts, as well as the information content, including messages for what to do in the case of an air action day. EPA will contributed generic content that the localities are free to use if desired. This could include metadata from the Systems of Registries to explain ozone or particulate data.

EnviroFlash was rolled out in August 2004 for the cities of Cincinnati, Nashville, St. Louis, Detroit, Winston-Salem, and Phoenix. The concept can now be nationalized as more localities join the system. In the future, other alerts can be added, such as notifications of beach closings or drinking water quality violations. The EnviroFlash Notification system is being built as a free-standing tool that will be incorporated into the suite of available Portal applications.

Geospatial and Analytic Tools Project

Because of the existing configuration of OAR's data, such as that contained in AQS, AQS data is not linked to enterprise geospatial data assets and cannot be used by geospatial analytic tools. OAR analysts must gather disparate geospatial data and aggregate on their own to conduct geospatial analyses. In addition, OAR geospatial data (such as locations of emission points) cannot be easily accessed for use by the rest of the Agency. This project will address these problems through development of Web services-based connections between OAR assets and geospatial tools.

2.2.2 Research and Science Projects

The Environmental Science Portal and Subnet

The Environmental Science Portal will address the needs of distributed researchers and decision-makers seeking a single point of access to scientific data and computing resources within a secure environment. Portal users will find a wide variety of modeling tools, applications, data sources, and computing resources targeted to their special needs. The Environmental Science Portal will be an integrated part of the Enterprise Portal, providing this Agency stakeholder group with the capacity and collaboration tools and data specific to meeting EPA's research and science mission. Users of the Portal will enjoy the ability to customize and personalize their desktops with the scientific tools, applications, and data they routinely use without multiple log-ins, thus adding to the efficiency of the research process. The Science Subnet also provides a segmented part of the

[17] NEI manages data on emissions from all types of air pollution sources.

Environmental Science Portal for EPA's external collaborators to quickly and easily exchange large volumes of data, and to collaborate using tools that will effectively erase the physical miles between research centers. Further enabling collaboration will be a planned suite of tools that will allow collaborative communications, data visualization, evaluation, and decision making.

The Data Grid

EPA scientists have a continuing need for expanded data grid capabilities to support the processing of very large sets of data for modeling activities. A pilot group of projects under way this year involves leading scientists in the field of air quality modeling. This group is interested in expanding the use of air quality simulation and prediction at the state and regional level. Supported by grid computing and data grids, these state and regional partnerships will collaboratively develop, optimize, and enhance a series of air quality models. Access to computational capabilities and datasets needed to run state-of-the-science models will allow manipulations and analyses never before imagined.

State-of-the-science air quality tools and approaches will be transferred to the states and applied to solve real problems, such as enhancing states' ability to evaluate control strategies, expanding state-level forecasting capability to include more pollutants, enhancing states' ability to explore links between air quality and human health, and improving states' ability to assess the impact of regulatory and policy decisions.

2.2.3 Administrative Systems Architecture Projects

Further Integration of Financial and Performance Information

In December 2003, OCFO launched a business intelligence tool called the OCFO Reporting and Business Intelligence Tool (ORBIT). This application integrates EPA financial, administrative, and program performance information. It provides enterprise-wide, Web-based access to this information and assists Agency managers in making informed decisions about their programs and operations. It is based on the COTS software package BusinessObjects™.

ORBIT provides access to Agency information at all levels. The tool includes an executive dashboard that uses charts and graphs to provide senior level managers with big-picture information on finance and funds control management, the Agency operating plan, and payroll and full time equivalents (FTEs). Standard reports provide users with detailed financial, budget, grants (financial), and human resources information. The application's *ad hoc* capability allows users to customize reports with the data they need to answer specific questions.

Grants Management

EPA has developed an interface with Grants.gov/APPLY. System-to-system testing with Grants.gov is complete and EPA's Web services are being configured to receive electronic applications from Grants.gov through CDX. (EPA is the first agency to bind to the Grants.gov "Web service definition language".) Internally, EPA continues to modify the Integrated Grants Management Systems (IGMS) to accept and act on applications arriving through Grants.gov. Final requirements are in place to collect electronic applications and manage them prior to migration to IGMS for award preparation.

EPA has submitted a ramp-up schedule to Grants.gov/APPLY and anticipates six programs available for application through Grants.gov during FY 2005. One program was available during FY 2004 and successfully collected 30 percent of the total applications electronically through Grants.gov.

Financial Management

EPA will implement a fully integrated, COTS-based payroll/HR system (PeoplePlus™) in October 2004. PeoplePlus supports the full range of payroll and HR functions, including payroll processing, time and attendance, and labor distribution, without redundant data entry or record keeping. It will replace legacy systems, some over 30 years old, and automate several manual processes.

The implementation of PeoplePlus is a significant achievement. PeoplePlus represents a major improvement in efficiency: elimination of paper-based data collection, approvals, and clerical re-entry of data. It provides significant improvements in functionality, primarily in the areas of management reporting and control. The system also integrates seamlessly with HR-Pro, the system of record for all EPA personnel data. In October 2005, the Defense Finance and Accounting Service (DFAS) will handle EPA Payroll Processing while EPA will still be responsible for Time and Attendance and Labor Distribution.

2.2.4 E-Gov initiatives

e-Rulemaking and EDOCKET

EPA is the designated federal lead for eRulemaking, the cross-government initiative to establish a federal regulatory clearinghouse Web site and a centralized electronic rulemaking system based on EPA Dockets (EDOCKET).

The eRulemaking initiative will include the Regulations.gov Web site, which will enable the public to find and view any open federal rulemaking and submit comments electronically for those rules. EDOCKET is an online public docket and comment system based on EPA's ECMS. EDOCKET will expand access to documents in EPA's major dockets. Dockets contain Federal Register notices, support documents, and public comments to support regulations the Agency publishes, as well as various non-regulatory activities. Work to develop and expand EDOCKET began in FY 2004.

Additional E-Gov projects

EPA is actively involved in eight of the 23 federal E-Gov initiatives. In addition to the projects mentioned above, they include:

- **E-Training:** EPA will procure online training services using GoLearn from the Office of Personnel Management (OPM). Arrangements were made in August 2004.
- **Federal Asset Sales (FAS):** EPA is in negotiation with the General Services Administration (GSA) to use FAS to manage a portion of EPA's surplus personnel assets.

3. The Future

3.1 EPA's EA will continue to mature

In June 2004, EPA achieved a "Green" rating from OMB in both status and progress for the Agency's implementation of the E-Government initiatives on the PMA Scorecard. Green is the highest level an agency can achieve and signifies that EPA has met all the standards for success in the E-Gov category. It reflects significant changes in how the Agency manages grants, payroll, and travel, as well as continued success in meeting the operational requirements of capital planning and investment, earned value management, enterprise architecture, security, information access, and records management.

EPA's PMA Scorecard Green rating is supported by progress in maturing the Agency's enterprise architecture. Over the past several years, both the Government Accountability Office (GAO) and OMB have compiled government-wide ratings of EA maturity. GAO scores the maturity of management practices; OMB's approach is broader, evaluating both EA management and EA implementation (see Attachment F).

3.1.1 The GAO Enterprise Architecture Management Maturity Framework (EAMMF)

GAO first rated agencies in 2001, when EPA's CIO assessed EA maturity at Stage 3 out of a possible 5 on Version 1.0 of the GAO scoring criteria. After modifying its criteria (Version 1.1), GAO lowered EPA's maturity rating to Stage 1 in 2003. With the issuance of its Interim EA Policy and procedures and development of the Sequencing Plan (see Attachment D), the CIO has re-evaluated the Agency's EA maturity and has determined that the program has improved its rating to Stage 4 on the older scale and Stage 3 on the newer scale.

In FY 2005, EPA will approve detailed procedures for EA maintenance (formal management of EA change), create metrics for measuring EA benefits, and institute EA compliance measurement and reporting, thus achieving Stage 4 under Version 1.1.

3.1.2 The OMB Enterprise Architecture Assessment Framework (EAAF)

In June 2004, OMB issued the EAAF Version 1.0. Using this set of criteria—quite different from GAO's—OMB scored EPA at 3.13. Over the coming year, EPA will institute improvements to its program to raise its score to 4.0 on the OMB EAAF scale.

3.2 Priorities for the coming year

The general priority for the coming year is to continue to make the architecture an effective management tool at all levels, from up-front capital investment planning, to design and implementation of new systems, to maintenance and upgrading of existing systems.

3.2.1 Governance

The Interim EA Policy is now signed and in effect. The coming year will see formal review and approval of the final policy, the complete set of supporting procedures, changes to relevant organizational charters, and improvement of the operational implementation of the governance.

The procedures now under development will cover all aspects of the EA program. The process of making them operational, however, will be phased in over the next two years. In FY 2005, the Agency will focus on a limited number of procedures—those that have the most impact on day-to-day operations and that are most closely linked to the capital investment planning process.

Continue and expand the Strategic Process Alignment initiative

The Strategic Process Alignment initiative, begun in FY 2004, will continue and expand in the coming year. This year the Agency has aligned the EPA EA program in support of the EPA budget process, helping to improve resource allocation decisions by further unifying the capital investment planning process, EA, and the budget program. Next year the alignment will be extended to include strategic planning, grants management, contracts and acquisitions management, and human capital planning. Work will begin to align these processes with the EA program.

Expand Solution Standards

In FY 2004, EPA began work to create the integrated Solution Standards—the formal specification of the target EA in operational terms. With the conclusion of the initial round of Central Services implementation projects, the necessary information will be available to finalize the first version of the Solution Standards. The PMO and the Chief Architect will be able to more fully define the functionality, and (equally important) boundaries to the functionality, of the main Central Services portfolios—the Portal, CDX, System of Registries, and the Framework for Business Area Warehouses (the DIP plus data warehouses and data marts).

Deliver Component Standards

Next year will also see the development of Component Standards, essential technical standards within the EA governance framework. Component Standards define the standards used by individual component architectures within the "one architecture" of the EA enterprise.[18] They define the artifacts necessary to build component architectures, the necessary relationships among those artifacts, the modeling procedures necessary to document component architectures, the change management procedures required to update component architectures within the enterprise architecture, and other relevant information.

3.2.2 Architectural Development

In addition to advancing the governance of the EPA EA program, the architecture itself will be expanded and improved in FY 2005 on several fronts.

Evolve EA methodology toward model-driven, business pattern-oriented architecture

EPA's EA methodology will be expanded toward the goal of Model Driven Architecture (MDA), in which enterprise architecture solution development and actual system design and implementation become integrated.

Transition to a business pattern- and solution- oriented MDA will only be started in FY 2005, but it will help approach and support a number of goals, including integrating the use of performance

[18] As based on the FEA Service Component definition.

measures into the evaluation of the EA target architecture as well as the EA program and its products. It will help create a direct link between the modeling of business processes and the applications and technology of the target architecture, and increase the Agency's ability to interoperate with other federal components.

Update the Baseline Architecture

One of the first formal EA procedures to be implemented will be change management of the Baseline Architecture. All layers of the architecture—strategic, business, data, applications, and technology—will be updated in FY 2005. Particular emphasis will be placed on updating the technology layer, which has not been extensively updated since September 2002. This update of the baseline will extend to the regions and to the creation of new component architectures, such as an Emergency Management architecture to complement the Cleanup Target Architecture initiated this year by OSWER.

Re-baseline Security Architecture to the FEA Security Profile

The EPA Security Architecture will be re-baselined in FY 2005 through the use of the FEA Security Profile. For the first time, the Security Profile integrates the FEA reference models (BRM, DRM, SRM, and TRM) with the security architecture.

Update the target Business Architecture

The FEA Performance Reference Model (PRM) provides a model for linking technologies and applications to customer service and mission goals. OMB requires use of the PRM only for CPIC proposals for investments undergoing development, modernization, or enhancement (DME). The PRM's approach, however, is generic and can be used generally to link information technology and information management investments to business processes based on quantifiable performance objectives.

Work to enhance the EPA business architecture through the use of performance measures that link to the applications and technology layers of the architecture will begin in FY 2005. This is an essential step to enhance the maturity of the EA program as measured by OMB's EAAF.

Update the target Data Architecture: move toward an information assurance architecture

One of the EA efforts for FY 2005 will be advancing the data layer of the EA toward an information assurance architecture that provides systematic, multilevel protection of mission-critical information. Work this year to create an initial entity relationship diagram (ERD) from the FY 2003 Strategic Information Model will be expanded. The data classes and entities of the ERD must be validated against the data exchange standards that govern the Exchange Network (the XML schema and the Core Reference Model) and expanded to include the data requirements of the Environmental Indicators program. The integration of these three data areas—the EA ERD, the data exchange standards of the Exchange Network, and the Environmental Indicators—will then allow the Agency to develop the environmental management area of the FEA DRM.

Update the target Applications Architecture

The target architecture will be updated in FY 2005 to include initial solution-oriented and MDA concepts and models, and to formalize implementation of the DIP to link with data warehouses and data marts such as those developed this year for the Office of Air. Details of the DIP will be completed by the first quarter of FY 2005, permitting formal review and approval of this initiative. The Agency will then conduct detailed case studies of the major Agency applications, starting with those current in the DME stage of their system life cycle. Case studies will include an initial overview of business process modeling to address the potential for business process

improvements as these systems move toward DME. Based on the technical results of the initial implementation of the Portal and the Data Integration Platform, it will be possible to create system-specific investment plans for migrating Exhibit 300 systems to the Central Services. These will be incorporated in the target architecture and Solution Standards.

Update the target Technology Architecture

In FY 2005, the TASP, which documents the process for injecting emerging technologies into the EPA EA, will be updated (it was first developed in FY 2003). In addition, the EPA TRM, IT Roadmap, and the FEA TRM will be fully aligned. This will include linking individual standards of the IT Roadmap with the seven major areas of the EPA TRM.

Update the Sequencing Plan

Based on completion of the Solution Standards and the initial round of warehousing projects, one of the most important priorities in FY 2005 will be to update the Sequencing Plan to include detailed investment schedules of EPA's major mission-critical systems as they integrate with the Central Services. These schedules will be developed in collaboration with program offices, based on system case studies and updates to the target architecture described above.

The PMO-led construction projects will provide information necessary for the detailed sequencing of systems to the target architecture. If the completion of these projects is delayed for any reason, some necessary information may not be available in time for the FY 2007 CPIC cycle. However, because the revised applications architecture no longer assumes that every application must migrate as a unit, it will be possible to specify the sequencing of components individually. For instance, the FY 2007 CPIC cycle may be able to provide definitive sequencing information for CDX or the Portal, but not for the Framework for Business Area Warehouses.

The updated investment schedules of the major systems will be incorporated—at whatever degree of technical specificity is possible—into the FY 2007 CPIC submissions developed over the third and fourth quarters of FY 2005.

Create a regional architecture for Region 4

Although the target EA already includes the programs, organizations, applications, data flows, and technologies of the EPA regional offices, in FY 2004 the EA Team is launching an initiative with EPA Region 4 (Atlanta) to investigate regional architectural issues in detail. This project will continue in FY 2005 and produce a regional architecture at least for Region 4, and possibly a template for expanding regional architectures as components covering other EPA regions.

Expand business domain architectures: Cleanup Target Architecture, Emergency Management

FY 2005 will continue to see activity in the development of component architectures such as the Cleanup Target Architecture and a new Emergency Management architecture. Additional component architectures may be initiated for other business areas.

Continue work on the RSA

As noted previously, the target RSA will be expanded through Joint Architecture Development (JAD) sessions that will complete the details of the above focus areas. Upcoming priorities include the completion of the RSA BRM and initiation of an RSA DRM.

3.2.3 Architecture Build-out

The build-out of the EPA target architecture by the PMO will continue and accelerate in FY 2005. The following projects and initiatives directly support the build-out and are referenced in the EIIPD FY 2005 CPIC proposal.

Complete the PMO-led architecture build-out projects

The five PMO-led build-out projects, including the AQS Analysts Portal and the AQS Data Mart (which will have the most technical significance in developing the Solution Standards), will continue in the coming year. They are currently scheduled for completion in January 2005.

Complete the ASA build-out projects

The ASA Data Initiative, Strategic Process Alignment, and other ASA initiatives will continue in FY 2005 and are currently scheduled for completion in that fiscal year.

Continue the RSA build-out projects

The RSA Science Subnet and Data Grid initiative are long term projects that are expected to continue throughout FY 2005 and beyond. The initial version of the Environmental Science Portal is expected to be operational with FY 2005, though it—like all the other program-specific portals—will continue to be built and enhanced continuously.

Design and deploy new data model registry

The Central Services currently include multiple data registries identified in the EPA target architecture: the Data Element Registry (the EDR); the new XML Registry (implemented in FY 2004); and the Data Set Registry (based on the existing Environmental Information Management System (EIMS)). This project will create and deploy the initial version of the new data model registry in FY 2005.

Launch the Data Integration Platform in production environment with AQS Data Mart

The Data Integration Platform now exists in a development environment. In FY 2005, the development version of the platform will be completed and transferred to a production platform to serve the new AQS Data Mart and other data warehousing and data mart development efforts.

Prototype infrastructure implemented; initial applications integrated with infrastructure

As scheduled in EPA's *Enterprise Architecture Status Report 2003,* the PMO is committed to complete the initial functioning version of all elements of the Central Services infrastructure by the end of FY 2005. At that time, initial applications will be using all elements of the infrastructure. This schedule will be met by the end of FY 2005 through the completion of new Central Services elements and the enhancement of existing elements.

Deliver financial information in dashboard format

One of the initial Portal offerings will be implementation of business intelligence services. Building on the capability of the existing ORBIT system, the PMO will implement a financial information "dashboard" for senior EPA managers. This dashboard will provide high-level information on the execution of the FY 2005 budget and other financial indicators, such as grants and acquisitions management.

Develop Version 1 of the Portal

The first quarter of FY 2005 will see the launch of the first version of the EPA Portal Framework and the operation of at least one functioning Portal. The Portal Framework is a software system that supports the creation of multiple Portals and allows for the sharing of sub-portal elements (or

"portlets") across multiple Portals. Version 1 of the Portal will draw on existing LAN identity management capabilities pending the completion of the full Identity Management and Access Control system, scheduled for implementation in the fourth quarter of FY 2005.

Phase II implementation of Enterprise Content Management System (ECMS)

Work in Phase II will expand ECMS to assist other regions in their ECMS deployment plans, and provide programs, regions, and laboratories with Documentum software access for development and deployment. EPA anticipates acquiring an enterprise license to the software, building the enterprise hosting environment and installing adequate storage capacity, and upgrading necessary network connections. EPA will conduct an alternatives analysis to determine the most cost-effective and operationally reliable architectural solution.

3.2.4 Communications

Comprehensive communications and training program

Completion of the Interim EA Policy and the upcoming review and approval of the EA procedures increased the need for a comprehensive communications outreach program. All sectors of the Agency must hear consistent messages at all levels—from the PMO as well as the architecture staff. In addition, FY 2005 implementation of four of the EA procedures (change management of the baseline, change management of reference models, alignment of EA with CPIC process, and IT System Compliance) will require aggressive training programs to ensure compliance.

ART maintenance and content management

The ART model is the source reference of the baseline and target architectures. ART maintenance and content enhancement are therefore a continuing priority of the EA Team. In FY 2005, in addition to updating the baseline and target architectures, ART will be enhanced to include representation of integrated planning programs, including strategic planning, budget planning and execution, capital investment planning, and potentially grants management and acquisitions management.

Support for the cross-federal architecture community

EPA's participation in the federal enterprise architecture community includes the CIO co-chairing the Architecture and Infrastructure Committee (AIC) of the CIO Council, leading the Architecture Plus Seminar Series, and participating in the AIC Governance, Components, Emerging Technologies subcommittees and Chief Architects' Forum. This participation requires continued staff support from the EA Team throughout the year. EPA also provides a liaison from the AIC to the line of business initiatives. In addition, EPA will actively work to support implementation of the E-Gov initiatives.

3.3 Much remains to be done

Although much has been accomplished, it is clear from the above list of upcoming tasks and architectural enhancements that much remains to be done. The enterprise architecture field is evolving even as EPA's implementation approaches the goals set two years ago. The Agency is, however, now in a position to target its activities efficiently and effectively to areas that will pay off most—increasing the EA's practical utility and methodological maturity at the same time. EPA is committed to making the EA program actionable—an integral part of the Agency's high-level planning processes and a management tool at the system life-cycle level to ensure more efficient, performance-based information management investments

ATTACHMENTS

This page is intentionally left blank.

Attachment A: Interim EA Policy

This page is intentionally left blank.

U.S. EPA
Interim Enterprise Architecture Policy

Issue Date: September 9, 2004
Review Date: One year from approved date.

Subject	Establishment of Enterprise Architecture Policy
Purpose	The purpose of this U.S. Environmental Protection Agency (EPA) Enterprise Architecture Policy is to ensure the Enterprise Architecture Program affords the Agency the opportunity to maximize the business value of EPA's investment in information technology and minimize the amount of unnecessary redundancy resulting from disparate planning and development efforts. More specifically, this policy: a. Establishes the governance of EPA's Enterprise Architecture Program; b. Sets direction for how the Enterprise Architecture will be developed and maintained; c. Establishes how information technology (IT) investments will be made and maintained in compliance with the Enterprise Architecture; d. Facilitates alignment of EPA's current business processes, strategic planning, human capital planning, organizational structures, information management programs, and budget and capital investments with the EPA's mission, goals, and objectives; e. Supports changes and enhancements to the Agency's business, as directed by Congress or the President, and in response to the Administrator's priorities and program office and regional initiatives; f. Supports and enhances information system development and practices to allow EPA and its partners to assess environmental and human health information with a holistic view; and g. Promotes public access to environmental information through the effective use of information technology.
Policy	**I. Governance and General Practice** There is one authoritative Enterprise Architecture for the Agency for which the CIO establishes a governance structure. The Enterprise Architecture is an authoritative representation of the Agency's strategic direction, organizational programs and projects, lines of business, information technology portfolio (i.e., data, applications, and technologies), security measures, and the inter-relationships among them. It is maintained for

the purpose of supporting the Agency's strategic planning, budget formulation and execution, information technology capital planning, information technology acquisition, human capital, and security planning processes.

The Enterprise Architecture serves as a sub-component of the Federal Enterprise Architecture and is therefore maintained in alignment with Federal Enterprise Architecture and e-Government requirements.

This Policy is supplemented with procedures, technical standards, and guidelines as established and published by the CIO or designee. Compliance with the procedures and technical standards is mandatory. Compliance with the guidelines is mandatory unless a senior official in the sponsoring organization has approved an alternate practice that is preferred and beneficial for the particular business concern.

II. Enterprise Architecture Development and Approval

All Assistant and Regional Administrators or their designees participate in the development and maintenance of the Enterprise Architecture and its subcomponents.

Security is a primary consideration in the development and implementation of the Enterprise Architecture.

The Enterprise Architecture and any subdivisions are developed utilizing a common framework and methodology that ensures conformance with the Federal Enterprise Architecture.

The Enterprise Architecture is approved by the Chief Information Officer.

III. Enterprise Architecture Maintenance and Toolset Usage

The Enterprise Architecture is updated to align with the Agency's strategic and budgetary plans and the information technology capital planning process.

The Enterprise Architecture is maintained under version and configuration control within the Agency's Enterprise Architecture repository and toolset.

Any subdivisions of the Enterprise Architecture repository and toolset must be interoperable with, and recorded within, Agency's Enterprise Architecture repository and toolset.

IV. Enterprise IT Investment Portfolio Compliance

Compliance with Enterprise Architecture Policy, procedures, technical standards and guidelines is mandatory for all EPA organizations, including contractors that direct, develop, maintain, and operate information technology systems in support of the Agency.

All EPA information management and technology acquisitions must comply with the Enterprise Architecture.

All new and modified information technology development projects must

	include the development of a Solution Architecture (see Definitions), that ensures adherence to the Enterprise Architecture and the Enterprise Architecture Principles identified in A. Solution Architectures must be approved by the Chief Architect prior to project development. All IT systems, applications, data, and metadata, must be recorded within an authoritative inventory.
Background	The Clinger-Cohen Act requires federal agencies to establish Enterprise Architecture programs and focus on the results achieved through capital investments, while streamlining the federal information technology procurement process. EPA has implemented an Enterprise Architecture that must be maintained and updated under formal direction and governance, align with the Federal Enterprise Architecture, and support Presidential initiatives and Executive Orders. In December 2002, EPA published proposed policy for establishing an Enterprise Architecture Practice, Enforcement Function, and Enterprise Architecture Tool Usage (U.S. EPA, *Target Architecture*, December 2002). This Policy further establishes that proposal. In December 2002, EPA also adopted guiding principles for the Enterprise Architecture. Those principles, and any subsequent revisions to them, will serve as the foundation for this Policy, related procedures, guidance materials, and all future revisions (see Appendix A).
Audience	This policy applies to all EPA personnel, agents or others authorized to work or conduct business for EPA. Further, this policy applies to Agency partners, EPA contractors and recipients of EPA financial assistance (e.g., grants, and Interagency Agreements (IAGs)).
Authorities	See Appendix B
Related Documents	See Appendix C
Waivers	Published procedures will provide for a waiver process from selected provisions of associated procedures, technical standards, and guidelines.
Roles and Responsi-bilities	**EPA Administrator:** The Administrator is the champion of Enterprise Architecture, responsible for communicating its value as an enterprise management tool. **Assistant Administrators and Regional Administrators:** The Assistant Administrators and Regional Administrators ensure that their organizations actively participate with the Chief Architect and comply with the target architecture. They may also develop Component Architectures in alignment

	with the Enterprise Architecture. **Chief Information Officer (CIO):** The CIO has the responsibility and authority for the Enterprise Architecture Program, providing strategic direction, and enforcing its requirements. The CIO establishes, maintains, and approves the Enterprise Architecture. The CIO, or designee, supplements this Policy by approving procedures, technical standards, and guidelines. **Chief Financial Officer (CFO):** The CFO is the responsible authority for: (a) all architectural considerations required under the Chief Financial Officers Act of 1990 (the CFO Act) and (b) ensuring the Enterprise Architecture and the Capital Planning and Investment Control processes are integrated with strategic and budget planning. **Chief Technology Officer (CTO):** The CTO is responsible for procedures, technical standards and guidelines associated with the technology and security architectural components. **Chief Acquisition Officer (CAO):** The CAO ensures information technology services contracts contain requirements for compliance with the Target Architecture through the EPA Acquisition Regulation. **Chief Architect (CA):** The CA is responsible for providing direction to the Enterprise Architecture development and maintenance, and ensuring its coordination with the Federal Enterprise Architecture and EPA's IT collaborations with state, local, and tribal partners.
Definitions	**Agency Applications Inventory:** The authoritative registry that uniquely identifies EPA's diverse information resources, including computer application systems, Web sites, databases, models, and data sets. **Architecture:** The structure of components, their interrelationships, and the principles and guidelines governing their design and evolution over time. **Baseline Architecture:** The set of products that portray the existing enterprise, the current business practices, and technical infrastructure. Commonly referred to as the "as-is" architecture. **CPIC Process:** The Capital Planning and Investment Control (CPIC) process, mandated by the Clinger-Cohen Act of 1996, is a systematic approach to selecting, managing, and evaluating information technology investments that requires federal agencies to focus on the results achieved through information technology investments. **Enterprise:** An organization (or cross-organizational entity) supporting a defined business scope and mission. An enterprise includes interdependent resources (e.g., people, organizations, and information technology) that must coordinate their functions and share information in support of a common mission (or set of related missions). **Enterprise Architecture:** A strategic information asset base that defines the

	mission, the information and technologies necessary to perform the mission, and transitional processes for implementing new technologies in response to changing mission needs. An enterprise architecture includes a baseline architecture, target architecture, and a sequencing plan. **Enterprise Architecture Artifacts:** The graphics, models, and/or narrative that depict the enterprise architecture. **Information Technology:** Applied computer systems, both hardware and software, and often including networking and telecommunications, usually in the context of a business or other enterprise. Often the name of the part of the enterprise that deals with all things electronic. **Repositories and Tools:** A collection of databases, architectural and modeling tools, and other electronic support for developing, modeling, managing, analyzing, and publishing the Enterprise Architecture baseline architecture, target architecture, and sequencing plan. Collectively, the Enterprise Architecture repositories and tools comprise the strategic information asset base of the Enterprise Architecture. **Sequencing Plan:** A document that defines the strategy for changing the enterprise from the current baseline to the target architecture. It schedules multiple, concurrent, interdependent activities and incremental builds that will evolve the enterprise. **Solution Architecture:** An architecture that transforms the business and use cases of the Strategic and Business Architectures into the requirements for systems architects to specify to system developers how to design and implement the Application, Data and Technology Architectures. The Solution Architecture demonstrates compliance with the Target Architecture. **Target Architecture:** The set of products that portray the future or end-state enterprise, generally captured in the organization's strategic thinking and plans. Commonly referred to as the "to-be" architecture.
Additional Information	For further information about this Policy, see www.intranet.epa.gov/architec, or contact the Chief Architect in the EPA Office of Environmental Information, Office of Technology Operations and Planning, Information Technology Policy and Planning Division.
Review Date	One year from approval date.

Appendix A
Enterprise Architecture Principles

The following Enterprise Architecture Principles, published in the *U.S. EPA Target Architecture*, December 2002, provide the conceptual foundation of the EPA Enterprise Architecture Policy.

General Principles

a. Information resource management activities comply with applicable laws, regulations, policies, procedures, and standards.

b. Interoperability is a goal of infrastructure, data management, and applications development.

c. Enterprise-wide access to information, based on users' business need for, and rights to, that information, is the rule rather than the exception.

d. Individual system development or implementation efforts must not dictate enterprise architecture, enterprise data models, enterprise object models, enterprise business process models or enterprise business rules without an independent review from an enterprise level.

Business Architecture Principles

a. Business objectives must be well-defined before information technology solutions are initiated. Business processes (along with security requirements) drive the data architecture, applications architecture, and technical architecture.

b. Maximizing value is the main objective when making information technology decisions.

Data Architecture Principles

a. Information is not only an organizational asset but also a national asset for which our employees are the stewards.

b. All enterprise-level data have a shared enterprise-wide definition and are accessible to all organizations in the enterprise. Enterprise-level data are data that are collected or compiled by the Agency and which support its mission to protect human health and the environment.

c. Data is captured once and validated at the source or as close to the source as possible. Data will not be gathered multiple times from the same source.

d. Data names will be standardized between applications and across database platforms.

e. Data access is to be independent of physical data location.

Application Architecture Principles

a. Cross-functional applications are highly encouraged where appropriate.

b. Business and information requirements should be met using commercial off-the-shelf (COTS) technology rather than customized or in-house solutions, whenever practical.

c. Applications are designed to be highly granular and loosely coupled. In other words, applications should be flexible to allow for the possibility of future re-partitioning and/or reuse of different application components to adapt to changing business needs and requirements.

d. The integration of information must adhere to good business practices.

e. Applications should derive from established needs of internal and external customers and be able to effectively respond to emerging needs.

f. Applications should minimize data collection burden and use resources efficiently.

g. Applications should provide effective and economical access to EPA and other data.

h. Applications should authenticate the sources of information.

i. All applications will incorporate approved data standards by the recommended date.

j. The applications architecture should support effective data exchange between data partners of widely varying resources.

k. Applications should be robust enough to handle changing business needs.

l. Applications should be deployed across an N-Tier, distributed computing environment.

m. Flexibility should be designed into the architecture to accommodate continuing business changes and improvements in technology.

n. Established industry vendors, best practices, technology standards and products should be invested in being at the forefront of current trends.

Technology/Security Architecture Principles

a. The technology architecture's ability to adapt to user needs is paramount.

b. The technology architecture is based on open technology and should rely on reusable components.

c. Security is designed into all architectural components, balancing accessibility, availability and ease of use with protection of the integrity of the data and, where appropriate, protection of the confidentiality of the data.

d. A level of security should be applied to resources commensurate to their value to the organization and sufficient to contain risk at an acceptable level.

e. External systems are assumed to be insecure and internal security features should be planned accordingly.

f. Application security should be based on open standards and use security features and services consistent with open standards where possible.

g. Layered protections should be applied to mitigate vulnerabilities in COTS products, isolate system components and limit or contain vulnerabilities.

h. Information security is primarily an issue of governance and process, not technology; the applicable organizational goals and policy must therefore serve as the framework on which the remainder of the security architecture is built. Effective change management is key to maintaining a secure environment.

i. Proactive identification of vulnerabilities and assessment of the associated risks is essential.

j. Audit and monitoring mechanisms should be used to secure information resources.

k. Public access systems should be isolated from mission critical resources.

Appendix B
Authorities

a. The Clinger-Cohen Act of 1996 (also known as the Information Technology Management Reform Act of 1996) (Pub. L. 104-106, Division E);

b. The Government Performance and Results Act of 1993 (GPRA) (Pub. L. 103-62);

c. The Chief Financial Officers Act of 1990 (31 U.S.C. 3512 et seq.);

d. The Federal Information Security Management Act of 2002 (which amends the Computer Security Act of 1987 (Pub. L. 100-235));

e. The Paperwork Reduction Act of 1995 (Pub. L. 104-13);

f. The Government Paperwork Elimination Act of 1998 (Pub. L. 105-277, Title XVII);

g. The E-Government Act of 2002 (Pub. L 107-347);

h. The Rehabilitation Act of 1998 (Pub. L. 105-220);

i. The Federal Managers Financial Integrity Act (FMFIA) of 1989 (Pub. L. 97-255);

j. The Federal Financial Management Improvement Act (FFMIA) of 1996 (Pub. L. 104-208);

k. The Privacy Act, as amended (5 U.S.C. 552a);

l. The Budget and Accounting Act, as amended (31 U.S.C. Chapter 11);

m. The Federal Acquisition Streamlining Act (FASA) of 1994;

n. The President's Management Agenda, Office of Management and Budget, Fiscal Year 2002;

o. Executive Order 13011, Federal Information Technology, FR 61-140, July 19, 1996;

p. OMB Circular A-11, Preparation, Submission and Execution of the Budget, revised July 25, 2003;

q. OMB Circular A-94, Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs, revised January 22, 2002;

r. OMB Circular A-123, Management Accountability and Control, dated June 21,1995;

s. OMB Circular A-127, Financial Management Systems, dated July 23, 1993;

t. OMB Circular A-130, Management of Information Resources, dated November 28, 2003.

Appendix C
Related Documents

a. U.S. Environmental Protection Agency: 2003 Strategic Plan, September 30, 2003;

b. U.S. Environmental Protection Agency: Human Capital Strategic Plan, December 2003;

c. U.S. Environmental Protection Agency: Baseline Architecture, September 2002;

d. U.S. Environmental Protection Agency: Target Architecture, December 2002;

e. U.S. Environmental Protection Agency: Enterprise Architecture Status Report 2003, September 2003;

f. Performance Reference Model (PRM) Version 1.0, Vols. 1 & II, Federal Enterprise Architecture Program Management Office, September 2003;

g. Business Reference Model (BRM) Version 2.0, Federal Enterprise Architecture Program Management Office, June 2003;

h. Data Reference Model (undated draft);

i. Service Component Reference Model (SRM) Version 1.0, Federal Enterprise Architecture Program Management Office, June 2003;

j. Technical Reference Model (TRM) Version 1.1, Federal Enterprise Architecture Program Management Office, August 2003;

k. U.S. EPA IT Roadmap and Standards Profile;

l. Information Resources Management Policy Manual, EPA Directive 2100B8;

m. Interim System Life Cycle (SLC) Policy, Directive 2100.4;

n. Interim System Life Cycle (SLC) Procedure, Directive;

o. Data Standards Policy;

p. EPA Acquisition Regulation, Information Resources Management clause;

q. Capital Planning and Investment Control (CPIC) Policy, Directive 2100.2al;

r. Procedure: *Obtaining a Waiver from an EPA IT Requirement* (in draft).

This page is intentionally left blank.

Attachment B: Framework for EA Procedures Development

This page is intentionally left blank.

U.S. Environmental Protection Agency
Framework for Enterprise Architecture Procedures Development

I. INTRODUCTION

This framework identifies and describes in brief the procedures required to implement EPA's Enterprise Architecture (EA) Policy. The Chief Architect, working in collaboration with the appropriate committees and workgroups (e.g., Enterprise Architecture Coordination Committee (EACC), Administrative Domain Executive Steering Committees, Technology Architecture Work Group (TAWG)) will define the governance structure, process steps, and roles and responsibilities for each procedure outlined in this framework. New procedures will be added to this framework as requirements are identified by the Chief Architect. The Chief Information Officer (CIO), or designee, will approve the procedures and associated technical standards and guidelines.

As stated in the Enterprise Architecture Policy, compliance with the Enterprise Architecture procedures is mandatory.

Figure B-1 depicts EPA's overall model for the EA policy structure as a four-tier sequence with policy at the apex, supported by procedures, technical standards, and guidelines. This figure also reflects the nine currently identified procedures described in this framework.

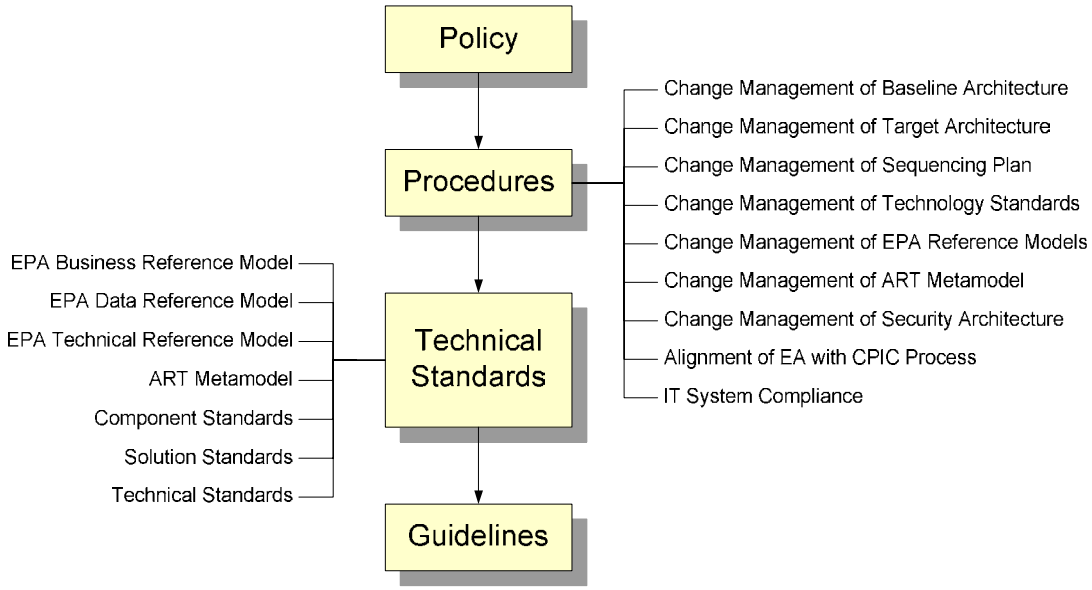

Figure B-1: Enterprise Architecture Policy Structure

II. GOVERNANCE

The Enterprise Architecture Policy authorizes procedures for development and maintenance of the EA and its sub-components. These procedures will provide the Agency the approach, methodology, framework and tools required to successfully govern the EA Program. Development, maintenance, and compliance with the EA procedures, and associated technical standards and guidelines, will be administered by a collection of steering committees, boards and work groups sanctioned by the CIO. The following governance structures have been established.

- **Quality and Information Council (QIC):** serves as reviewing and governing authority for the Enterprise Architecture and endorses all changes to the target architecture and sequencing plan.
- **Domain Executive Steering Committees:** guide development and approval of domain architectures (e.g., Administrative Systems Architecture, Research and Science Architecture).
- **Enterprise Architecture Coordination Committee:** provides Agency-wide input to the EA Core team with respect to policy, procedures, standards, work products and management practices as they relate to the development, maintenance and leveraging of a unified Agency Architecture.

Roles and responsibilities, including individuals and groups will be provided as part of each procedure document developed within this framework.

III. PROCEDURES

The procedures are grouped into the three categories identified in the EA Policy: Enterprise Architecture Development and Approval, Enterprise Architecture Maintenance and Toolset Usage, and Enterprise IT Investment Portfolio Compliance.

Procedures for Enterprise Architecture Development and Approval

Enterprise Architecture Development and Approval include procedures for changing the baseline architecture, target architecture, and sequencing plan, including special focus area architectures such as the Security Architecture. Each procedure is described below.

Change Management of the Baseline Architecture
As the target architecture is implemented, the baseline architecture must be updated to reflect the new "as-is" state. This procedure will identify the process that operates over the course of the year, aggregating information from detailed program office information to the more general view of the as-is state of the architecture at the enterprise level. This process will be keyed to the annual capital investment cycle, producing an updated baseline architecture in the Architecture Repository and Tool (ART).

The procedure will define the formal steps the Chief Architect will take to revise the baseline.

Change Management of the Target Architecture
The target architecture is the "to-be" state of the enterprise and specifies the target state of the Agency's business, programs, data and information technology infrastructure. The target architecture and any subdivisions must be developed under a common framework and methodology that ensure Federal Enterprise Architecture (FEA) conformance. The enterprise and subdivision target information must be captured and maintained in ART. Updates to the target architecture must occur annually in conjunction with the Agency's strategic and budgetary processes.

The procedure will define the formal steps the Chief Architect will take to update the target architecture, in collaboration with the EACC, and obtain approval through the appropriate senior-level official or governing body.

Change Management of the Security Architecture

Security is integrated into all levels of the EA, and aligned with the emerging requirements of the FEA Security Profile. This procedure will provide the process for change management of the security architecture in compliance with the FEA Security Profile and in alignment with the broader EPA enterprise architecture.

The procedure will define the formal steps the Chief Architect will take to update the security architecture, in collaboration with the EACC, and obtain approval through the appropriate senior-level official or governing body.

Change Management of the Sequencing Plan

The sequencing plan reflects the integration of funding decisions made in the budget and capital investment planning processes with the modernization plans of information management systems. It defines the milestones and sequence of activities necessary to migrate from the baseline to the target architecture. The sequencing plan describes the activities necessary to move the Agency from the baseline to the target and must be adjusted according to the funding decisions and target architecture. All programs and architecture subdivisions must comply with the updates to the target architecture.

The procedure will define the formal steps the Chief Architect will take to update the sequencing plan, in collaboration with the EACC, and obtain approval through the appropriate senior-level official or governing body.

Procedures for Enterprise Architecture Maintenance and Toolset Usage

Enterprise Architecture Maintenance and Toolset Usage include procedures for maintaining the Technology Standards, EPA Reference Models, and ART Metamodel. Each procedure is described below.

Change Management of Technology Standards

Technology Standards are an important element in the implementation of an Enterprise Architecture. EPA must have a procedure for updating its standards to accommodate technology advancements in the industry and the Agency's changing business requirements.

EPA maintains an IT Roadmap that lists individual products the Agency has approved for use. Part of the roadmap—the Standards Profile—lists our core product inventory. This procedure will define the process for merging the annual IT Research Agenda, which infuses new technologies into EPA's EA, with the process for reassessing and updating existing IT standards.

The IT standard-setting portion of the procedures will define new target standards, will transition existing target standards to operational standards and operational standards to legacy standards, and will sunset legacy standards.

Change Management of EPA Reference Models

EPA has defined an EPA Business Reference Model and an EPA Technical Reference Model. Additional EPA-specific reference models may be required and will be developed in the future. The EPA reference models must be maintained in alignment with their counterpart FEA reference models. A change management procedure is therefore required to maintain the EPA reference models.

This procedure will formalize the process for reviews and updates of the EPA reference models in response to changes in the Federal reference models. It will also support modifications required as a result of new or amended Agency business functions or further decomposition of the models to meet program needs.

Specifically, the procedure will define the formal steps the Chief Architect will take to update the EPA reference models, in collaboration with the EACC, and obtain approval through the appropriate senior-level official or governing body.

Change Management of the ART Metamodel
The Enterprise Architecture Policy requires maintenance of the EA under version and configuration control within the Agency's Enterprise Architecture Repository and Toolset—ART. The EA Program uses a series of architecture artifacts, stored in ART, to capture and maintain enterprise and subdivision architecture information

The ART metamodel provides the essential shared artifacts for architectural modeling at all levels, from the enterprise down to the program level. This procedure will formalize the process for ensuring timely response to program office modeling needs.

The procedure will define the formal steps to be taken by the Chief Architect to update the EA artifacts, and their representation in the ART metamodel, in collaboration with the EACC, and obtain approval through the appropriate senior-level official or governing body.

Procedures for Enterprise IT Investment Portfolio Compliance

Enterprise IT Investment Portfolio Compliance includes the procedures for IT Systems Compliance, and Alignment of EA with Capital Planning and Investment Control (CPIC) Process. These procedures are described below.

IT Systems Compliance
EPA's Systems Life Cycle (SLC) process will become the foundation for EA implementation and compliance. This procedure will formally define the touch points between EA and SLC. It will also define the reviews necessary to ensure that individual systems comply with the architecture through effective solution architecture[1] planning and execution. This procedure also defines reviews for architecture exemptions and describes the process for resolving non-compliance within a reasonable time period.

Solution architectures must be approved by the Chief Architect or designee (e.g., business Component Architecture domains, program offices, regions, laboratories), as determined by the CIO.

Alignment of EA with CPIC Process
This procedure will describe the reviews and approvals necessary for integrating EA requirements into the capital investment planning process. It will also define the requirements for generating timely guidance to the investment planning community. It will describe the roles of OEI, Programs, and Regions for managing IT investment monitoring practices relating to enterprise architecture. Further, this procedure will establish the actions required to ensure that unnecessary redundancies in information technology investments are eliminated.

[1] A Solution Architecture is a blueprint of an individual IT project

Attachment C: Target Research and Science Architecture (RSA)

This page is intentionally left blank.

Table of Contents

List of Tables and Figures

Attachments

Executive Summary

The target Research and Science Architecture (RSA)[1] Goals were derived directly from the higher level goals of the Agency's Enterprise Architecture but are specific to the Environmental Protection Agency's (EPA) research and science work. Those higher level Agency goals are commonly known as the Agency's five "GPRA Goals" because they reflect Agency direction set forth in accordance with the Government Performance and Results Act of 1993 (GPRA). In addition to the GPRA Goals, the target RSA also incorporates specific research and science direction that may be associated with the related EPA "Cross Goal Strategies." With that in mind, the principal goal for the target RSA is expressed as follows:

> RSA is laying the technical foundation for building the infrastructure that will meet the Information Technology needs/requirements of EPA researchers and scientists.

A baseline RSA was developed to describe the current research and science environment, providing the basis on which to build the blueprint for its future Research and Science Architecture. This baseline covered the business and functional models in depth and provides a listing of high level data types, major applications, and technologies.

The baseline RSA findings generated objectives that guided the development of the target RSA, including the following:

- Manage scientific information to provide accurate results quickly.
- Provide the appropriate scientific computing services infrastructure on demand.
- Leverage technology to ensure that information of known quality is used by Agency decision-makers.
- Implement effective, efficient and secure collaboration.
- Manage geospatial assets and support the EPA Geo Blueprint.
- Leverage existing investments via a catalog of reusable EPA research science technologies and coordination of the evaluation and piloting of emerging technologies.

To develop the target RSA in accordance with these findings, Focus Areas were formed to address each of the above objectives. These six target RSA Focus Areas became the primary means by which the target RSA was developed to meet its goals, as well as the Agency's goals.

[1] In September 2002, the EPA Baseline Enterprise Architecture included the baseline Research and Development Architecture (RDA), which was developed primarily by EPA's Office of Research and Development (ORD). In FY 2004, the RDA was expanded beyond ORD to include all science activities across the Agency. It was accordingly renamed the Research and Science Architecture (RSA).

This page is intentionally left blank.

1. Introduction

The U.S. Environmental Protection Agency relies on science, technology, and scientifically defensible data and models to evaluate risk, develop and defend protective standards, anticipate future health and environmental threats, and identify their solutions. Consequently, understanding how EPA conducts research and science and communicating these structures and processes to internal and external stakeholders alike is vitally important.
In particular, key drivers such as the Government Performance and Results Act of 1993, the Clinger-Cohen Act (also known as Information Technology Management Reform Act) of 1996, and a collection of Executive Orders and Office of Management and Budget (OMB) Directives are placing increasing emphasis on developing outcome-based metrics that can be linked to budget formulation and expenditures. In response to these requirements and drivers, the Agency is transforming the way it conducts business and manages environmental information.

1.1 Target RSA Domain in the Context of EPA's EA

To respond to these drivers, the Agency has embarked on a comprehensive effort to create an EPA-wide EA. Initial EA efforts focused on three functional domains – the Administrative Systems Architecture, the Environmental Health and Protection Architecture, and the Research and Science Architecture – and several cross-cutting domains, including the Technology and Security Architecture and the Geospatial Architecture. These business-focused domain architectures contribute to and are coming together with the Agency EA. Both the EPA EA and the target RSA are defined in terms of five layers of the Enterprise Architecture Framework "pyramid" so that the target RSA is naturally integrated with the EPA EA as depicted in Figure C-1 below.

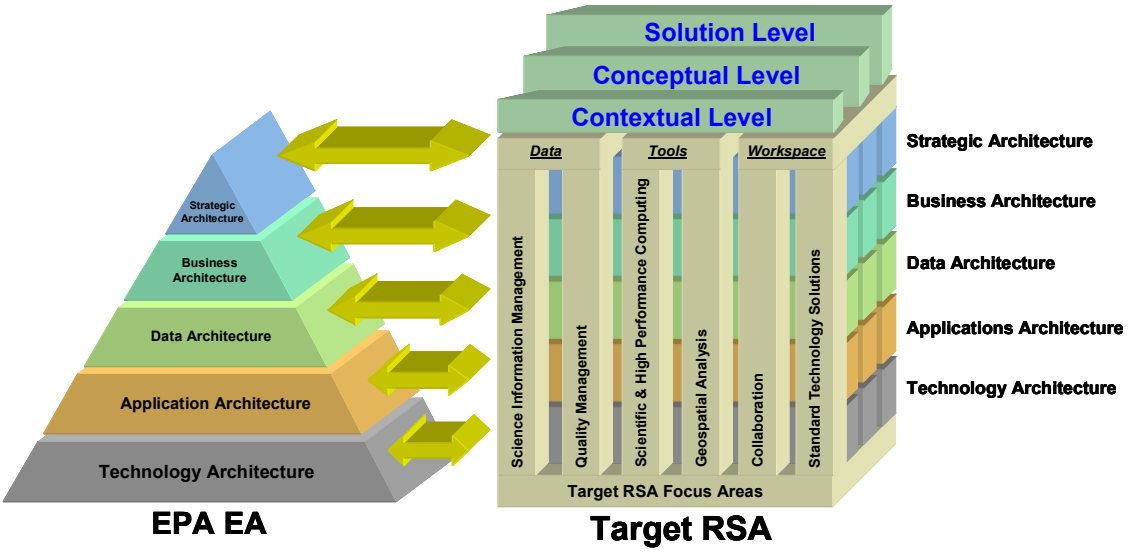

Figure C-1: Target EPA EA and RSA Alignment

1.2 Target RSA Overview

The RSA will provide a comprehensive structure to coordinate Agency research and science information management and technology resources holistically, efficiently, and effectively. The RSA includes all components of EPA engaged in science, from applied research and science for improving the fundamental understanding of human health and environmental relationships to problem-driven research and science to resolve issues of high risk or scientific uncertainty. The RSA contributes to strategic and tactical science planning by creating a collaborative work environment where the workspace, data, and tools required to create and to manage knowledge to support EPA's mission are easily found, accessed, evaluated, and used to meet existing and emerging environmental challenges.

To manage the magnitude and complexity of the Research and Science Architecture in its entirety, the target RSA was approached from the perspectives of six crosscutting Focus Areas (see Figure C-1) aligned with the activities and resources required to support and enhance EPA's research and science business processes. Each RSA Focus Area presents a unique view of the target RSA and covers issues specific to that perspective. The six Focus Areas are characterized as follows:

- **Scientific Information Management**
  - Electronic capture and management of data
  - Integrated management of science data
- **Quality Management**
  - Electronic pedigree of research and science information
- **Scientific and High Performance Computing**
  - On-demand access to high-end science and research tools
  - Shared system resources through Grids and clusters
- **Geospatial Analysis**
  - Uniformly managed and accessible Geospatial data
  - Indexed and catalogued Geospatial data
- **Collaboration**
  - Collaboration Workspace for EPA researchers and scientists
- **Standard Technology Solutions**
  - Catalog and manual of technology solutions
  - Re-use and re-purposing of available EPA technology

1.3 Target RSA Governance

Contributions to the RSA are first shaped by leads for each of the six Focus Areas, who present products to a staff-level RSA Workgroup (WG) for input, and finally vet approved RSA materials through an RSA executive steering committee (ESC). Focus Area leads act as stewards for all the RSA activity related to their assigned Focus Area. The RSA WG, composed of functional and IT experts from across program offices and regions, is responsible for detailed planning and coordination of RSA data collection and product development. Members of the RSA ESC review products that have been recommended for approval by the RSA WG. Following RSA ESC review and approval, RSA products are sent through the Agency's EA Coordination Committee (EACC) for incorporation within the EA. The Agency EA governance process culminates with executive

review by the Quality and Information Council (QIC). This relationship is shown in Figure C-2 below.

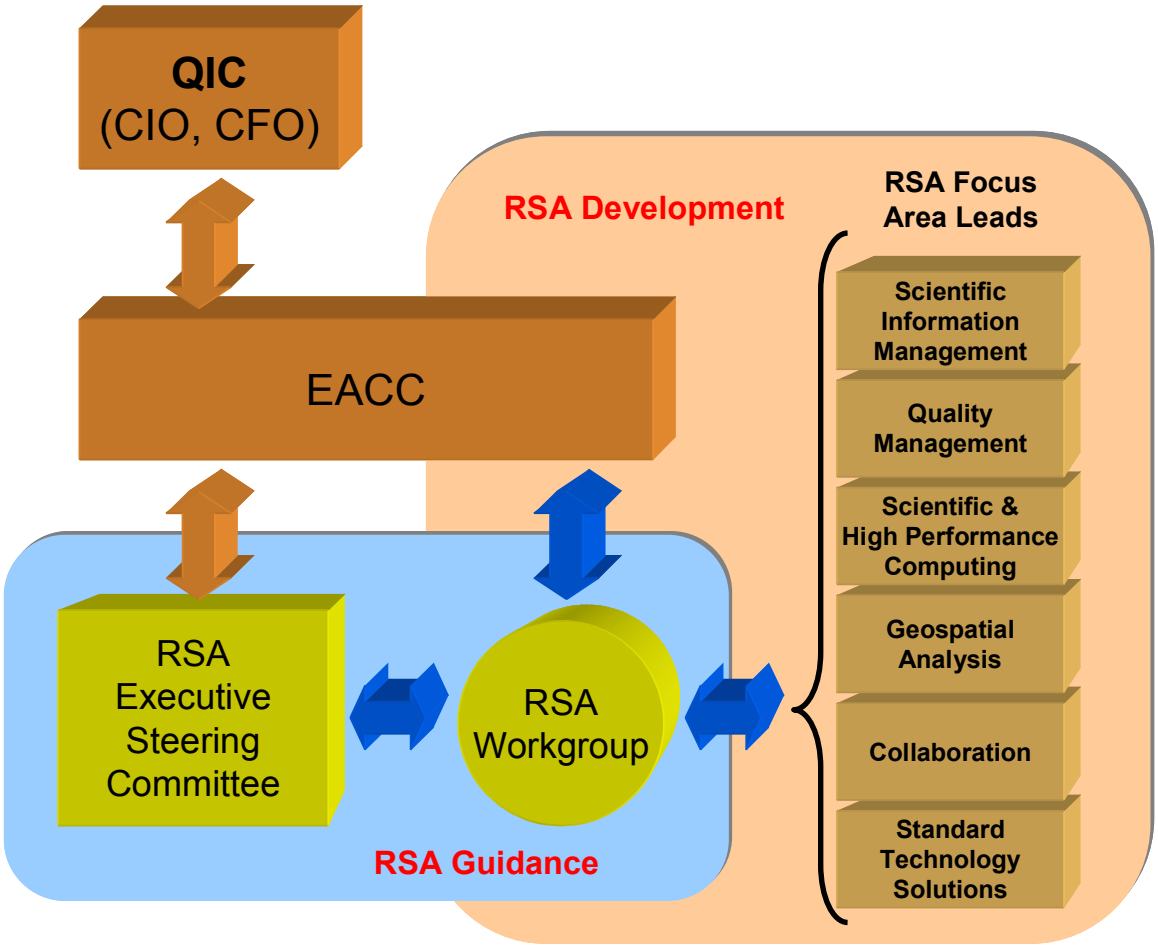

Figure C-2: Target RSA Governance

1.4 Objectives and Scope of this Document

The RSA represents EPA's Research and Science business domain and responds to the Agency's cross-goal objective to "Enhance Science and Research" as identified in the *EPA Strategic Plan 2003-2008*. The RSA is fully aligned with Agency- and government-wide efforts to standardize planning and architecture.

This document provides a Mission Context overview of the first target RSA which will support EPA research and science with the technology needed to better support EPA's mission. The target RSA further provides a description of the intended Business Functions and Defined Solutions of the target RSA that will be developed in subsequent RSA releases. As shown in Figure C-3 below, these target RSA Levels will be built successively. More specifically, the target RSA provides the Agency with the information necessary to:

- Develop the RSA Transition Sequencing Plan
- Support IT investment review and capital planning for research and science assets

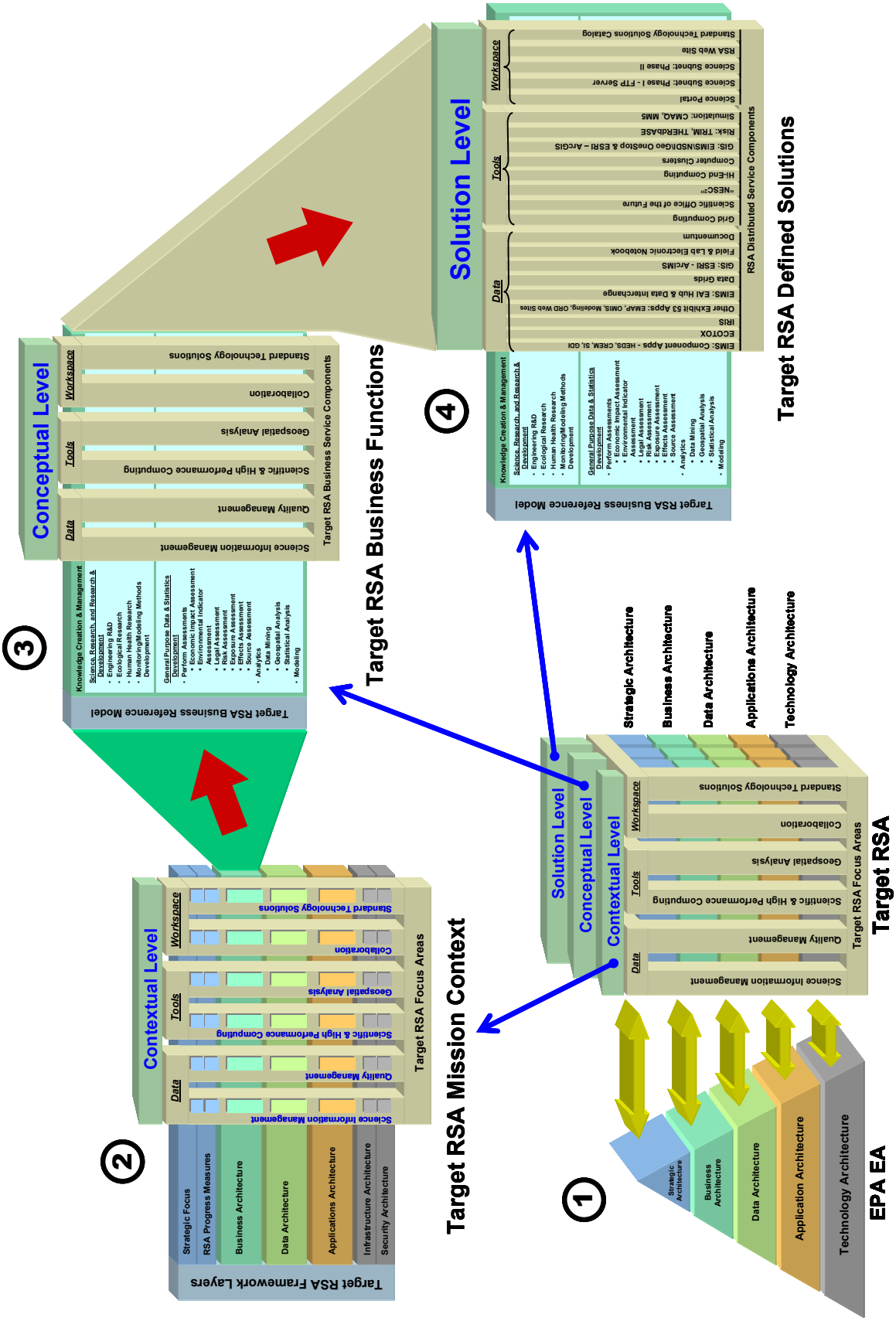

Figure C-3: Target RSA Context, Concept, and Solutions

2. Target RSA Mission Context

The target RSA is defined in terms of the five layers of the EPA Enterprise Architecture Framework "pyramid" as discussed before, but a couple of the Framework layers are further subdivided for the target RSA as depicted in Figure C-4, below.

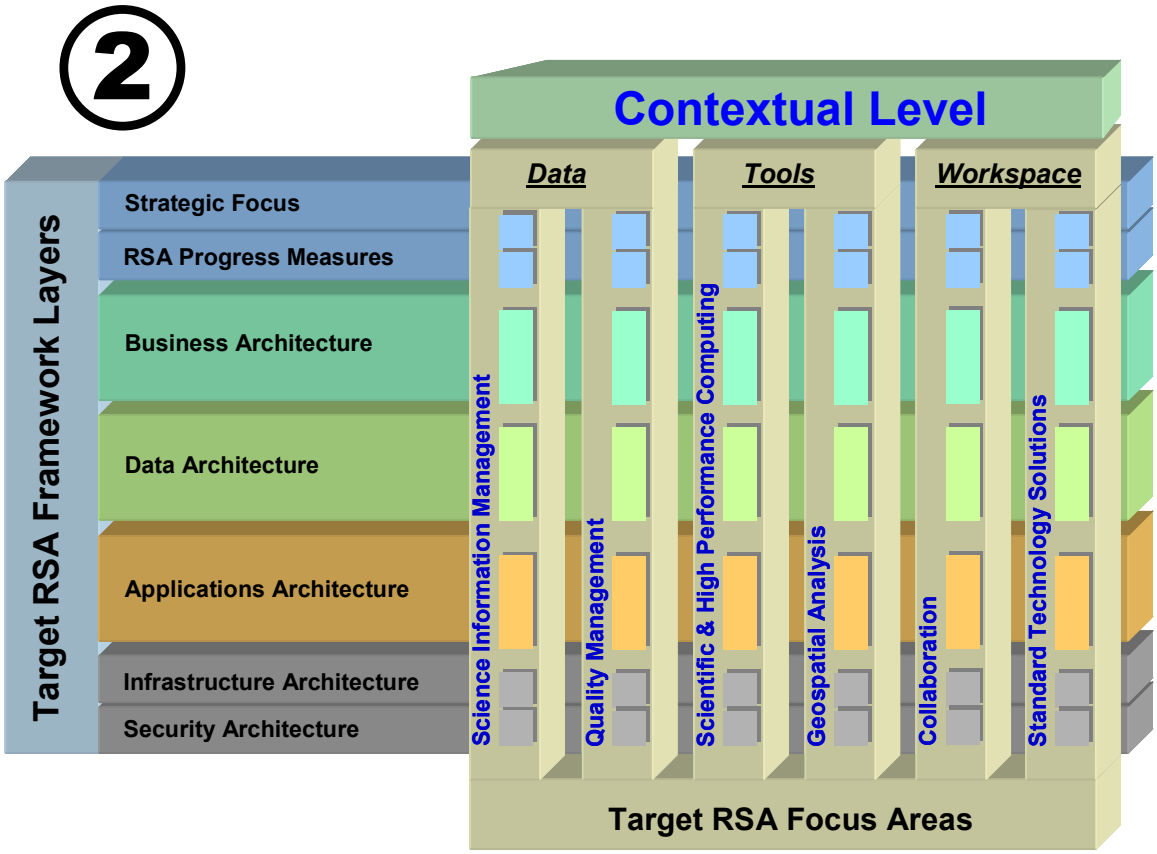

Figure C-4: Target RSA Mission Context

For the target RSA, the Mission Context is the highest and most general level of specification in the Enterprise Architecture representation hierarchy. While the subsequent, lower levels (i.e., the Business Functions and Defined Solutions) provide more detailed architectural elements and mechanics, the Mission Context simply captures and communicates the Conceptual Level of each layer of the architectural framework (i.e., Goals, Business, Data, Applications, and Technology). To further understand what enables each Framework layer, the target RSA also has another dimension in addition to that of the layers which consists of six Focus Areas.

The following subsections describe first, the essential features of each of these Framework layers for the target RSA, then the nature of each of the target RSA Focus Areas, and finally the properties that each target RSA Focus Area contributes to each target RSA Framework layer. These relationships are graphically depicted in a matrix (see Section 2.3) highlighting the features of these intersections.

2.1 Target RSA Framework Description

The main purpose of the target RSA Framework is to provide a disciplined and structured means to guide, manage and improve the way that information management can support research and science conducted at EPA. This is in an effort to generate, access and reuse reliable scientific information that is based on science priorities, rooted in sound practices, and performed in an effective and efficient manner. In the sub-sections that follow, each target RSA Framework layer will be discussed with respect to its own purpose and some of the key characteristics that the six target RSA Focus Areas contribute to their effectiveness.

2.1.1 Strategic Architecture Layer

The Strategic Architecture layer describes how the target RSA is to be positioned to meet its mission and goals. It provides the focus for accomplishments and means to measure progress toward achieving them.

Strategic Focus

The principal goal is further described via the following, supporting goal statements:

- Perform priority-driven research and science.
- Produce sound and defensible research and science.
- Make research and science information resources available both within the EPA and to external customers and partners.
- Effectively coordinate research and science efforts and resources.

RSA Progress Measures Sub-Layer

The governance of the RSA has to be conducted with progress measures in mind. These measures illustrate how well the target RSA is fulfilling its role for EPA as it is developed and used.

Some of the key progress measures provided by the Focus Areas in the target RSA are:

- The extent of EPA science records that have been made available
- The extent of quality document and metadata pedigrees validated
- The extent of ready access to Scientific and High Performance Computing tools
- The extent of geospatial governance with respect to EPA science and research
- The extent of availability of data and tools behind the firewall to beyond the firewall
- The extent of best emerging technology applied to EPA science and research

2.1.2 Business Architecture Layer

At the Contextual Level, the Business layer of the research and science community involves executing its Science Life Cycle process, which will be described in a later section. This is done by members from EPA program offices, regions, centers, laboratories, and non-EPA organizations with partnerships and other collaborative relationships with the Agency, including public and private U.S. and international entities such as other Federal Agencies, tribes, and academia.

Some of the notable Business layer features provided by the Focus Areas in the target RSA are:

- The ability to monitor model methods as they are employed at EPA
- The ability to use Quality Management (QM) procedures for research and science inside EPA
- The ability to develop new models and applications for EPA research and science
- The ability to use geospatial approaches when conducting research and science modeling and monitoring at EPA
- The ability to collaborate on experiments for EPA research and science
- The ability to discover and propagate new technology solutions

2.1.3 Data Architecture Layer

The target RSA's Data layer is generally organized to serve the six target RSA focus areas in an interoperable way that was identified and investigated to develop the target RSA. This represents a logical grouping of the major subject areas or classes of data that are projected to encapsulate research and science information required by the Business layer. There is overlap across these categories that will be resolved as a natural product of refining the Data Architecture which will result in a data representation that manages any overlaps in an effective and transparent way.

Some of the highlights of the data layer provided by the Focus Areas in the target RSA are:

- The data related to science information repositories
- The data related to document quality pedigrees
- The data contained in very large scientific data sets
- The data contained in geospatial data sets (including vector and raster data)
- The data related to collaborative experimentation
- The data related to implementing the IT Roadmap

2.1.4 Applications Architecture Layer

The target RSA Applications layer is currently organized as a high-level summary of major application classes that are projected to encapsulate the applications required to support and work respectively with the Business and Data layers above. The RSA baseline identified over 300 different applications, many of which are expected to be included in the target RSA; however, at the Contextual Level, identifying application types and key enterprise solutions rather than specific narrowly defined applications is required to yield a technology independent Applications layer that supports the layers above effectively.

Some of foremost elements of the Application layer provided by the Focus Areas in the target RSA are:

- Environmental Information Management System (EIMS)
- Electronic Field and Lab Information Management Systems
- Simulation Models such as Community Multi-scale Air Quality (CMAQ)
- EPA-tailored Geographic Information System (GIS) Visualization tools
- Science Portal for EPA's researchers, scientists and their collaborators
- Literature search tools

2.1.5 Technology Architecture Layer

The Contextual target RSA Technology layer identifies the technologies required to satisfy the Infrastructure and Security requirements of the layers above. Such technologies will ultimately be integrated into the Agency's Technology Architecture and IT Road Map through continued coordination with the Technology Architecture Work Group (TAWG). This layer is also expected to evolve as research and science technology requirements are gathered and collectively represented and satisfied across the Agency through more RSA driven efforts in the future. In the meantime, the RSA development team will continue working with groups, such as the TAWG, and the research and science community to collaborate accordingly and consequently shape a technology environment that best supports EPA research and science in view of the Agency Enterprise Architecture.

The types of technologies and/or Commercial Off-The-Shelf (COTS) applications are a sampling of what has been identified as relevant, Contextual Level target RSA requirements, based on an initial analysis of the six target RSA Focus Areas, including:

Infrastructure Sub-Layer

- Data Grid and EPA-adapted ESRI-ArcIMS for information management
- Documentum for quality management
- Grid Computing and Computing Clusters for scientific computing
- High Speed, High Bandwidth networks for transferring geospatial data sets
- Subnet access behind the EPA firewall for researcher and scientist collaboration
- Standard utilities and COTS applications for routine system support and applications

Security Sub-Layer

- Security Management
  - Identity Management
  - Role-Based Access Management
  - Policy-Based Access Management
  - Information Assurance
  - Releasability Data Tagging
- Secure Hosting
  - High Performance Computing
  - Grid Computing
  - Computer Clusters
  - High-end Platforms
  - Communications Equipment

2.2 Target RSA Focus Areas Description

As described earlier, each of the six target RSA Focus Areas is a distinctive grouping of the information management and technology characteristics that enable the target RSA to be effective, while also covering issues specific to that perspective. These Focus Areas can be also grouped according to their function in providing services in the target RSA into one of three categories. That is, they are associated with Data, Tools, or Workspace types of functionalities in

the target RSA. These six target RSA Focus Areas will each be discussed in turn below with respect to what comprises them and some of the key solutions they provide for the target RSA.

2.2.1 Data Group

These are perspectives on the distinctive and often unique data generated by research and science activities.

Science Information Management Focus Area

Under the target RSA, scientific information will be managed efficiently, transparently, and holistically. The RSA community must cope with a geometrically-expanding proliferation of data, gathered and developed by collaborating scientists within the Agency and across the country. Different scientists and institutions have different working styles, needs, and approaches to managing data, but regulatory-based science demands transparent and reliable documentation and access. These different approaches must be integrated into a single architecture.

The target solution lies in the strategic use of modern Science Information Management (SIM) tools, some of which are already in use to a certain extent.[2] Effective integration of these tools with emerging new applications that improve the management of scientific data and geospatial information will lead to more consistent and thorough metadata and the establishment of more repeatable data management procedures across the research and science community. The RSA SIM initiative addresses the information management problem holistically, forming a foundation for achieving the other RSA objectives discussed below.

Some examples in the Science Information Management focus area are repository systems that are based on the underlying functionality of the data management application. For instance, GIS includes place-based display and analysis, and Document Management Systems focus on text search and version management. A "derived collection" refers to arrangements where data, which are available for wider access, have been structured and organized for secondary use. Sub-categories of derived collections include portals, warehouses, data marts, and information products from partner organizations. Example applications corresponding to derived collections are the Science Inventory (SI), ECOTOXicology (ECOTOX) database, Human Exposure Database System (HEDS), and the other "Specialized Information Management Systems." As these systems are enhanced to become integrated and interoperable, SIM data structures will be modified to accommodate those changes.

Quality Management Focus Area

Under the target RSA, the quality of science will be more efficiently controlled and documented. Science data of known quality are the foundation of sound environmental policy and regulatory decisions. The Agency's science activities must be planned, managed, and documented appropriately to provide products of known quality. Therefore, leveraging science information management technology to better support the EPA Quality System is a major theme of the RSA effort. Improving data quality is not a new problem, but it is central to the RSA. This issue relates to Scientific Information Management, but it focuses less on systems for data management than on explicit RSA support for the EPA Quality Program's policies, procedures, and standards.

For example, in the Quality Management focus area, "sound" science is science that is appropriate to the Agency mission, reproducible, and defensible. The RSA makes a significant contribution to sound science at EPA by ensuring that tools are readily available to facilitate the capture, management, and retrieval of critical Quality Assurance (QA) information. It is

[2] The Environmental Information Management System is already being adapted to use within the Central Services. Documentum™ has been selected as the basis for E-Content Management.

anticipated that three Agency standard and pending standard applications, Documentum, Waters-NuGenesis (pending), and the EIMS, will be configured to provide the required functionality. Each of these run on Agency standard technology and thus do not require the enhancement of the Technology layer of the architecture. In addition, improving the data capture in the pursuit of 'lossless' documentation will require the Electronic Field and Lab Notebook technology currently in development.

2.2.2 Tools Group

These are perspectives on the extremely complex and very powerful types of tools that research and science activities require.

Scientific and High Performance Computing Focus Area

The target RSA will make high-performance computing available for many more uses than it is today. Many of EPA's concerns are amenable to so-called *in silico* approaches—those that rely on environmental simulations and computational modeling. Air dispersion analysis, climate change, waste load allocation in water, subterranean pollutant migration and many other problems cannot be adequately addressed other than through large-scale computer modeling. *In silico* science is especially adaptable to problems that are either too large or too small in scale to study on the bench, i.e., *in situ*. Furthermore, a host of analyses that used to be done on the bench can now be done on computers, often more appropriately and cheaply. The goal of the RSA is to make the most of costly high-end computing resources, efficiently allocating access to high-end machines and networks and assigning tasks appropriately to different classes of computing.

EPA will participate in such special purpose Grids as the Biomedical Information Research Network (BIRN), the United Nations Environmental Programme (UNEP), or the Access Grid Network (AGN). Under the RSA, as the power of individual platforms increases, and as computing Grid services and data Grid services become more sophisticated (including the increasing ability of scientific instrumentation of all types to talk to computers), and as computing models become more integrated and interoperable, the scope and practicality of *in silico* science will increase dramatically. The cost efficiencies of this approach are obvious; the practical impacts on day-to-day scientific operations will be transformative.

Within EPA, most of the Scientific and High Performance Computing (S&HPC) computing platforms scaled larger than workstations (i.e., distributed computing applications, large scientific instruments, clusters, and high-end computers) will be "Grid aware", meaning that they will be identified as a Grid asset and administered and shared through Grid services, or Grid Computing. Local groups of computing and storage assets will be managed within "campus Grids", for example, within individual Labs and Centers. The campus Grid assets, along with other Agency computing and storage assets, will make up EPA's "enterprise Grid". This enterprise Grid will give scientists access to aggregated computing and storage capacity across the entire Agency that they otherwise would not have. Through Grid services, scientists will submit their jobs and allow the Grid administration to manage the job scheduling, submit jobs to the right size platform at the right time, manage job queues, and notify the scientist when a job is done. Optionally, scientists can provide information to restrict the jobs (e.g., must complete by a certain time, must run only on a particular platform).

Geospatial Analysis Focus Area

The RSA envisions the complete integration of geospatial information into the analysis and display of environmental information through the use of better tools and high speed networks. Although geospatial data, analysis, remote sensing and scientific visualization are cross-cutting components of the entire EPA Enterprise Architecture, they are of particular concern to the RSA.

Environmental research and science efforts are inherently oriented toward geospatial information and rely heavily on geospatial data tools.

Geospatial data is defined as vector and/or raster information that identifies, depicts, or describes the geographic locations, boundaries, or characteristics of inhabitants and natural or constructed features on the Earth. This includes such information derived from, among other sources, socio-demographic analysis, economic analysis, land information records and land use information processing, statistical analysis, survey and observational methodologies, environmental analysis, critical infrastructure protection, satellites, remote sensing, airborne imagery collection, mapping, engineering, construction, global positioning systems, and surveying technologies and activities. It also includes individual point or site specific data that are referenced to a location on the earth and digital aerial imagery of the earth.

As geospatial data and tools become increasingly available, existing *ad hoc* project-specific approaches for developing and/or procuring must give way to more comprehensive and systematic approaches across the entire Agency. Additionally, many EPA offices are starting to use spatial data and tools in their everyday business. Because of escalating uses of spatial data and tools, and the need to access databases and websites with spatial data on a regular basis, cost and time efficiency in data purchases, data storage and retrieval, and spatial analysis applications can only be achieved through a carefully developed EPA program on geospatial governance and architecture.

Establishing an effective governance structure to articulate research and science related geospatial requirements to the broader Agency and interagency geospatial community is key in the Geospatial focus area of the RSA. In FY 2004, ORD appointed a senior scientist for spatial analysis and modeling. The goal is for the senior scientist to work across ORD and with EPA Program and Regional offices to help catalyze development of spatial analysis approaches and tools. In doing so, a number of issues related to geospatial infrastructure and capabilities are likely to arise. The senior scientist will work closely with the EPA GIS working group, OEI, the QIC, and the Agency Geospatial Information Officer to ensure that geospatial requirements for science and research are articulated and addressed. Ready and efficient access to the wide array of remote sensing, GIS, and visualization tools and their associated data and metadata are being pulled together through Web service portlets in the Science Portal under development. The Science Portal must be serviced by exceptionally high-speed networks (i.e., increased bandwidth) for rapid exchange of the extremely large data sets that GIS generates. The Geospatial One-Stop Portal will also be a key source of Geospatial Data generated by partners that is necessary for research and science efforts.

2.2.3 Workspace Group

These are perspectives on the enormous need to support and facilitate collaboration as an integral part of research and science activities.

Collaboration Focus Area

Under the target RSA, scientists across the nation will be able to collaborate more efficiently and conveniently through electronic means. Collaboration is the norm in science, but it is made difficult by necessary security requirements and the large number of dispersed research and science projects that EPA conducts simultaneously. The RSA target addresses collaboration through a number of solutions, including the Environmental Science Portal, the Scientific Office of the Future, and the Science Subnet.

The Environmental Science Portal will be the scientific community's central point of access to environmental information and EPA's computing resources. It will be developed in collaboration

with the Program Management Office (PMO) as a sub-element of the Agency's general Portal, becoming the central point of access to Grid computing systems, data sets, and analytical tools. It will serve the entire RSA community, including researchers, scientists, laboratory staff, managers, support contractors, international partners and stakeholders, states, tribes, private industry, academia, and external contractors and consultants.

The Scientific Office of the Future is an initiative to upgrade the standard hardware and software configurations available to EPA scientists. These systems will support small-scale collaboration efforts at the desktop, mid-size collaboration in which multiple scientists gather to conduct team science with remote partners, and ultimately allow shared access to large-scale and high-end collaborative facilities containing the latest technologies for dynamic, multi-site team research and science activities.

The Science Subnet ties the Office of the Future to other systems. It will be a secure network environment that moves the Agency toward creation of virtual "Collaboratories". The Science Subnet will provide scientists with freer access to tools, data, and each other while protecting secure information systems at their current level.
For example in the Collaboration focus area, collaborative partners of the Agency will include other federal agencies, states, tribal and local governments, industry and environmental interest groups, academia, and international partners. Security considerations will continue to be balanced with collaboration requirements to ensure the protection of EPA's confidential business information (CBI). Collaboration will occur both through Grid and non-Grid collaboration channels. However, many of the science collaboration activities will eventually be conducted through the recently deployed Science Subnet (and not through EPA's public access network) because the fully-implemented subnet will provide a more robust collaboration environment with compensating security controls. The Environmental Science Portal (under development) will also simplify access to appropriate Agency science information products and computing resources for EPA researchers and scientists and their collaborators, making the details of the process invisible to the user. The RSA Science Subnet and Data Grid initiatives are long term projects that are expected to continue throughout FY 2005 and beyond. The initial version of the Environmental Science Portal is expected to be operational in FY 2005, though a beta version has already been demonstrated in FY2004. It is being developed in open source JetSpeed. Like all the other program-specific portals—it will be built and enhanced continuously.

Standard Technology Solutions Area

The target RSA will make it easier for scientists to get access to the technical tools they need by identifying, streamlining, and standardizing existing and emerging research and science technology. Developing a Standard Technology Solutions (STS) catalog will enable the RSA community to efficiently find technology solutions that satisfy their requirements, or to determine when new solution must be sought. It will also help to identify partial solutions that may be leveraged through modification or integration to avoid redundant investments or duplication of effort. This approach maximizes the potential for reuse and collaboration, and directly supports the goals of solution-oriented architecture.
For example, in the Standard Technology Solutions focus area, with the current development of the STS Catalog and its intended distribution and data mining, realizing the STS vision will enable the research and science community to effectively and efficiently find and leverage standard technology solutions that satisfy their requirements and determine when those requirements have not been previously addressed and a new solution must be sought. This will also help identify partial solutions that may be leveraged through modification or integration to further avoid redundant investments and duplication of effort while maximizing reuse and collaboration potential. To assist with the use of the catalog and general document creation, maintenance and distribution, Documentum is one of EPA's standard applications under the STS

vision that is intended to provide an agency-wide cross-cutting technical foundation for document management.

2.3 Focus Area Properties in the Framework Layers

The contributions by each target RSA Focus Area to each target RSA Framework layer are shown in detail in Table C-1 below in terms of selected key properties they provide. Some of these properties are a combination of the characteristics of the Focus Areas that apply to the layer and also solutions that are used from that Focus Area for that layer. Footnotes and acronyms are provided on the next page.

BOLD ITEMS are Sample RSA Solutions To Date In Development or Deployed	Data		Tools		Workspace	
	Science Information Management	Quality Management	Scientific & High Performance Computing	Geospatial Analysis	Collaboration	Standard Technology Solutions
Strategic Focus	**Easy Access and Use of Scientific Data, Information, & Tools**	**Defensible Science through support of EPA Quality System**	**Maximum leverage of computational tools in support of EPA science**	**Easy Access and Enterprise Licensing of Geospatial Data & Tools**	**Maximal Collaboration Among EPA & With Trusted Partners**	**Best Technology ID and Use for EPA Science on a Uniform Basis**
RSA Progress Measures[20]	•Accessible Science Records •Metadata Integrity •Defined Data Admin Processes	•Quality Doc/Meta Data Pedigrees •QM Workflows Supported	•"In Silico" Science Support[4] •Access to Tools •Effective Use of Computing	•Available & Accessible Geo Data and Tools[18] •Metadata Integrity •Established Geo Governance	•Colleague & Partner Sharing •Data & Tools Availability Behind Firewall[17]	•Best Emerging Tech Applied •Apply Uniform Set of Standards
Business Architecture	•Enviro-Science •Enviro-Engineering •Monitoring Model Methods •Assessments	•QM Procedures Inside EPA •QM Procedures Outside EPA	•Develop Models & Applications •Provide Resources and Services	•Logistics •Engineering •Modeling •Monitoring •Assessments	•Metadata Analysis •Experiment •Analyze & Discuss •Publish	•Discover •Innovate •Standardize •Communicate •Coordinate
Data Architecture	•**EIMS** •**Geospatial Data Warehouses** •Science Info Repositories[1] •Derived Info[2]	•**EIMS** •Metadata Quality Pedigrees •Document Quality Pedigrees	•Data Wizards •Very Large Scientific Data Sets •Complex Schema	•Metadata (**EIMS\GDI**)[13] •Vector & Raster[12] •Georeferenced Data •Model Derived •Tool Derived (**ESRI**)[9]	•**Data Grid/EIMS** •Streaming Data •Government Data •Academic Data •Private Data	•IT Roadmap •Technologies under review •Managed Discussion DB
Applications Architecture	•**EIMS component applications[15]** •**Representative RSA systems[16]** •**IRIS** •**GIS**	•Doc, Metadata & Workflow Mgt. •Field & Lab Info Mgmt Systems	•Tool Wizards •**Simulation**[5] •Risk Analysis[6] •Statistical[7] •Info Mining[8]	•**GIS/Visualization**[9] •Geostatistical •Geo-enabled Models and Tools	•**Science Portal** •Workgroup Tools •Web Services •Application & Data Sharing	•**EIMS** •Moderated Discussion Group •Literature search tools
Infrastructure Technology Architecture	•**Data Grid** •**Documentum**[11] •**EIMS** •**ESRI – ArcIMS**[9] •LIMS	•**Documentum** •EPA Standard Technology •**Electronic Notebook**[3]	•**Sci Off of Future**[19] •**"NESC[2]"** •**Grid Computing** •**Hi-End Computing** •**Compute Clusters**[10]	•Imaging Technology •High-Speed Net •GPS Technology •HPC Technology •Wireless/mobile	•**Sci. FTP Server** •**Science Subnet** •Enhanced Subnet •High-Speed Net •AV Conferencing	•**RSA Web site**[14] •Standard Utilities •Specialized COTS (e.g., Notes app.)
Security Technology Architecture	•EPA ID Security Management •EPA Role-Based and Policy-Based Access Management •NTSD (EPA's National Technology Services Division) Models •Information Assurance with Releasability Data Tagging					

Table C-1: Focus Area Properties Aligned with Architectural Layers (footnotes on next page)

TABLE FOOTNOTES:

1. These include Scientific Data Management Systems, Geographic Information Systems, Document Management Systems, Metadata Management Systems, Data Modeling Tools.
2. These Derived Information Collections include include Portals, Warehouses, Information from Partners.
3. This is specifically the Field & Lab Electronic Notebook technology.
4. This is a third leg of science combined with the two traditional ones of Theoretical and Experimental and is also often referred to as Computational. "In silico" Science is brought to bear in the research, science and decision-making activities of EPA where the problems are too large or too small to resolve by ordinary empirical means.
5. These include air quality forecasting, air dispersion models (CMAQ), and meteorological models (MM5).
6. These include TRIM and Total Human Exposure Risk database and Advanced Simulation Environment (THERdbASE).
7. These include commercial products such as SAS applications, ArcView, StatLIb, and RS/1, and EPA-developed tools such as applications for Dunnett's Procedure, Probit Analysis, Trimmed Spearman-Karber Method, and Linear Interpolation Method.
8. These include data mining, information pattern recognition, and knowledge management applications.
9. The predominant GIS applications at EPA rely on the ESRI suite of tools, including ArcView, ArcGIS, and ArcIMS. However, there is also a large collection of other visualization tools available for specialized purposes including Eiffel, AVS, Houdini, Shade Tree, Pixar Renderman, MAYA, and others.
10. These clustered computing platforms are mostly comprised of highly connected, rack-mounted systems that serve as small-scale high-end computing platforms.
11. This is one of EPA's standard applications intended to provide an agency-wide cross-cutting technical foundation for document management.
12. In general, Geospatial Data means vector or raster information that identifies, depicts, or describes the geographic locations, boundaries, or characteristics of inhabitants and natural or constructed features on the Earth. Please refer to the "Geospatial Analysis Focus Area" discussion for a more detailed definition.
13. Involved with Geo OneStop too.
14. Includes STS catalog.
15. EIMS related systems include:

    a. partners using the EIMS database to support a topical web presence, for example the Science Inventory,
    b. partners using EIMS to database information for organizational web-sites, for example NCEA,
    c. partners using EIMS for certain types of information aligned with their own database, for example HEDS,
    d. partners who do not have an individual web presence, but who use EIMS to store information in collections.
    A full list of EIMS partners can be found at www.epa.gov/eims
16. FY06 Exhibit 53 Systems: IRIS, EMAP, EMAP Surface Waters, ECOTOX, HEDS, Modeling (Health & Ecological Research), Modeling (Compliance & Stewardship Research), OMIS, EIMS, ORD Websites (Research Results), ORD Websites (Grants & Fellowships).
17. Establish firewall guidelines to support effective collaboration while managing risk.
18. Key Geospatial data sets have been identified and during FY2005, the science community will set priorities and write business cases.
19. Scientific Office of the Future will soon be known as Scientific Group Workspaces.
20. The RSA Progress Measures are the first step toward developing a complete set of RSA Performance Measures.

TABLE ACRONYMS:

ArcIMS – ESRI solution for delivering dynamic maps and GIS data and services via the Web.

CREM – Council on Regulatory Environmental Modeling knowledge base

ECOTOX – ECOTOXicology database

EIMS – Environmental Information Management System

EMAP – Environmental Monitoring and Assessment Program

ESRI – Environmental Systems Research Institute

GIS – Geographic Information Systems

HEDS – Human Exposure Database System

This page is intentionally left blank.

3. Target RSA Business Functions

The Business Functions of the target RSA are under development and will center on the Owner's Perspective (see Figure C-5 below), as opposed to the Mission Context which is devoted to the Planner's Perspective[3]. The Owner's Perspective describes the Conceptual Level of models, architectures and descriptions used by the individuals who are the owners of the business model. They focus on the usage characteristics of the target RSA products. On the other hand, the Planner's Perspective is devoted to defining the Mission Context of the models, architectures and representations that provide the boundaries for the organization and describe what senior executives must consider when they think about the organization and how it interacts within itself and with the external world.

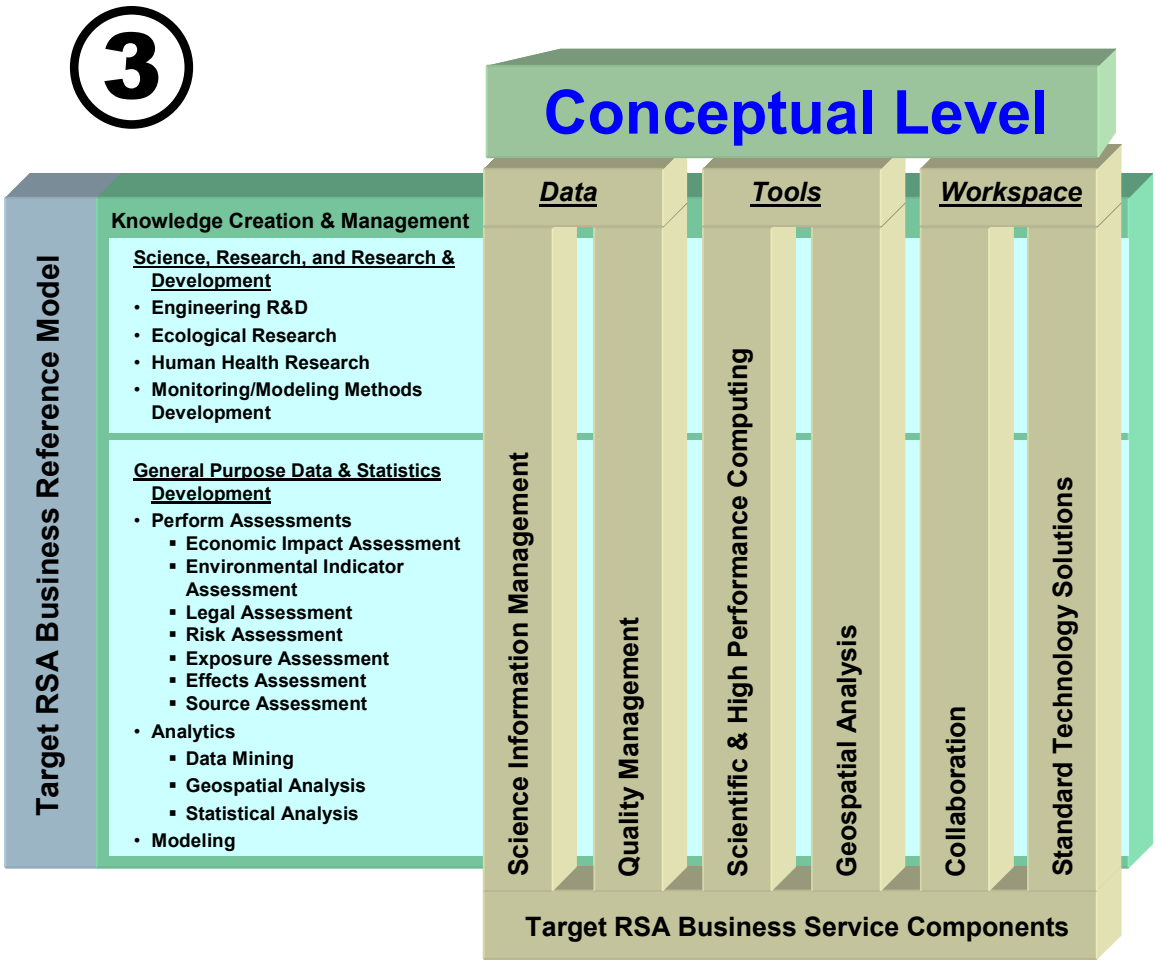

Figure C-5: Target RSA Business Functions

In Figure C-5 above, there are two distinctive changes from the Mission Context in Figure C-4 shown earlier. First, this stage of the target RSA is dedicated to developing the target RSA Business Reference Model. Second, anticipating a greater role for the six target RSA Focus

[3] John A. Zachman, "A Framework For Information Systems Architecture", IBM Systems Journal, Vol26, No3, 1987, 1999, pp 462-466. NOTE: Later versions by Zachman substituted "Planner" for "Ballpark".

Areas, they can now be identified as target RSA Business Service Components as part of building the target RSA toward a Service Component-Based Architecture and aligning the target RSA with the OMB's Federal Enterprise Architecture (FEA) Service Component Reference Model (SRM). This aspect of the target RSA will be described further in a following section.

3.1 Target RSA Business Architecture Layer Details

The target RSA Business Architecture is centered on the target RSA Business Reference Model (BRM). The RSA BRM is also successfully mapped back to the FEA BRM through the EPA BRM, the details of which will be shown in a later section of this report. In the larger picture, the OMB FEA is comprised of five Reference Models. To date, EPA has defined its own version of one of them, the FEA BRM, and the target RSA BRM has been developed as a sub-model within EPA's BRM to form the core of the target RSA Business Architecture.

3.1.1 Target RSA Business Drivers

The RSA Business Architecture in general is greatly influenced from outside the Agency principally by the various initiatives, regulations and legislation that appear and disappear in the Federal government. Strategies developed within EPA with respect to its research and science programs are directly affected by these various Federal government directives as principal business drivers. More specifically, new and changing legislation, Congressional mandates, and Federal initiatives, such as the President's Science Agenda, GPRA, and OMB Circulars, are often the reason EPA needs to undertake certain research and science projects. For example, the President's FY 2004 Budget set forth a research and development (R&D) agenda for that forthcoming fiscal year that reflected the President's priorities and important opportunities for discovery and development while sustaining the basic R&D machinery needed for continued U.S. leadership in science and technology (S&T).[4] These priorities include research and science to combat terrorism, develop nanotechnology, develop networking and information technology, provide a molecular-level understanding of life processes, and further basic knowledge on energy and the environment.

In addition, other outside circumstances such as current public events, high profile areas of public interest, emergency disaster situations, and hot topics in the media also shape the direction research and science takes. For instance, the events of September 11, 2001 caused EPA to change its research and science priorities, emphasizing the need for rapid risk assessments from its scientists and creating a strategic need for a scientific expert database.
Within EPA, collaborative initiatives across EPA Program Offices with Regional and Key Partners are another set of business drivers for EPA research and science. For example, EPA's National Vehicle and Fuel Emissions Laboratory (NVFEL) collaborates with key partners in the automotive industry to achieve mutually agreed upon research goals and drive research and science into the future. NVFEL uses the strategy of forming Cooperative Research and Development Agreements (CRADA) with industry partners to increase the suite of research and engineering resources available. Increasing resources through partnerships provides NVFEL with an increased scope of possibilities and an opportunity to increase the quality of science at EPA.

3.1.2 Target RSA Business Process

The target RSA business processes and organization layer defines the research and science business process and is used by the organizations responsible for achieving research and science goals as stated above. The generic research and science lifecycle (see Figure C-6) reflects the

[4] FY 2005 Interagency Research and Development Priorities, June, 2003

highest level of major activities that are typically performed by the research and science community while conducting business on a day-to-day basis across EPA.

The sequence of the process events is illustrated in Figure C-6 by the circled numbers in the schematic. The cycle begins with 'Plan Projects & Resources' and finishes with 'Assess & Rate Performance' which feeds into the beginning of the cycle each time to provide it with progressively greater findings on which to base further research and scientific work.

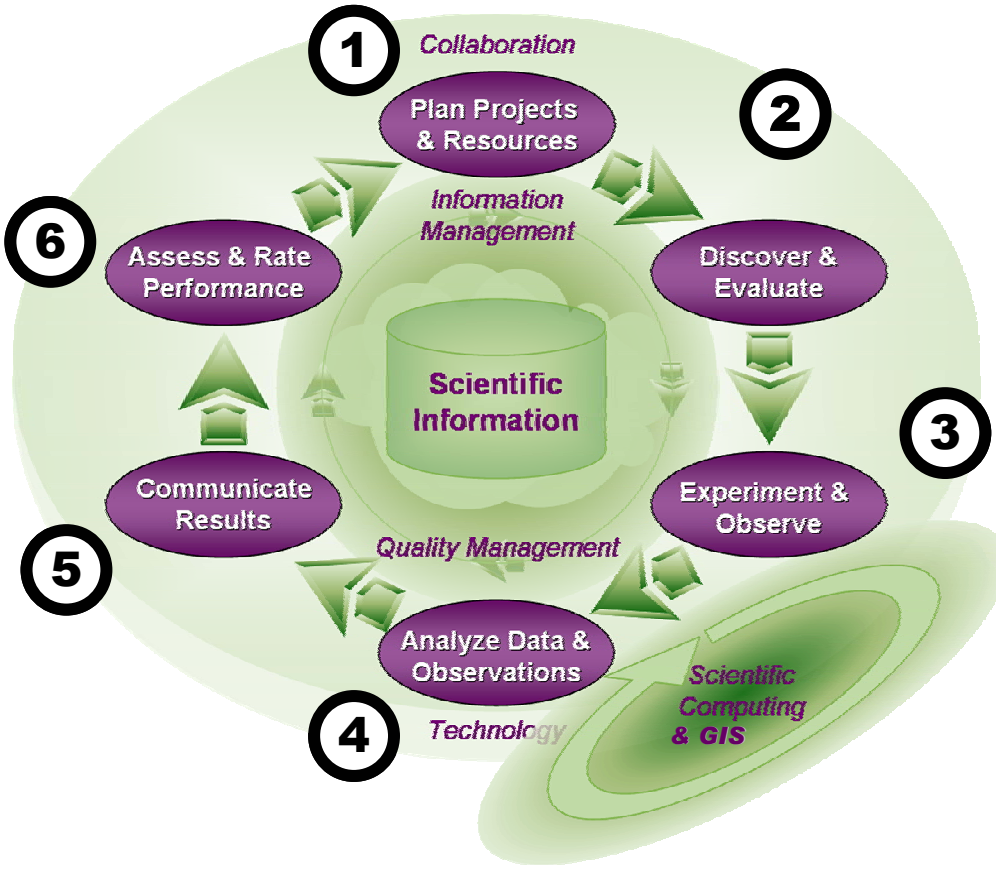

Figure C-6: Research and Science Lifecycle

3.1.3 Target RSA Business Reference Model – Knowledge Creation and Management (FEA and EPA Line of Business)

Knowledge Creation and Management covers programs, capabilities, and activities that make up the core and essential target RSA processes. These are processes in which the Agency creates or develops a body or set of knowledge, the manipulation and analysis of which can provide inherent benefits for the Agency and its stakeholders and further the Agency's delivery of Environmental and Health Protection services.

Science, Research, and Research and Development (FEA and EPA Sub-Function)

All programs that produce original scientific knowledge and technologies to support Agency programmatic needs.

Engineering R&D Capabilities

Capabilities to research and develop engineering solutions, such as civil innovations or mechanical tools, for pollution prevention or pollution control.

Ecological Research Capabilities

Capabilities to develop original scientific information on non-human species and ecosystems within the environment.

Human Health Research Capabilities

Capabilities to develop original scientific information on human health issues related to environmental stresses and exposures to pollutants and wastes.

Monitoring/Modeling Methods Development Capabilities

Capabilities to develop environmental and biological variables, relationships, and specific mathematical formulas. Models enable and indicators validate assumptions about the current state and forecasts of the future state of the environment or human health.

Protocols Development Capabilities

Capabilities to develop protocols and approaches to enhance the replicability and standardization of scientific and technical activities in such areas as monitoring media, using models, and assessing and interpreting risks and hazards.

General Purpose Data and Statistics (FEA and EPA Sub-Function)

All programs providing empirical, numerical, and related data and information pertaining to the current state of the environment, the environmental or health effects of substances, the associated social or economic impacts of proposed environmental actions, and other formal evaluations conducted in delivery of Agency services.

Perform Assessments Capabilities

Capabilities to do analytical work of all kinds leading to actionable assessments on matters of concern to regulatory programs, remediation and pollution prevention.

Economic Impact Assessment Activities

Activities to assess the economic impacts of a proposed action or observed trend.

Environmental Indicator Assessment Activities

Activities to assess the relationships between Agency actions and changes in environmental quality or human health.

Legal Assessment Activities

Activities to assess the legal issues of a proposed action or topic of interest

Risk Assessment Activities

Activities to assess risks to human health or the environment associated with a substance or environmental stressor of interest.

Exposure Assessment Activities

Activities to assess the exposure of receptors, including human species, non-human species (plants and animals), or the built environment, to pollutants or other environmental stressors.

Effects Assessment Activities

Activities to assess the effects of exposures on human species, non-human species (plants and animals), or the built environment.

Social Impact Assessment

Activities to assess the social impacts of a proposed action or observed trend.

Source Assessment Activities

Activities to assess sources or classes of sources that discharge or emit pollutants or other substances of interest.

Analytics Capabilities

Capabilities to do analysis of information for Agency purposes.

Data Mining Activities

Activities to perform the extraction of implicit, previously unknown, and potentially useful information from data.

Geospatial Analysis Activities

Activities to do analysis using mapping, photography, and automated processes that manipulate and display data related to location.

Statistical Analysis Activities

Activities to do analysis of data through the application of statistical techniques.

Modeling Capabilities

Capabilities to do analytical work to model environmental and health processes, such as through the use of computer simulations.

3.1.4 Target RSA BRM Mapping to EPA BRM and FEA BRM

In Table C-2 below, the whole (i.e., core and essential parts combined with supportive portions) target BRM is mapped to the whole Core EPA BRM. Together, the target BRM and the Core EPA BRM are mapped to the whole OMB FEA BRM. Significantly, the six target RSA Focus groups are mapped as part of Information Management, positioning them as key pieces of a Service Component-Based Architecture as the target RSA works through the details of the Conceptual Level.

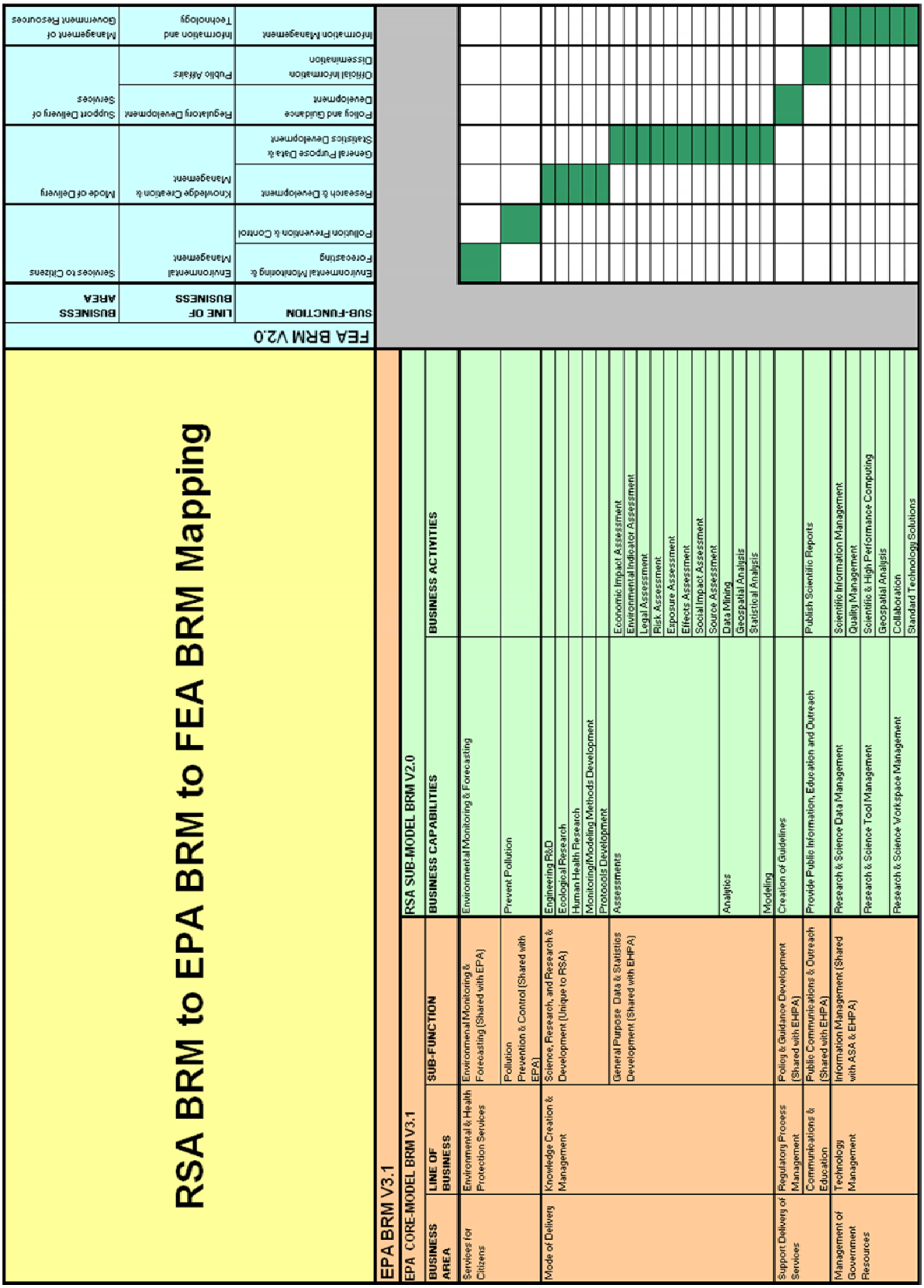

Table C-2: RSA BRM to EPA and FEA BRM

3.1.5 Target RSA BRM Capabilities as Business Service Component Systems

The development of the target RSA BRM Capabilities is not only successful in determining the core and essential business things that the target RSA has to accomplish, but when mapped against the target RSA Focus Areas at a deep enough granularity level (as depicted in Table C-2), they also provide the target RSA with "a set of cooperating business components assembled together to deliver a solution to a business problem."[5] Accordingly, the target RSA BRM Capabilities look like good candidates to function as Business Service Components at the Conceptual Level of the target RSA in a Service Component-Based Architecture (SCBA) (see Appendix A: FEA and Service Component Based Architectures).

3.1.6 Target RSA Focus Areas as Business Service Components

In a way similar to the target RSA Capabilities situation discussed above, the target RSA Focus Areas describe the nature of integral and intrinsic information management resources that the target RSA can use to provide service to the target RSA BRM Capabilities. As a result, the target RSA Focus Areas appear to 'represent the implementation of an autonomous business concept, business service, or business process where they consist of all the technology elements (i.e., software, hardware, data) necessary to express, implement, and deploy a given business concept as an autonomous, reusable element of a large information system.[6] As such, they can most probably be viewed as Business Service Components at the Conceptual Level of the target RSA in a Service Component-Based Architecture.

[5] Service Component-Based Architectures, Version 2.0, CIO Council, June 2004, p.4.
[6] Ibid. p.4 paraphrased.

This page is intentionally left blank.

4. Target RSA Defined Solutions

The target RSA Defined Solutions (see Figure C-7 below) are either under development or deployed and center on the Designer's Perspective as opposed to the Mission Context which is devoted to the Planner's Perspective or the Business Functions which are the Owner's Perspective[7]. The Designer's Perspective describes the Solution Level for the models, architectures and descriptions used by the engineers, architects and those who mediate between what is desirable and what is technically possible. They focus on the logical characteristics of the target RSA products.

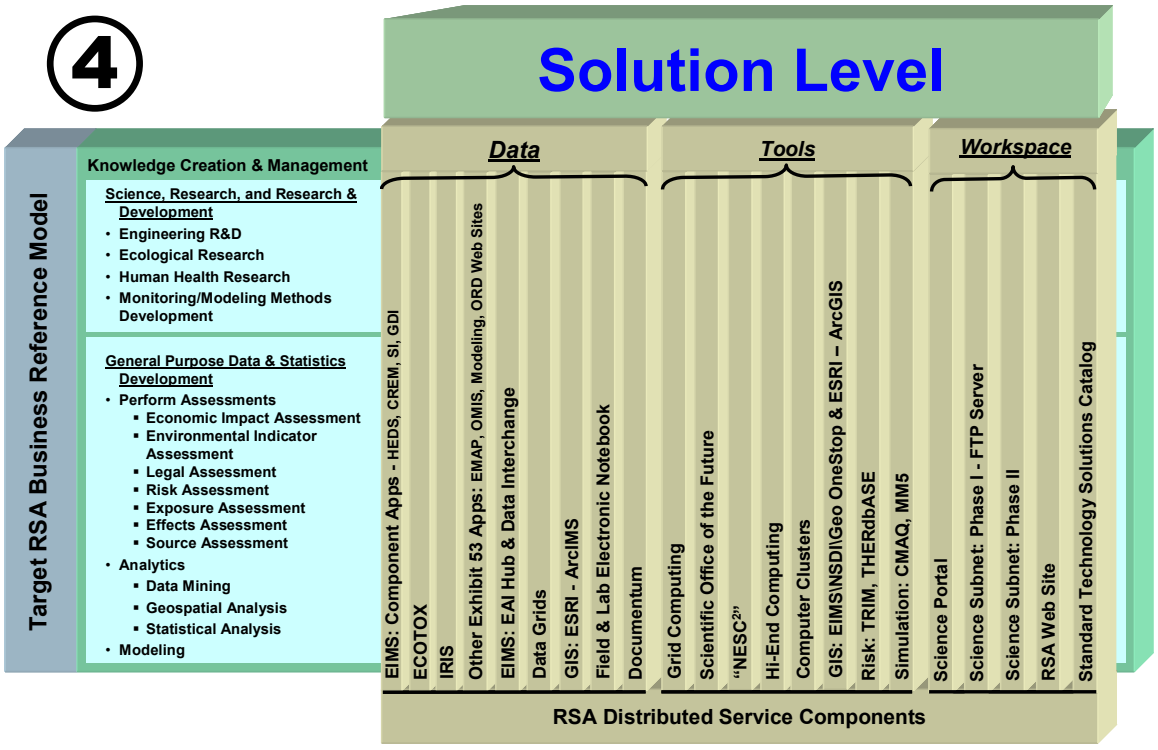

Figure C-7: Target RSA Defined Solutions

In Figure C-7 above, there is one distinctive change from the Mission Context in Figure C-5. As with the Business Functions, this stage of the target RSA is also dedicated to developing the target RSA Business Reference Model. However, anticipating a greater role for the target RSA Defined Solutions (either in development or deployed now) out of the target RSA Focus Areas, they can now be identified as target RSA Distributed Service Components. This will be part of building the target RSA toward a Service Component-Based Architecture and aligning the target RSA with the OMB's FEA Service Component Reference Model. As such, the target RSA Defined Solutions can also be seen as target RSA Defined Services hereafter. This aspect of the target RSA was briefly described further in the preceding section.

[7] John A. Zachman, "A Framework For Information Systems Architecture", IBM Systems Journal, Vol26, No3, 1987, 1999, pp 462-466.

4.1 Target RSA Defined Solutions Descriptions

Research and science Defined Solutions are environmental system products that are developed to assist EPA researchers and scientists in creating written publications and more data sets. These solutions use the data sets to test their development efforts, and researchers and scientists publish written documents on their findings from solution development efforts. Research and science solutions are created through all of the capabilities of the target RSA BRM.

The types of EPA research and science solutions include:

- **Engineering Solutions:** The hardware tools, databases, applications, engines, and other scientific equipment and technologies being developed by EPA research and science. This type of information is usually limited to the metadata about the product because the product is either a physical tool, such as the metadata for a new alternative-fuel engine, or actually part of an application, such as a new web portal for other types of information.
- **Indicators**: Information that signifies the presence of a particular substance, condition, or change. Examples include characteristics of a community or organisms that are proven to signify the presence of particular substances. Through the use of geospatial software and environmental databases, EPA Region 2 is advancing representing the relative burden within a selected community using the following derived indicators: TRI Air Emission, Air Toxics, and Facility Density.
- **Models**: Computational algorithms used to make inferences or predictions about the state of the environment or human health. Examples of this information include physiologically based pharmacokinetic (PBPK) models, benchmark dose (BMD) models, and quantitative structure-activity relationship (QSAR) models. Currently, the state of modeling information is being evaluated by the Council on Regulatory Environmental Models (CREM). The council is setting up a knowledge base to improve the quality and use of models and the information they produce. One significant goal is to make models interoperable, such that relevant data from one model job can be passed seamlessly to models for other, related, jobs.
- **Visualizations**: Graphical data files saved and stored separately from the original source data (e.g., occurrence data sets) used to create the image. These products include GIS and other spatial software produced images. Examples of these data include the three-dimensional digital images of crop formations created by ORD using USGS data.
- **Methods and Protocols**: Scientifically sound procedures and approaches for accomplishing specific tasks, such as measuring and estimating the presence and concentration of specific stressors, their properties, and their effects. Computational toxicology methods are increasingly critical information used to perform assessments without as much animal testing.

4.1.1 Data Services

Published research and science data sets provide the details that researchers and scientists use to perform their meta-analysis, conduct their experimentation, analyze their data, draw their conclusions, and publish their results. Data sets are created in the General Purpose Data and Statistics line of business in the BRM and the experimentation and observation step in the program/project lifecycle. As a result, Data Services are used across all research and science at EPA.

EIMS: Component Apps - HEDS, CREM, SI, GDI

EIMS itself is described below. One of its component applications is the CREM knowledge base, which provides collaborative information on the use and developmental history of EPA environmental models. As a part of the EIMS family of systems, it is the Agency's repository of models. HEDS is also part of the EIMS family of systems which provides a one stop shop for human exposure data. SI is a part too that is a searchable catalog of science activities and scientific and technical work products. Lastly included is GDI, (Geospatial Data Index) which provides a means for EPA employees to quickly locate information about EPA geospatial resources and activities. As EPA's computing models are enhanced to become integrated and interoperable, research and science data structures will be modified to accommodate seamless passing of data between models and computing jobs.

ECOTOX

ECOTOX is a single searchable front for three smaller databases maintained in disparate locations (AQUIRE, PHYTOTOX, and TERRATOX) to provide single chemical toxicity data sets for aquatic and terrestrial life.

IRIS

The Integrated Risk Information System (IRIS) contains an inventory of chemical exposure assessment publications (i.e., health effects) with references the supporting toxicity data sets, in support of the Assessments line of business. It is the main repository for risk-related information in the Agency.

Other Exhibit 53 Apps: EMAP, OMIS, Modeling, ORD Web Sites

The Environmental Monitoring and Assessment Program (EMAP) is a research program to develop the tools necessary to monitor and assess the status and trends of national ecological resources. The ORD Management Information System (OMIS) supports ORD research and science by providing comprehensive resource management info, including planning, resource allocation, and performance management. The ORD Web Sites provide access results of research performed at ORD's labs and centers, as well as grants information.

EIMS: EAI Hub and Data Interchange

EIMS is the Agency's inventory of environmental information in support of assessment activities. The system provides users with the capability to identify and locate metadata about data sets, databases, projects, models, and documents needed to support environmental assessment activities. Descriptive information stored within EIMS is consistent with the Federal Geographic Data Committee (FGDC) metadata content standards for spatial data. A significant enhancement of these standards, however, is the addition of a hierarchical metadata framework that organizes detailed scientific data and documentation, and accommodates customized information at the catalog level to facilitate a review of the different types of metadata in EIMS.

Data Grids

Data Grids provide the capability to remotely access scientific and research data sets.

GIS: ESRI – ArcIMS

ArcIMS from ESRI is a geospatial information management system that allows development and deployment of web-based applications to display and query data that can be stored in multiple very large geodatabases.

Field and Lab Electronic Notebook

The Field and Lab Electronic Notebook is a system to extend electronic recordkeeping to field and lab activities to support defensible research and science at EPA.

Documentum

Documentum is the EPA standard for document management, and as such, is used extensively to create, maintain, and store documents as well as provide access to documents when and as needed.

4.1.2 Tool Services

EPA's Technical Reference Model (TRM) describes three major services that comprise EPA's Enterprise Technology Architecture (ETA) – User Environment, Common Infrastructure, and Applications. The first two services apply to the Technology Architecture. The third service, Applications, provides distinct application types that are used to describe the roles of applications in EPA research and science. These applications are the tools, systems, and other information resources used by the Agency to provide and use the research and science data described in the data architecture through Tools Services.

Grid Computing

Grid computing is the shared use of scientific and research tools where EPA computing resources agency-wide are administered collectively and shared communally. State pilots have been instituted using Grid computing to enable CMAQ optimizations, tool and computational resource sharing and preprocessing of meteorological and emissions input. It is also comprised of coordinated clusters integrated into the scientific Grid while supporting data Grids of scientific and research data-sets.

Scientific Office of the Future

Scientific Office of the Future (soon to be known as Scientific Group Workspaces) consists of high power workstations for scientific collaboration with scientific computational power at the desktop.

"NESC2"

The National Environmental Scientific Computing Center ($NESC^2$) will soon become a part of a new organization called the Center for Environmental Computing (CEC). $NESC^2$ has the following scientific computing assets:

- Computational Resources – high performance computers, clustered computers, and Grid computing used to run applications such as simulations and models
- Data Management Resources – used to manage the large volume of data accessed and created by applications running on scientific computing systems
- Networking Resources – such as Gigabit Ethernet, High Performance Parallel Interface, Fiber Direct Data Interface, and other high performance networking formats

Hi-End Computing

Hi-End (or High Performance) Computing (HPC) combines very powerful computer processors, massive amounts of data, high speed networks, and mathematical models to create visualizations and simulations. In reference to the Research and Science Data Architecture, HPC uses data sets in combination with methods and models, to produce more data sets.

Computer Clusters

Several EPA research and scientific sites have established clusters using existing PC assets to make some level of high performance computing readily available at their own site. This type of solution provides extreme scalability and flexibility and is highly cost effective (lower cost, by far, than specialized high-end platforms like supercomputers). A cluster can be tuned for the unique processing characteristics of a particular model or application. In this way, the technology can be adapted to meet application needs.

GIS: EIMS\NSDI\Geo OneStop and ESRI – ArcGIS

EPA's GIS Program manages, coordinates, and promotes EPA's integrated use of spatial data. GIS is part of the EIMS family where EIMS is used for geospatial data interchange in compliance with the National Spatial Data Infrastructure (NSDI) standards and as a node in the FGDC. ESRI produces ArcGIS as an integrated collection of GIS products to manage enterprise-wide geographic information systems.

Risk: TRIM, THERdbASE

Total Risk Integrated Methodology (TRIM) is a tool suite for assessing environmental risk where its uncertainty and variability features (e.g., sensitivity and Monte Carlo analysis) augment TRIM's capability for performing iterative analyses. There are three individual modules that cover the movement of chemicals (TRIM.FaTE), tracking of human exposure (TRIM.Expo), and estimates of human exposure levels or doses (TRIM.Risk). THERdbASE is a collection of databases and models that are useful in conducting assessments of human exposure to chemical pollutants, especially for inhalation exposure to indoor airborne volatile organic compounds (VOCs).

Simulation: CMAQ, MM5

CMAQ is a modeling system designed to approach air quality as a whole by including state-of-the-science capabilities for modeling air quality issues, including tropospheric ozone, fine particles, toxics, acid deposition, and visibility degradation. Tue Mesoscale Meteorological Model (MM5) is the primary tool for providing meteorological input for the CMAQ simulations. These, and other scientific computing models, will be enhanced to integrate their functions and allow for interoperability. Model interoperability allows the results from one model job to be seamlessly passed as input to another, related, model job.

4.1.3 Workspace Services

Collaboration, with respect to the target RSA, is broadly defined as a means for EPA (internal collaboration) and non-EPA (external collaboration with domestic and international partners and stakeholders) researchers and scientists to share information and tools/applications, access information resources (e.g., online libraries, data and metadata repositories), and freely communicate on subjects that will advance EPA's environmental mission through virtual research communities. Workspace Services support collaboration to help fulfill this mission for EPA with the target RSA research and science capabilities.

Science Portal

The Environmental Science Portal is designed and maintained to simplify access to appropriate Agency science information products and computing resources for EPA researchers and scientists and their collaborators, while making the details of the process invisible to users. As a subset of the Agency's Web Portal, the Environmental Science Portal also provides access to the Agency's Scientific and High Performance Computing information and computing services. It is designed to be customizable per user, and will comprise an array of 'portlet' applications.

Science Subnet: Phase I - FTP Server

The EPA has launched a Science FTP Server (SFS) as Phase I of its Science Subnet effort to meet the needs of the EPA scientists collaborating with outside researchers. The SFS has the capacity to accommodate large multi-Gigabyte file transfers and support both inbound and outbound file transfers from authorized collaborators. A web-based system is used for user setup and authorization. It provides major secure network server connections outside the EPA firewall to improve EPA research and scientific collaboration as well as deliver secure transfer of research and scientific files between research and science collaborators.

Science Subnet: Phase II

The Science Subnet in its first phase provides an FTP facility for transferring the very large data-sets and files outside the EPA firewall that EPA researchers and scientists along with their collaborators generate in their work. Phase II is designed to make Grid Computing available to collaborators as they work with their datasets.

RSA Web Site

EPA has established an intranet site for the purpose of sharing information management and technology news and solutions in the RSA area. It is a subset of the Agency's "CIO Central" intranet site, the EPA's Chief Information Officer Web Site. It is dedicated to showing how the RSA guides, improves, and manages the way EPA's information technology supports the conduct of the Agency's research and science and how it makes it easier to develop successful business cases for investment decision-making.

Standard Technology Solutions Catalog

EPA has articulated its Standard Technology Solutions (STS) vision to inventory, streamline, consolidate, and standardize it research and science technology solutions in order to efficiently reduce the time, financial costs, and overall effort required for technology activities, including that of identifying requirements-based technology solutions. To promulgate these results throughout the RSA, EPA has initiated the development of a Standard Technology Solutions Catalog. It is intended to provide coordination with other EA activities and technology groups with EPA and to keep them abreast of existing and emerging technologies as well as developing and communicating research and science requirements and trends.

4.2 Target RSA Defined Solutions as Distributed Service Components

In a way similar to the target RSA Focus Area situation discussed above, the target RSA Defined Solutions describe what is being made available to service the distinctive and complex information management requirements that the target RSA needs to have fulfilled to provide service to the target RSA BRM Capabilities. As a result, the target RSA Defined Solutions appear to be 'software elements that can be called at run-time with a clear interface and a clear separation between interface and implementation where they are autonomously deployable.'[8] As such, they can likely be viewed as Distributed Service Components at the Solution Level of the target RSA in a Service Component-Based Architecture.

[8] Service Component-Based Architectures, Version 2.0, CIO Council, June 2004, p.4 paraphrased.

5. Target RSA Next Steps

The target RSA is expected to be extended from the Contextual Level down into the details of the Conceptual and Solutions Level over the next year. To do this, it is quite likely that the target RSA will migrate toward a Service Component-Based Architecture.

5.1 Joint Architecture Development (JAD) Sessions

The effort to build out the target RSA is planned to begin with JAD sessions with key stakeholders to press down into the use and utilization issues of the target RSA.

5.1.1 Elaborate the Details of the RSA

The details of the target RSA beyond the Contextual Level will be developed to the appropriate degree of granularity. The participants in the JAD sessions will assist considerably in determining what the effective granularity level is to accomplish the development of a meaningful and usable target RSA.

Conceptual Level

The Conceptual Level will develop the details necessary for the target RSA BRM to map successfully to the OMB FEA BRM and SRM from the Owner's Perspective.

Solution Level

The Solution Level will develop the details necessary for the target RSA BRM to map successfully to the OMB FEA BRM and SRM from the Designer's Perspective. It will also identify how the target RSA Service To Date can be reused and repurposed.

5.1.2 Contribute to Science component of Region 4 EA

Parallel with the JAD sessions, work has begun and will continue to use EPA's Region 4 as a pilot to begin the process of developing Regional EA that reflects the EPA Core EA as well as the sub-domain EAs, including the target RSA.

5.2 Target RSA Sequencing Plan Development

To support the Core EPA EA Sequencing Plan as well as establish the best path for the detail development and utilization of the target RSA, a target RSA Sequencing Plan will also be undertaken to provide guidance to not only the RSA in development, but also to inform and to integrate with the Core EPA EA Program.

This page is intentionally left blank.

Appendix A: FEA and Service Component-Based Architectures[9]

To facilitate efforts to transform the federal government into one that is citizen-centered, results-oriented, and market-based, the Office of Management and Budget (OMB) is developing the Federal Enterprise Architecture (FEA), a business-based framework for government-wide improvement. The FEA is being constructed through a collection of interrelated "reference models" designed to facilitate cross-agency analysis and the identification of duplicative investments, gaps, and opportunities for collaboration within and across federal agencies.

The Components Subcommittee of the Federal CIO Council Architecture and Infrastructure Committee has developed Service Component-Based Architecture guidelines, which compliment specifically OMB's FEA Service Component Reference Model (SRM) by discussing the concepts of service component-based architectures and placing them in the context of the FEA.

Service Components and Service-Oriented Architectures

The pros and cons of component-based architectures have been discussed and debated for some time. With the evolution of web-based technologies and the growing complexity of IT solutions, increased attention is being paid to the benefits of component-based development. Not surprisingly, the successful implementation of a component-based architecture requires the application of sound architecture principles, such as integration, interoperability, adaptability and the management of scale and flexibility.

Service Component Definition

The term "service *component*" can represent many things to many people. It can describe a complete business line such as U.S. Treasury's PAY.GOV, the U.S. Government's Central Contractor Registry (CCR), a business service supporting the validation of a Social Security Number, an application to support Content Management, or a capability that may be accessed through a technology or business interface. With multiple types of components available in industry and across governments, it became critical to the success of the SRM to define "service component" and to clarify the level of granularity that will reside within the SRM. We expand on the notion of component granularity to include return on investment (ROI). A Component is defined as "a self contained business process or service with predetermined functionality that may be exposed through a business or technology interface."

Service Component Levels

The effective identification, assembly, and usage of service components allows for aggregate business services to be shared across agencies and governments. These business services provide the functionality and execution of business processes, which in turn sustain the FEA Business

[9] All of the material in this appendix is directly, but selectively excerpted from Service Component-Based Architectures, Version 2.0, CIO Council, June 2004, with minimal paraphrasing for narrative consistency.

Reference Model (BRM) sub-functions. Service component aggregation will enable rapid building and implementation of components to support a given initiative or investment.

Federated Business Service Components
A set of cooperating system-level components federated to resolve the business need of multiple end users often belonging to different organizations. Can be expressed as an IT 300 exhibit or a federation of IT 300 exhibits.

Business Service Component System
A set of cooperating business components assembled together to deliver a solution to a business problem. Can be expressed as an IT 300 exhibit.

Business Service Components
Represents the implementation of an autonomous business concept, business service, or business process. It consists of all the technology elements (i.e., software, hardware, data) necessary to ex-press, implement, and deploy a given business concept as an autonomous, reusable element of a large information system. It is a unifying concept across the development lifecycle and the distribution tiers. Normally not expressed as an IT 300 exhibit, but as a sub-component of a larger business component system.

Distributed Service Components
The lowest level of component granularity. It is a software element that can be called at run-time with a clear interface and a clear separation between interface and implementation. It is autonomously deployable. Normally not ex-pressed as an IT 300 exhibit. A distributed component provides low ROI for capital planning purposes.

Language Class Service Components
A class in an object-oriented programming language to build distributed components. This is NOT considered an SRM component. Normally not expressed as an IT 300 exhibit. A language class provides very low ROI for capital planning purposes.

Desirable Features of a Service Component

A successful service component-based architecture requires the application of sound architecture principles to the definition and composition of components. The components in the architecture should exhibit basic features:

- Encapsulation
- Consumability
- Extensibility
- Standards-Based
- Industry Best Practices and Patterns
- Well Documented
- Cohesive Set of Services
- Well-Defined and Broadly Available Licensing or SLA

Service-Oriented Architectures

Business improvement and integration challenges are major concerns with the majority of large private businesses and government organizations. Enterprise architects and capital planners must

collaborate to meet the demands of today's business objectives and challenges. To accomplish this, an enterprise architecture is responsible for supporting the development of new business capabilities and must be congruent with business capital planning. An architecture that provides for reuse of existing business services and rapid deployment of new business capabilities based on existing capital assets is often referred to as a service-oriented architecture (SOA). Successful SOAs ensure that the service components and the business services that they de-scribe are consistent and complete across the enterprise. Experience with SOA has shown that reuse is successful when the reuse efforts focus on harvesting the existing capabilities and capital assets within business lines and elements.

This page is intentionally left blank.

Appendix B: Target RSA Document References

- *The Business Reference Model, Version 2.0, A Foundation for Government-wide Improvement*, FEAPMO, June 2003.
- *The Service Component Reference Model (SRM), Version 1.0, A Foundation for Government-wide Improvement,* OMB FEAPMO, June 2003.
- *Service Component-Based Architectures, Version 2.0,* The Federal CIO Council, Architecture and Infrastructure Committee, Components Subcommittee, in collaboration with the OMB FEAPMO and the IAC, June 2004.
- *Federal Enterprise Architecture Program Management Office, Office of Management and Budget, Using the FEA Throughout the Budget Process: Summary and Requirements*, OMB FEAPMO, May 6-7, 2003.
- *FY06 A-11 FEA Additional Instructions & Guidance,* OMB FEAPMO, June 4, 2004
- *The U.S. Environmental Protection Agency, Enterprise Architecture Status Report 2003,* EPA, September 8, 2003.
- *The U.S. Environmental Protection Agency, Enterprise Architecture Status Report 2002,* EPA, September 8, 2002.
- *The U.S. Environmental Protection Agency, Target Architecture, Version 1.0,* Environmental Health Protection Architecture, EPA, December 16, 2002.
- *The U.S. Environmental Protection Agency, Target Architecture, Version 1.0,* Administrative Systems Architecture, EPA, December 16, 2002.
- *The U.S. Environmental Protection Agency, Enterprise Architecture, Version 2.0, Baseline Research and Science Architecture, Main Report EPA,* September 8, 2003.
- *The U.S. Environmental Protection Agency, Enterprise Architecture, Version 2.0, Baseline Research and Science Architecture, Appendix B – Application Inventory,* EPA, September 8, 2003.
- *The U.S. Environmental Protection Agency, Enterprise Architecture, Draft 'Blueprint' Version 0.32, Target Research and Science Architecture,* EPA, June 2, 2004.
- *Enterprise Architecture Reference Models: A Shared Vision for Service-Oriented Architectures,* Tim Bass & Roy Mabry, Draft Version v0.81, for submission to IEEE MILCOM 2004, March 17, 2004
- *Linking Requirements Goal Modeling Techniques to Strategic e-Business Patterns and Best Practice,* Steven Bleistein, Aybuke Aurum, Karl Cox and Pradeep Ray, IEEE, January, 2004.
- *Succeeding with Component-Based Architectures,* IAC, Enterprise Architecture SIG, slide presentation for colloquia, 2004.

- *A Framework for Information Systems Architecture,* John Zachman, IBM Systems Journal, vol. 38, Nos. 2&3, 1999, pp.454-470.
- *The Zachman Framework and the OMG's Model Driven Architecture*, Business Process Trends, September, 2003.

Appendix C: List of Acronyms

Acronym	Acronym Definition/Description
AGN	Access Grid Network
AQUIRE	ECOTOX database for AQUatic toxicity Information REtrieval
ArcGIS	ESRI Geographic Information System
ArcIMS	ESRI Information Management System
AVS	Advanced Visual Systems
BIRN	Biomedical Information Research Network
BMD	Benchmark Dose
BRM	Business Reference Model
CBI	Confidential Business Information
CCR	Central Contractor Registry
CEC	Center for Environmental Computing
CFO	Chief Financial Officer
CIO	Chief Information Officer
CMAQ	Community Multi-Scale Quality
COTS	Commercial Off-The-Shelf
CRADA	Cooperative Research and Development Agreements
CREM	Council on Regulatory Environmental Models
EACC	Enterprise Architecture Coordinating Committee
ECOTOX	ECOTOXicology
EIMS	Environmental Information Management System
EMAP	Environmental Monitoring and Assessment Program
EPA	Environmental Protection Agency
ESC	Executive Steering Committee
ESRI	Environmental Systems Research Institute (a company)
ETA	Enterprise Technology Architecture
FEA	Federal Enterprise Architecture
FEAF	Federal Enterprise Architecture Framework
FEAPMO	Federal Enterprise Architecture Program Management Office
FGDC	Federal Geographic Data Committee

Acronym	Acronym Definition/Description
FTP	File Transfer Protocol
GDI	Geospatial Data Index
GIS	Geographic Information Systems
GPRA	Government Performance and Results Act of 1993
HEDS	Human Exposure Database System
HPC	High Performance Computing
IAC	Industry Advisory Council
IM	Information Management
IRIS	Information Risk Information System
IT	Information Technology
JAD	Joint Architecture Development
MAYA	Alias Products visualization system
MDA	Model Driven Architecture
MM5	Mesoscale Meteorological Model
NESC^2	National Environmental Scientific Computing Center
NSDI	National Spatial Data Infrastructure
NVFEL	National Vehicle and Fuel Emissions Laboratory
OEI	Office of Environmental Information
OMB	Office of Management and Budget
OMG	Object Management Group
OMIS	ORD Management Information System
ORD	Office of Research and Development
PBPK	Physiologically-Based Pharmacokinetic
PHYTOTOX	ECOTOX database for Terrestrial Plant and Agriculture Toxicity
PMO	Program Management Office
QA	Quality Assurance
QIC	Quality and Information Council
QM	Quality Management
QSAR	Quantitative Structure-Activity Relationship
R&D	Research and Development
ROI	Return on Investment
RSA	Research and Science Architecture
S&HPC	Scientific and High Performance Computing

Acronym	Acronym Definition/Description
S&T	Science and Technology
SAS	Statistical Application Systems (a company)
SCBA	Service Component-Based Architectures
SFS	Science FTP Server
SI	Science Inventory
SIG	Special Interest Group
SIM	Science Information Management
SLA	Service Level Agreement
SOA	Service-Oriented Architectures
SRM	Service Component Reference Model
StatLib	Collaborative website hosted by Carnegie Mellon University
STS	Standard Technology Solutions
TAWG	Technology Architecture Work Group
TERRATOX	ECOTOX database for Terrestrial Wildlife Toxicity
THERdbASE	Total Human Exposure Risk database and Advanced Simulation Environment
TRI	Toxic Release Inventory
TRIM	Total Risk Integrated Methodology
TRIM.Expo	TRIM for tracking of human exposure to chemicals
TRIM.FaTE	TRIM for tracking dispersion movement of chemicals
TRIM.Risk	TRIM for estimates of human exposure levels or doses
TRM	Technical Reference Model
UNEP	United Nations Environmental Programme
USGS	United States Geological Survey
VOC	Volatile Organic Compounds
WG	Work Group

This page is intentionally left blank.

Attachment D: 2004 Sequencing Plan

This page is intentionally left blank.

Table of Contents

List of Figures and Tables

This page is intentionally left blank.

Introduction

Over the past three years, the U.S. Environmental Protection Agency (EPA) has been developing its Enterprise Architecture (EA) program consistent with the requirements of the Clinger-Cohen Act of 1996 and in compliance with the Office of Management and Budget's (OMB) directives. This Sequencing Plan describes the major steps taken from the Agency's initial architecture submission in September 2001 to the current activities in Fiscal Year (FY) 2004 and projects the sequence of steps to be taken in coming years.

This is the first year of construction of the EPA Central Services infrastructure and marks the start of a second architectural phase.

In Phase I, EPA laid the foundations for its architecture, convened the necessary design groups, and appointed a Chief Architect. During this phase, several major elements of the current target architecture were built: the National Environmental Information Exchange Network (NEIEN), the Central Data Exchange (CDX), the Facilities Registry System (FRS), and the Substances Registry System (SRS). In addition, the Agency developed baseline and target architectures and deployed Architecture Repository and Tool (ART). EPA began to integrate EA into the budget and other major planning processes. The Agency continued implementation of HRPro to revitalize the financial management and human capital systems. EPA also created a Program Management Office (PMO) to oversee construction of the target Central Services infrastructure to begin the integration of EPA's mission-support information systems.

Phase II started in FY 2004. Initiating construction of the Central Services infrastructure is EPA's main emphasis during this phase. The PMO is building the first version of the EPA Portal with its supporting Identity Management and Access Control system. EPA is also constructing the first Environmental and Health Protection Architecture (EHPA) data mart for a subset of data within the Office of Air and Radiation (OAR) that is particularly useful to researchers. By the end of FY 2005, the new Central Service infrastructure will be proven through the PMO's work with OAR. In addition, the Agency's Grants.gov interface is now operational. Another important initiative is the upgrading and integration of EA governance and planning systems. EPA is putting in place an EA Policy with a set of supporting procedures and standards, and the Office of Environmental Information (OEI) and the Office of the Chief Financial Officer (OCFO) are taking initial steps to integrate the five major EPA planning efforts: strategic planning, budgeting, human capital planning, enterprise architecture, and the capital planning and investment control (CPIC) process.

In Phase III, starting after the culmination of Phase II projects, governance activities will continue to include implementation of the EA Policy, procedures, technical standards and guidelines. Work on the security and data architectures will also continue in this phase, and the EA will be expanded to include an EPA region. In addition, Phase III includes migration of the remaining EPA systems to the Central Services.

This page is intentionally left blank.

1. Phase I: Establishment of Baseline and Target Architectures

Phase I was the start-up phase of the EA. Lasting from January 2001 to the submission of last year's *Enterprise Architecture Status Report 2003* in September 2003, it included creation of the NEIEN, CDX, and initial data registries; all work to organize and establish the EA program; design and review the initial version of the EPA baseline, transition, and target architectures; and creation of the PMO to begin construction of the target architecture.

1.1 Establishing the Architecture Program

EPA's Enterprise Architecture program began in January 2001. Phase I operations established models and modeling procedures, and developed and published EPA's initial baseline, transition, and target architectures. Phase I also included all work between January 2001 and the update of the target architecture in December 2003, set-up of the program, and integration with ongoing architecture-related efforts for improving data collection among EPA's partners (i.e., states, localities, and tribes).

1.1.1 Foundation Architecture-Related Efforts

The architecture program benefited from the work of a number of preexisting committees and information technology development groups in the following specialized areas:

- **The Information Management Work Group (IMWG)**: State environmental agencies, organized through the Environmental Council of States (ECOS) and the EPA, had formed the State/EPA IMWG. Composed of senior leaders from EPA and state environmental agencies, the IMWG had initiated an approach to address joint state-EPA information management issues. Through the IMWG, the states and EPA committed to a partnership to build locally and nationally accessible, cohesive, and coherent environmental information systems. These systems ensure that both the public and regulatory agencies have access to the information they need to document environmental performance, understand environmental conditions, and make sound decisions to ensure environmental protection.

- **The Environmental Data Standards Council (EDSC)**: This group develops environmental data standards to promote the exchange of information among states, Native American tribes, and EPA. The Council identifies those areas of information for which data standards will render the most value in achieving environmental results. The Council involves tribes, other state and federal agencies, business groups, nongovernmental organizations, and other interested parties as the standards are developed. The EDSC was formed by a resolution of the State/EPA IMWG in November 1999. The EDSC has 10 members, including four state, four EPA, and two tribal representatives.

- **The National Environmental Information Exchange Network (NEIEN)**: The IMWG commissioned a Blueprint for a National Environmental Information Exchange Network (also known as the Exchange Network) to serve as the conceptual design of the Exchange Network. The Exchange Network marks the beginning of the transition from the "data push" era to a new "data pull" era in which data resides at its source and can be accessed by those who need it, anywhere, any time. In February 2001, the IMWG chartered an

Interim Steering Group to develop the Exchange Network's implementation. Implementation was completed in early 2002. The IMWG subsequently chartered the Network Steering Board to oversee and guide the Network's implementation.

1.1.2 Defining the Technical Approach to EPA's Architecture

In 2001, EPA modeled the EA program based on the Federal Enterprise Architecture Framework (FEAF). The Agency appointed a Chief Architect and supporting EA Team to direct all architecture development efforts, which was done with extensive cross-Agency involvement and domain lead efforts.

The EA program divided EPA's architecture into three domains. Mission-level programs to protect the environment and human health were incorporated into the Environmental and Health Protection Architecture (EHPA); administrative functions were directed to the Administrative Systems Architecture (ASA); and research and development functions became the Research and Science Architecture (RSA). In addition, the program created cross-cutting component architectures for geospatial services, document management, and security.

1.1.3 Governance Structures

In Phase I, EPA appointed work groups and executive steering committees to guide development of the three major domains. EPA deferred development of the document management architecture and focused its attention on the EHPA. It convened the Enterprise Architecture Coordination Committee (EACC) to oversee and review the development efforts. The Quality and Information Council (QIC), an existing group responsible for the development and governance of EPA's information technology operation, assumed the architecture steering role. The QIC reports its recommendations to the Chief Information Officer (CIO) for approval.

1.1.4 Creation of the Architecture Repository and Tool

In the initial stages of the EA program, the EA Team modeled existing business processes using Popkin's System Architect™ software. In 2002, the EA Team reviewed available architecture modeling tools and selected Metis™, published by Computas, as the COTS software for the EPA Architecture Repository and Tool. In the first project year, EPA customized the Metis metamodel to create the initial version of ART, which is the repository database.

ART was then populated with the EA Baseline Architecture 1.0, as it was submitted to OMB in September 2002. In 2003, EPA expanded the tool to include the EPA Target Architecture 1.0. ART was further enhanced to support EPA's CPIC process for both Form 300 and Form 53 submissions. In September 2003, the Agency provided a read-only version of the ART tool to OMB along with its formal CPIC and EA submissions—the first time a federal EA program had reported its architecture to OMB in electronic form.

1.1.5 Development of EPA's Version 1.0 Baseline Architecture

In 2001, EPA developed an initial baseline architecture by building on existing information technology documentation efforts. The Agency gathered available business process models, network and systems documentation, and data standards information. EPA used this information to prepare a preliminary baseline document, which it submitted to OMB in March 2001.

In 2002, EPA's baseline architecture was greatly expanded through further data calls, business process modeling, data entity definition, and technology architecture documentation. EPA submitted a complete *Enterprise Architecture Baseline Architecture (version 1.0)* to OMB in September 2002, covering all domains and components of the architecture. Further, this version

documented all layers of the EPA architecture: strategic goals, business functions, data, applications, and technology.

1.1.6 Development of the EPA's Version 1.0 Target Architecture

EPA's target architecture has, from the beginning, been strongly influenced by the data integration efforts of preexisting EPA information technology initiatives, the most important being the *Model for Information Integration* (M4I), developed by the Information Integration Program (IIP) and accepted by the Agency in July 2002. The M4I was a technical strategic framework that proposed an integration of data, applications, and technology across the Agency. In the initial December 2002 target architecture submission to OMB, EPA presented the basic concepts of the M4I as the elements of its EHPA target applications architecture. The ASA domain also developed a target applications architecture that was based on upgrading the Agency's administrative systems and integrating them through use of an Enterprise Application Integration (EAI) tool. The Research and Development Architecture (RDA) did not develop a target architecture in 2002.

The other layers of the initial target architecture were more integrated across the EA domains. The Agency developed a Business Reference Model (BRM), a Technical Reference Model (TRM), and Standards Profile to be consistent with the relevant FEA reference models. It also developed its own data model, a taxonomy of generic Agency data entities covering all three domains.

During 2003, EPA expanded and refined the initial target architecture, incorporating a target geospatial architecture and a target technology architecture and developing a Strategic Information Model (SIM). The target security architecture and SIM are further defined below.

1.1.7 The 2003 Strategic Plan

Last year, EPA revised the 2000 Strategic Plan by consolidating the Agency's 10 strategic goals into the five depicted in the 2003 Strategic Plan and Budget structure outlined in the 2003 Strategic Plan.

One of EPA's architecture goals is to align its five major planning processes: strategic planning, budget planning, EA, CPIC, and human capital planning. In Phase I, the Agency aligned the EPA BRM with the activities identified in the new strategic budget architecture. This formed the basis for further integration of Agency planning processes in Phase II.

1.1.8 Initial Integration of Budget and EA Planning Processes

One of EPA's architecture goals is to align its five major planning processes: strategic planning, budget planning, EA, CPIC, and human capital planning. In Phase I, the Agency aligned the EPA BRM with the activity codes used in the budget planning process. This formed the basis for further integration of Agency planning processes in Phase II.

1.1.9 The Strategic Information Model

A SIM is a high level entity-relationship diagram (ERD) that identifies and defines the information required to implement strategies as described in an organization's strategic plan. A SIM consists of entities, attributes, and relationships created in response to information requirements referenced in the Agency's strategic plan. During 2003, EPA created the initial draft of the SIM based on the EPA Strategic Plan of 2000. The EA Team continued to refine it into early 2004. The SIM can serve as the basis of a comprehensive EPA enterprise data model.

1.1.10 Target Security Architecture

The EA security approach includes technologies for identity management, perimeterization, data confidentiality and integrity assurance, data availability, surveillance, audit, and forensics.

The Phase I target technology architecture directions for security include:

- Dynamic, two-factor authentication as the minimum requirement for all data communications with employees and trusted partners
- Encryption for all sensitive messages and files
- Security policies and practices to ensure data confidentiality, data integrity and availability
- Management of user identities, classes, roles, and authorities
- Defined security perimeters, both within and outside EPA
- Capabilities for monitoring, auditing, and incident forensics

Further, EPA's security strategy includes comprehensive identity management, with simplified and reduced sign-on; advanced auditing services; file and message encryption; firewall centralization; perimeterization of Agency information technology resources; and piloting of more secure remote access methods.

1.2 Initial Implementation of the Architecture

Phase I implementation of the architecture was carried out through existing development efforts under such projects as the Exchange Network and the FRS, the central repository of information on facilities and sites of interest to EPA's programs. Phase I also included the inauguration of extensive work to consolidate administrative applications under the ASA domain, primarily through the development of a modernized Financial Replacement System (FinRS). The CIO convened a Program Management Office within OEI to coordinate development of the EHPA target architecture, but work under the PMO did not commence in earnest until Phase II (after September 2003).

The major implementation achievements of the architecture in Phase I are as follows:

1.2.1 CDX and the Exchange Network

The Central Data Exchange is EPA's node on the Exchange Network, developed under the IMWG. The achievements of CDX and the Network during Phase I included:

- **Making CDX Operational at Research Triangle Park:** CDX has been operational at the Agency's National Technical Systems Division (NTSD) at Research Triangle Park since 2000.

- **Establishing a Process for E-Authentication and E-Authentication Portal:** CDX requires electronic authentication of all partners that operate on the Exchange Network. Data exchanges require electronic signatures to validate the identity of parties involved. Authentication became operational in 2001.

- **Implementing State Nodes and Enrolling Users:** All 50 states, about 30 tribes, plus Puerto Rico, Guam, and the District of Columbia became active on the Exchange Network. More than 20 states are now actively building, installing, or testing their own nodes; another 15 states are in the planning stage.

- **Developing Data Standards:** Data standards were developed for the core exchange of information through the EDSC.

- **Establishing the Core Reference Model:** The IMWG established the Core Reference Model (CRM) with the assistance of four states—Michigan, Arizona, Nebraska, and Delaware. The CRM's data elements were gathered by examining all environmental reporting forms used by any office within the four states. In addition, consolidating data elements among states that exchange data with EPA is a valuable aid in metadata reconciliation.

- **Implementing Six Data Flows into CDX:** Six data flows became operational in Phase I. These were the Air Quality System (AQS), eBeaches, FRS, the National Emission Inventory (NEI), the Permit Compliance System (PCS), and the Toxics Release Inventory (TRI).

- **Developing the XML Schema Registry:** Network operations are based on XML schema for exchanging data via Web services protocols. Phase I included development of six XML schemas for the data flows listed above.

Actual data flows achieved during Phase I of CDX are shown on Table D-1.

Production Data Flows	Data Format	Source	Production Date
Safe Drinking Water Access and Review System/Unregulated Contaminant Monitoring Regulation (SDWARS/UCMR)	Web forms, flat files, XML files	Labs, Public Water System, States	October 2001
Air Quality System (AQS), File Transfer	Web forms, Flat files	States, Tribes, Locals	June 2003
eBeaches	Web forms, flat files, XML files	States	August 2003 (Node May 2004)
Facility Registry System (FRS)	XML files	States	September 2003

Table D-1: Phase I CDX Capabilities and Data Flows

1.2.2 Administrative Systems Architecture

During Phase I, the Administrative Systems Architecture work group evaluated EPA's multiple legacy administrative systems. Business units, in conjunction with the ASA work group, launched an initiative to consolidate and modernize its portfolio by:

- Developing a plan to implement the Financial Replacement System (FinRS) that proposes the replacement of legacy financial systems. FinRS includes all of EPA's financial business architecture functions and has been designed to meet OCFO's Information Resources Management (IRM) strategic vision. The components of this architecture, while in varying stages of their application's life cycle, collectively constitute a single Agency-wide source of official financial information easily accessible by authorized customers.

- Continuing to invest in and modify the Integrated Contracts Management System (ICMS) to comply with requirements of the Integrated Acquisition Environment (IAE) E-Government project. ICMS represents the "back office" system in the overall IAE architecture.

- Continuing to invest in the Integrated Grants Management System (IGMS) to support the Grants.gov E-Government project. Short term goals included linking with Grants.gov to accept electronic grant applications.

- Implementing a fully integrated, COTS-based payroll/HR system (PeoplePlus™). PeoplePlus supports the full range of payroll and HR functions, including payroll processing, time and attendance, and labor distribution, without redundant data entry or record keeping. It replaces legacy system, some over 30 years old, and automates several manual processes.

- Beginning implementation of an Enterprise Applications Integration (EAI) system to move administrative data among the various ASA systems. The EAI tools will simplify and standardize the interactions between existing systems and facilitate the migration to new systems that utilize Internet, e-commerce, extranet, and other new technologies.

- Constructing an Administrative Data Warehouse (ADW) structured to maximize performance, usability, data integrity/quality, data integration, and data management. The ADW will deliver a common view of EPA administrative data organized to support analysis, reporting, and EPA's strategic decision-making.

1.2.3 Technology Standards and the IT Roadmap

Technology standards are an essential building block of the Agency's target architecture. EPA's technology standards program is documented in its IT Roadmap, a database that catalogs current, target, interim, and legacy standards. The IT Roadmap was incorporated into the procedures of EA during Phase I. EPA also formalized the Technology Architecture Change Management (TACM) and the Research Agenda processes, whereby new technologies are proposed, evaluated, and integrated into the target Technology Architecture on an as needed basis.

1.2.4 Building Registries

Data registry systems containing uniform information definitions are of common interest to multiple programs and one of the core features of the target EHPA. Four such data registries were already under development and were formally incorporated into the target architecture during Phase I:

- **Facility Registry System**: lists sites and facilities of interest across EPA programs

- **Substances Registry System**: contains uniform definitions of chemicals, wastes, and other substances of scientific and regulatory interest to the Agency

- **Environmental Data Registry:** provides an ISO 11179-compliant listing of metadata definitions for data entities used throughout EPA information technology systems

- **Terminology Reference System:** inaugurated as a single source of environmental terminology, compiled from collections of terms from EPA and other sources.

1.2.5 Development of EDOCKET

EPA is actively involved in 17 of the 25 E-Gov initiatives in support of the President's Management Agenda. EPA is also the designated federal lead for eRulemaking, the cross-government initiative to establish a federal regulatory clearinghouse Web site and a centralized electronic rulemaking system based on EPA Dockets (EDOCKET).

The eRulemaking initiative will include the Regulations.gov Web site, which will enable the public to find and view any open federal rulemaking and submit comments electronically for those rules.

EDOCKET is an online public docket and comment system designed to expand access to documents in EPA's major dockets. Dockets contain Federal Register notices, support documents, and public comments to support regulations the Agency publishes, as well as various non-regulatory activities. Work to develop and expand EDOCKET began in Phase I.

1.2.6 Electronic Records and Document Management System

EPA's Electronic Records and Document Management System (ERDMS) enables employees to capture, manage, and dispose of electronic information in a way that promotes information sharing, collaboration, and managing electronic records in a legally acceptable manner. In Phase I, the Agency launched prototype and proof-of-concept pilots to support the E-Records Management E-Gov initiative.

This page is intentionally left blank.

2. Phase II: Governance, Central Services Infrastructure, Initial Program Office Warehouses, and Target Research & Science Architecture

Phase II is the current phase of EPA's EA and includes all operations since the September 2003 submittal of EPA's *Enterprise Architecture Status Report 2003* to OMB. Its highlights include the development of comprehensive governance policies and procedures to institutionalize EPA's EA, construction of the Central Services infrastructure by the PMO, the development of initial data warehouses under the target applications architecture, and creation of the target Research and Science Architecture. Phase II will end when all Central Services projects are complete.

2.1 Expansion of Governance Processes

Governance during Phase I was centralized under the QIC and its subcommittees, supported by the EACC and the Executive Steering Committees of the domain workgroups. Under Phase II, EPA is extensively expanding the scope and formality of its governance processes.

2.1.1 EA Policy

During Phase II, EPA put in place an Interim EA Policy, signed by the Administrator. The Interim EA Policy establishes the EA program, lays out a framework for EA development and maintenance, and provides high-level roles and responsibilities within the Agency. Specifically, the Policy identifies roles and responsibilities of officers with formal statutory responsibilities under the Clinger-Cohen Act, the Chief Financial Officers Act, and other legislation directly affecting the operation of enterprise architecture in the federal government. The policy has been submitted to the Directives Clearance Process for formal review throughout the Agency.

Under the EA Policy Framework, policies are supported, in hierarchical order, by procedures, technical standards, and guidelines. The CIO has overall responsibility for supplementing the policy with procedures, technical standards, and guidelines. The Chief Technology Officer (CTO) will have specific responsibility for updating procedures and technical standards associated with the technical and security components of the architecture. The Chief Architect will update the various EA guidance documents, often in cooperation with other EA officers or committees.

2.1.2 Procedures

Under the Framework for EA Procedures Development, nine procedures are under development, each establishing formal governance for a major architectural function or process. As illustrated in Figure D-1, these include:

- **Procedures for EA Development and Approval:** these guide the development, implementation, and maintenance of the target architecture.

    1. Change Management of the Baseline Architecture
    2. Change Management of the Target Architecture
    3. Change Management of the Security Architecture
    4. Change Management of the Sequencing Plan

- **Procedures for EA Maintenance and Toolset Usage:** these maintain the technical standards of the policy framework.

5. Change Management of Technology Standards
6. Change Management of EPA Reference Models
7. Change Management of the ART Metamodel

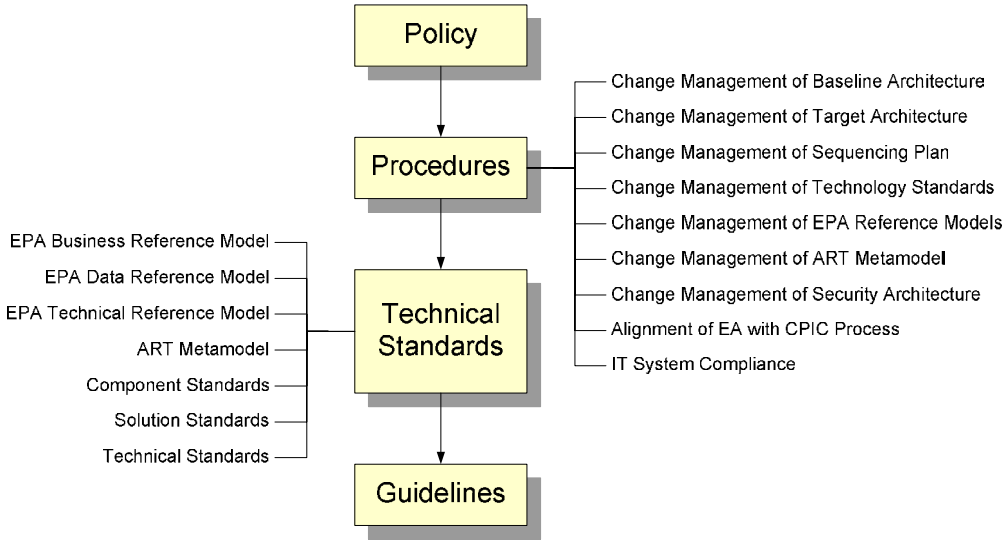

Figure D-1: The Enterprise Architecture Policy Structure

- **Procedures for Enterprise Information Technology Investment Portfolio Compliance:** Interfaces of architecture processes with related information technology capital investment and management and system life cycle processes.

8. Alignment of EA with CPIC Process
9. information technology System Compliance

2.1.3 Further Integration of Planning Functions

In Phase I, the Agency integrated the mapping of the EPA BRM with its budget activities categories. In Phase II, EPA is analyzing strategic business processes to better understand their interconnections and identify opportunities for better alignment with the capital planning and investment control process. These processes involved include strategic planning; budget planning, formulation, and execution; human capital planning and management; financial management; grants management; and contract and acquisitions management. This alignment will allow management to plan for and make better strategic decisions by understanding the relationship between the Agency's strategic processes and how they affect one another.

2.1.4 Development of Solution Standards

As the target architecture evolves, the Chief Architect is preparing the Solution Standards to help prepare program and regional offices to migrate their systems to the target. The heart of the applications architecture is an automated data integration platform, powered by a system of metadata registries, coupled with a system of warehouses and data marts, and accessed through the new EPA Portal. Solution Standards will give progressively more detailed advice to the programs on the design requirements of all these elements, explaining the role of OEI in providing new centralized information technology services and the Programs' responsibilities for connecting to these services in a technically and programmatically compliant manner.

2.1.5 Update of the EPA Reference Models

EPA developed an Agency BRM and TRM to elaborate the internal architecture beyond the detail available from the equivalent FEA reference models. These will be updated to better reflect existing business and technical processes. In addition, the Chief Architect is investigating the potential need for an EPD-specific Service Component Reference Model (SRM) and is reviewing the integration of the Performance Reference Model (PRM) with existing EPA Government Performance and Results Act (GPRA) goals and internal administrative performance measures.

2.1.6 Establishment of the Geospatial Information Officer Position

As planned in the *EPA Geospatial Blueprint* (June 2003), EPA has defined and filled the position of Geospatial Information Officer (GIO). The GIO is accountable for providing increased Agency-wide coordination of Agency geospatial capabilities through a geospatial architecture governance structure. The GIO will communicate the role and potential of geospatial analysis throughout the Agency, provide guidance for the collection and maintenance of geospatial data, improve the quality and quantity of geospatial data for intended uses, create solutions for improving the performance of Agency decision-making through geospatial analyses, promote more effective use of geospatial data across Agency partners, coordinate with the Federal Geographic Data Committee (FGDC) and the Geospatial One-Step E-Gov initiative, and track best practices and emerging technologies from the private sector for adoption within EPA.

2.2 Enhancement of Central Services Infrastructure

A major initiative under Phase II is work by the PMO to begin construction of the Central Services, the eight "portfolios" of common infrastructure services defined in the target applications architecture. The EA initially defined these infrastructure elements as serving only the EHPA, but during Phase II the infrastructure's scope has been generalized to include the operation of the ASA and the Research and Science Architecture (formerly the RDA; now expanded to include science activities outside the Office of Research and Development (ORD)).

Phase II build-out of the target architecture proceeded on two fronts: (1) upgrading the native technical capabilities of individual elements of the Central Services under the PMO, and (2) building out specific projects that validate how these elements will work together to support program needs as depicted in Figure D-2. Some of these projects were carried out jointly by the PMO and the Office of Air; others are underway independently by ORD to build out elements of the RSA, and by OCFO to build out sections of the ASA.

2.2.1 Portal Development

The PMO is currently evaluating three COTS packages for developing the EPA Portal Framework (the hardware and software infrastructure of the Portal) and the various operating Portals within the framework. The selected package will share back-end functionality among Portals, enabling developers to deploy the same "portlet" applications from one to another. The Agency expects to select its preferred software package in September 2004.

Two Portals are currently under development: one to support the Office of Air (Air Analysts' Portal), another to support the RSA domain (Science Portal).

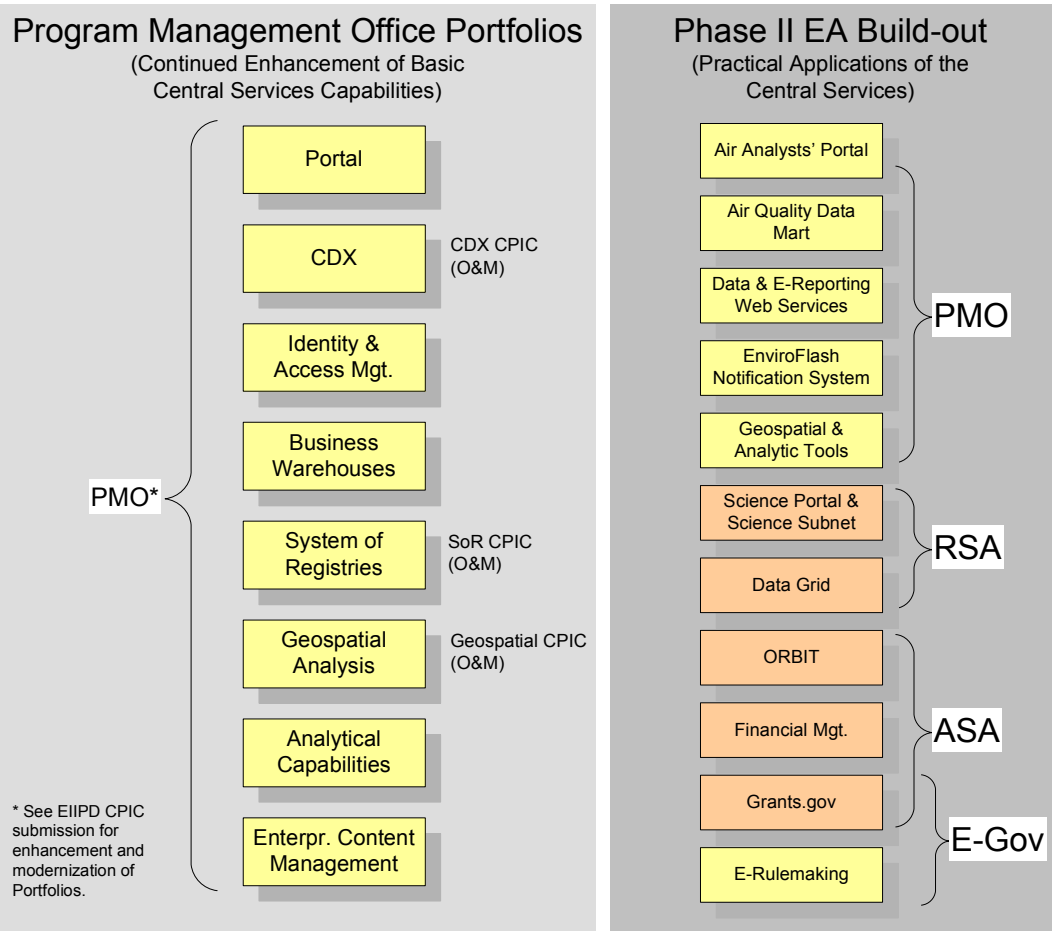

Figure D-2: The PMO Central Services and FY 2004 Build-out Projects

2.2.2 The Central Data Exchange

The Central Data Exchange is EPA's node on the Exchange Network, developed under the Information Management Work Group. The achievements of CDX and the Exchange Network during FY 2004 included the addition of the five new production data flows shown on Table D-2.

Production Data Flows	Data Format	Source	Production Date
Safe Drinking Water Information System (SDWIS)	Flat files	States	February 2004
Institutional Controls (ICTS)	Web forms	Regions	April 2004
National Emission Inventory	Web forms, flat files, XML files	States, Locals	April 2002 Node May 2004
RMP Web Reporting Center	Web forms	Industry	May 2004
Radionuclide NESHAPS (National Emissions Standards for Hazardous Pollutant)	Web forms, flat files	Industry	June 2004

Table D-2: Phase II CDX Capabilities and Data Flows

Node flows, the most highly automated form over exchange via CDX, are now in operation in 13 states. They include flows for the FRS, NEI, e-Discharge Monitoring Reports (e-DMR), and

Beaches. Nodes flows are in development for AQS, the Resource Conservation and Recovery Act Information System (RCRAInfo), and ICTS.

2.2.3 Enterprise Identity Management and Access Control

EPA plans to implement an Enterprise Identity and Access Management infrastructure, based on an enterprise approach to directory services, which the central infrastructure will leverage to provide user management/administration, authentication, and authorization services to a wide range of EPA applications. The Enterprise Identity and Access Management approach will result in improved security, significantly reduced user administration costs, simplified and accelerated application development, and enhanced user experience through self-service and reduced sign-on. The infrastructure will also be designed to support and integrate with the Federal E-Authentication architecture and will be designed to comply with OMB policy and National Institute of Standards and Technology (NIST) guidance.

2.2.4 Framework for Business Area Warehouses

In its update to the target architecture of September 2003, EPA's central Framework for Business Area Warehouses described a variety of technical methods for integrating data flows across various program warehouses and data marts. During Phase II, the technical design of the Central Services' Framework for Business Area Warehouses (FBW) has been defined. The FBW now incorporates the concept of a Data Integration Platform (DIP)[1] as the mechanism for populating data warehouses and data marts, controlling the creation and flow of metadata, and ensuring flexible and responsive data availability. This bus architecture[2] will be simpler, less expensive, and more quickly implemented than the approach envisioned one year ago. The EPA target applications architecture is now even more efficiently oriented than before to support solutions-driven, component-based systems.

The DIP relies on a modern extraction, transformation, and load (ETL) tool to manage incoming data, cleansing and transforming it to system specifications, and transferring it in proper form to the destination system or systems. EPA is expediting the review of available COTS ETL tools for use in creating the DIP for initial program office data warehouse projects. In Phase II, EPA will complete its evaluation of the main COTS products, select an Agency standard, procure the standard tool, install the tool in a development environment as a DIP solution, and conduct a test that populates at least one data warehouse. This system will extend the utility of the Agency's Data Registries by linking them to EPA data warehouses and data marts on a continuing basis.

The procedures and technologies involved in the proposed DIP are sufficiently different from the version 1.0 target that the Chief Architect will treat them as a formal EA version change. Program offices and the PMO will work with the Chief Architect over the next six months to define the solution's details and formulate formal change management documentation for submission to the QIC in November 2004. Interim work to deploy the Central Services will validate this proposed change to the target architecture and work out the necessary technical details for its implementation.

[1] A Data Integration Platform is a middleware strategy for accessing, integrating, and delivering data from any source, to any source. A DIP can utilize varying technologies.

[2] Bus architecture refers to a structural configuration for interconnecting data warehouses, distinguished from the hub and spoke configuration customary in traditional central data warehousing environments.

An important aspect of the refined applications architecture is that it allows systems to migrate to the target one component at a time.[3]

2.2.5 Enterprise Content Management

The E-Records Management E-Gov initiative that began in Phase I continued in Phase II, renamed the Enterprise Content Management System (ECMS). The initiative uses the ERDMS system as the foundation for a new approach to content management at the enterprise level, based on the Documentum™ COTS package.

2.3 Central Services Development of Program Office Projects

The PMO plans to prove the Central Services data infrastructure with its OAR partnership by the end of calendar year 2005. In parallel with the construction of the major new Central Services elements (e.g., CDX and the major registries are already in operation), the PMO is planning to build two operational data warehouses to support existing program operations.

2.3.1 Clean Up Architecture

The Office of Solid Waste and Emergency Response (OSWER) Clean Up Architecture is a major EPA initiative to integrate management of land-based cleanup efforts: Superfund, Brownfields, Underground Storage Tanks, and Hazardous Waste Management. The four OSWER programs, although authorized under different statutes with different program priorities, share many common data elements and technical implementation approaches. These programs form an excellent opportunity for a consolidated information technology architecture. The current OSWER effort will operate under a common data model and emphasize a component-based, distributed applications environment.

2.3.2 Office of Air Projects

The PMO is currently supporting integration of a subset of OAR data with the Central Services. Plans are under way to define the specifications and data flows to be integrated in the first data mart.

The initial data marts will validate the revised target applications architecture to be submitted to the QIC for ratification in the first quarter of FY 2005. The rapid deployment timeframe will also provide the necessary cost and resource reference for projecting the migration costs and schedules for integrating all the other major EPA program systems into the Central Services in Phase III.

2.4 Target Research and Science Architecture

In Phase II, the Research & Development Architecture was expanded beyond the ORD to include all science activities across the Agency. It was accordingly renamed the Research and Science Architecture and is currently completing its target architecture, building on the RDA baseline completed in September 2002.

[3] Under the defined applications architecture, migration of systems can be incremental, allowing systems to move forward independently of each other, rather than in a series of "waves," in which systems would be queued in groups to migrate to the target, as originally proposed in the *Enterprise Architecture Status Report 2003*.

The RSA target accommodates the technical needs of working scientists as well as the access and dissemination needs of consumers and producers of scientific information, inside and outside the Agency. Key elements of the RSA target include:

- **Scientific Information Management:** Under the target RSA, scientific information will be managed efficiently, transparently, and holistically. The RSA community must cope with a geometrically-expanding proliferation of data, gathered and developed by collaborating scientists within the Agency and across the country, but regulatory-based science demands transparent and reliable documentation and access. These different approaches will be integrated into a single architecture.

- **Quality Management:** Under the target RSA, the quality of scientific data—the foundation of science—will be more efficiently controlled and documented. Improving data quality is not a new problem, but it is central to the RSA: regulatory-oriented science activities must be planned, managed, and documented appropriately to provide products of known quality. This issue relates to Scientific Information Management, but it focuses less on systems for data management than on explicit RSA support for the EPA Quality Program's policies, procedures, and standards.

- **Scientific and High Performance Computing:** The target RSA will make high-performance computing available for many more uses than it is today. Many of EPA's concerns are amenable to so-called *in silico* approaches—those that rely relying on environmental simulations and computational modeling. A host of analyses that were previously done on the bench can now be done on computers, often more appropriately and cheaply. The goal of the RSA is to make the most of costly high-end computing resources, efficiently allocating access to high-end machines and networks and assigning tasks appropriately to different classes of computing

- **Geospatial Analysis:** The RSA envisions the complete integration of geospatial information into the analysis and display of environmental information through the use of better tools and high speed networks. Ready and efficient access to an array of many GIS tools and their associated data is being pulled together through Web service portlets in the Science Portal (under development), serviced by exceptionally high-speed networks (i.e., increased bandwidth) for rapid exchange of the extremely large data sets that GIS generates.

- **Collaboration:** The EPA Science Portal will be the scientific community's central point of access to environmental information and EPA's computing resources. It will serve the entire RSA community, including researchers, scientists, laboratory staff, managers, support contractors, international partners and stakeholders, states, tribes, private industry, academia, and external contractors and consultants.

  In addition, The Science Subnet ties the Office of the Future to other systems. It will be a secure network environment that moves the Agency toward creation of virtual "Collaboratories". The Science Subnet will provide scientists with freer access to tools, data, and each other while protecting secure information systems at their current level

- **Standard Technology Solutions:** The target RSA will make it easier for scientists to get access to the technical tools they need. Developing a Standard Technology Solutions (STS) catalog will enable the RSA community to efficiently find technology solutions that satisfy their requirements, or to determine when new solution must be sought. This approach maximizes the potential for reuse and collaboration, and directly supports the goals of solution-oriented architecture.

2.5 Updates to the Technology Architecture

During Phase II, the technology architecture is continuing along a number of fronts to meet the needs of the PMO, RSA, and ASA, as well as preparing for significant upgrades under Phase III.

- The IT Roadmap—the Agency's inventory of technology products and standards—is being updated on a frequent basis. Priority is being given to those standards and target technologies that are most critical to the needs of the three architecture domains (EHPA, ASA, and RSA) and to the construction of the Central Services by the PMO. The IT Roadmap will be extended with the addition of 26 new technology types and corresponding products based on the technology requirements of the EA business domains, PMO Central Services, and results of Research Agenda projects. New standards will be reflected in both the IT Roadmap and the Agency's standards profile.

- As soon as the target RSA is available, the technology architecture will begin to integrate the RSA's technology needs into a draft, revised Technology Architecture Sequencing Plan (TASP). The RSA is evaluating a number of technology solutions that could impact the TASP significantly, including:

  – A scientific information management system to address the need for improved metadata capture and maintenance. Bench scientists and researchers will use this system to record and document raw and published data, as well as to feed the data to scientific applications for analysis. In addition, this data could be used by other scientists inside and outside of EPA, decision-makers, other Agencies and collaborations, and the public. The system will reflect the Scientific Information Life Cycle, including quality assurance and peer review.

  – A Science Subnet to transfer large data sets across secure networks without being automatically blocked as apparent security violations. Using a dedicated FTP server located between two firewalls, this will greatly smooth collaboration and data exchange with research and scientific partners both inside and outside the Agency.

  – A Science Portal to provide one-stop access to EPA scientific data and computing resources and services.

  – A scientific "Office of the Future" that uses standard tools to support collaboration among EPA scientists and trusted partners. Enabling this collaboration will be video conferencing capabilities on each desktop, as well as emerging visual meeting technologies such as Webex. The new desktop uses multiple computer configurations based on job needs: a standard machine for staff conducting administrative tasks and a high-performance machine for scientific analysis, metadata capture, and visualization.

  – Integration of RSA strategies for scientific and high performance computing, including distributed computing that networks unused computing capacity for high-volume tasks, plus access to other supercomputing equipment, and management and archiving of data sets.

  – An Environmental Science Grid, providing the infrastructure to enable EPA scientists, states, industry and other researchers requiring high-capacity computing power to run models, visualization, and other decision-making tools. The Grid Pilot has been completed. Information resources and tools would be provided by the Science Portal. Initiatives like computational toxicology would use this shared computing power, while the Science Subnet would post and share large data sets and

research publications. A pilot of this grid proposal will be launched via the New York Air Quality Pilot.

- The revised TASP will also be updated with the status and sequencing of infrastructure projects and initiatives within each of the Technical Reference Model service areas. The projects represented in the TASP are enterprise-wide technology initiatives that will contribute to achieving the Agency's target Technology Architecture.

- The Technology Architecture team is assessing the EPA Technical Reference Model to align it better to the FEA TRM. This includes expanding some emerging technology categories to match those available in the FEA TRM. The technology architecture is also beginning to cross-walk elements of the IT Roadmap and standards profile to the EPA TRM in preparation for a full alignment under Phase III.

This page is intentionally left blank.

3. Phase III: Programmatic Solution Architecture Development; Migration of Major EPA Systems to the Central Services

Phase III of the EA architecture program will begin when the Phase II portal development projects and the initial program office data warehousing projects are complete. The projects will provide essential technical information for the development of the final sequencing plan to migrate the remaining EPA systems to the target environment. Individual plans for sequencing the Agency's major systems will be detailed beginning with the FY 2007 CPIC Proposals, and further detailed in subsequent annual CPIC submissions.

3.1 Planning Systems Integration

Work in Phase II to create a unified governance system for implementing all levels of EA—from the development of the target architecture to the implementation of the architecture through System Life Cycle planning at the systems level—will continue in Phase III.

3.2 Development of Component and Technical Standards

Component standards provide guidance for the preparation of sub-architectures at the program office and regional levels. Due for initial publication early in early 2005, the guidance will include modeling protocols below the enterprise level, with procedures for exporting high-level data of more detailed program office models to the enterprise level of ART. It will provide modeling guidance for all architecture elements at the program office or regional level, including modeling of data flows, business processes, applications structures, and technology infrastructure. It will describe how these groups will derive necessary architectural data needs through a review of program goals and objectives, as well as projected program or regional uses of architecture principles and processes. Finally, it will provide guidance on how to incorporate e-Gov and other cross-federal architectures into appropriate sub-architectural components.

Technical standards provide detailed guidance at the systems level for connecting components to each other and to the Central Services. It will be produced and updated over time by the PMO, initially drawing on the experience of the OAR warehousing projects of Phase II and the implementation of the DIP. Initial technical standards will be available in calendar year 2005.

3.3 Implementation of the Target Research and Science Architecture

In Phase III, the target RSA will begin implementation of the key elements of its strategy:

- Creation of the Scientific Information Management System (SIMS), incorporating the electronic notebook, science desktop, workflow management, and Laboratory Information Management System (LIMS) elements into a coordinated whole.
- Implementation of a new high-performance computing strategy, including the use of grid computing.

- Implementation of the science portal as a sub-area of the Agency Portal, including effective, efficient, and secure scientific collaboration tools as well as providing a common gateway for all users of EPA science information and products.

- Coordinated management of, and access via the Portal to, scientific information as referenced by EIMS.

3.4 Further Expansion of the Administrative Systems Architecture

In Phase II, the scope of the ASA will include regional and program office administrative functions and associated data and applications. This information will serve as the foundation for a complete EPD-wide baseline ASA and the development of the revised target ASA.

3.5 Further Expansion of the Technology Architecture

In Phase III, the Technology Architecture will continue its ongoing activities to update the IT Roadmap and Standards Profile, incorporating any new technology needs developed by the EHPA, ASA, or RSA domains. In addition, it will:

- Update the baseline technology architecture, first developed in 2002, to include a more current description of all Agency technology. This ongoing activity will coordinate closely with the development of the PMO's Central Services infrastructure, so that systems migrating to the target architecture over the next several years will have clear and current information on the status and design of existing EPA host computers, networks, security systems, telecommunications, desktops, and other infrastructure. The baseline will be incorporated into ART, with suitable data security to make the information available on a need-to-know basis to Agency personnel and support contractors.

- Implement governance systems to ensure the currency and accountability of all technology architecture artifacts, including the baseline and target technology architectures, TRM, the Technology Architecture Sequencing Plan, and the IT Roadmap and Standards Profile.

3.6 Migration of Major Systems

During Phase II, the working assumption was that once a system began to migrate from current state to target state—with data flowing through all components of the Central Services—that it migrate all at once. It would simultaneously flow all data through CDX, hook to all applicable data registries, build any necessary data marts to service all storage and query needs, and provide access to users through the EPA Portal.

The proposed version 2.0 target architecture will, when approved, relax the assumption of simultaneous modernization—and simultaneous modernization funding—of all system elements. Migration of systems can be incremental, allowing internal components of systems to move forward on their own schedules, as dictated by system life cycle logic, rather than as a series of "waves," in which systems would be queued to migrate as a whole to the target, as originally proposed in the *Enterprise Architecture Status Report 2003*. EPA's Version 1.0.

At present, it is possible to project the following milestones for the most established element of the Central Services, CDX. Table D-3 shows currently projected CDX data capabilities and flows for the beginning of Phase III.

Production Data Flows	Data Format	Source	Production Date
RCRAInfo XML for Node Exchanges	XML files	Industry	October 2004
RCRA Part A Permit Applications	Web forms	Industry	TBD
RCRA Site ID	Web forms	Industry	TBD
SDWIS XML Schema	XML files	States	September 2004
STORET Export	Flat files	States, Tribes, Locals	TBD
WCIT (Water Contaminants Information Tool)	Controlled Access	N/A	December 2004

Table D-3: Phase III CDX Capabilities and Data Flows

In addition, EPA projects the following CDX users will be exchanging data through the EPA node by the close of FY 2005:

- At least 24 and up to 40 states
- 15,000 to 30,000 company representatives
- 3,500 to 6,000 public water suppliers/laboratories
- 100 to 300 tribal and other local governments

In addition, CDX will include at least 12 implemented data exchanges, including one exchange to the E-Authentication Gateway and at least six new text data exchanges.

As systems migrate, they will incrementally add and delete functions and connections to elements of the infrastructure and/or to legacy systems that are in the processes of being "sunsetted." For example, some of the key milestones on the projected Integrated Compliance Information System (ICIS) development timeline are shown in Table D-4.

Milestone	Production Date
Implementation for NPDES (PCS) (v1.0)	12/31/2005
Turn off PCS	6/30/2007
Public access to ICIS NPDES Data (ECHO)	6/30/2007
Public Access to ICIS Air Data	9/30/2008
Turn off AFS	9/30/2008
Turn off IDEA	9/30/2009
Public Access to ICIS Toxics & Pesticides Data	3/31/2008
Turn off Legacy Toxics & Pesticides Systems	3/31/2008

Table D-4: Milestones for the Migration of ICIS to the Target Architecture

3.6.1 Evaluation of OAR Projects

At the completion of the initial data marts for the OAR projects, the PMO and the Chief Architect will review the results and define any consequent changes to the proposed version 2.0 target architecture. Change proposals will be submitted through prescribed channels of review and final approval.

3.6.2 Development of Technical Standards

The most detailed guidance for program office use in migrating applications to the target will come from the PMO, not the Chief Architect. The PMO and the portfolio managers will, on the basis of results from Phase II projects, define all necessary technical standards for publication to the program offices and regions in preparing their migration plans and CPIC proposals for FY 2007 and beyond. Technical standards will be product-specific and will define any interpretations of technology standards necessary to ensure consistent alignment of program systems to the Central Services.

3.6.3 Evolution of the Sequencing Plan

The sequence for major information technology investments applications migrating to the target architecture will take shape in the FY 2007 budget planning cycle. Major information technology investments undergoing development, modernization, or enhancement will document a transition schedule to the target architecture in their FY 2007 Exhibit 300 submissions. The Agency will also identify a subset of "CPIC Lite" investments (i.e., between $250,000 and $3 million) for which similar plans and documentation will be prepared in the FY 2008 budget submission.

Applications will migrate according to their own internal needs, connecting to one or more components of the Central Services (i.e., Portal, CDX, system of registries, and Framework for Business Warehouses), but not necessarily simultaneously. Instead, their investment proposals will support incremental migration to the target architecture in keeping with appropriate internal modernization schedules.

In addition, EPA's future Sequencing Plan will reflect the Agency's growing role in E-Gov, such as the hosting of cross-federal solutions (e.g., eRulemaking). Technical specifications and scheduling will emerge as these projects are developed further.

Attachment E: Portal Development Illustrations

This page is intentionally left blank.

Portal Development Illustrations

In December 2003, the PMO conducted a portal Operational Capabilities Demonstration (OCD) prior to developing the EPA portal. The goals of the OCD were to:

- Identify potential user groups of the EIIPD Portal;
- Determine sample functionality for each user group;
- Demonstrate the technical feasibility of a Portal solution within EPA's target architecture; and
- Obtain approval and solicit feedback for the EPA Portal project from senior management.

The attached figures are portal screen mock-ups that were generated during the portal OCD development process. These mock-ups are for illustrative purposes only. Actual portal screens will be developed in conjunction with requirements sessions with users for specific EPA portal functionality, such as the Office of Air and Radiation's Air Analysts' Portal.

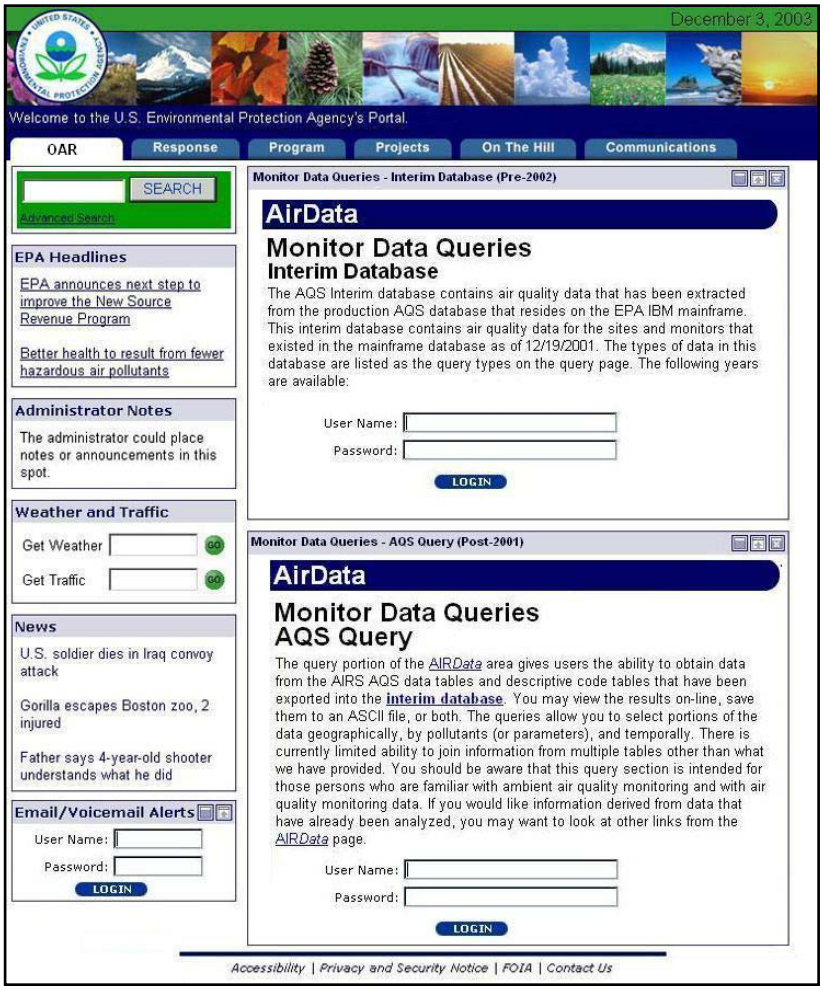

Figure E-1: EPA Portal Demonstration Screen – EPA Portal Common Look and Feel and Integrated Interface

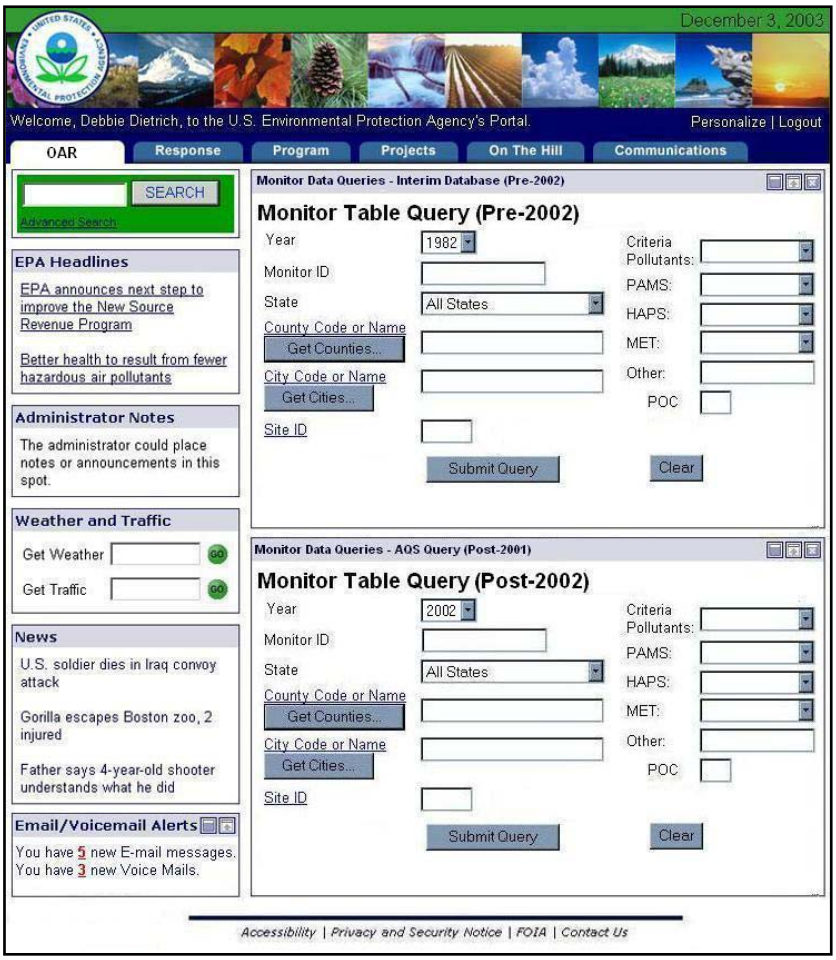

Figure E-2: EPA Portal Demonstration Screen – Single Sign-on, User Page Customization, Common Security Model

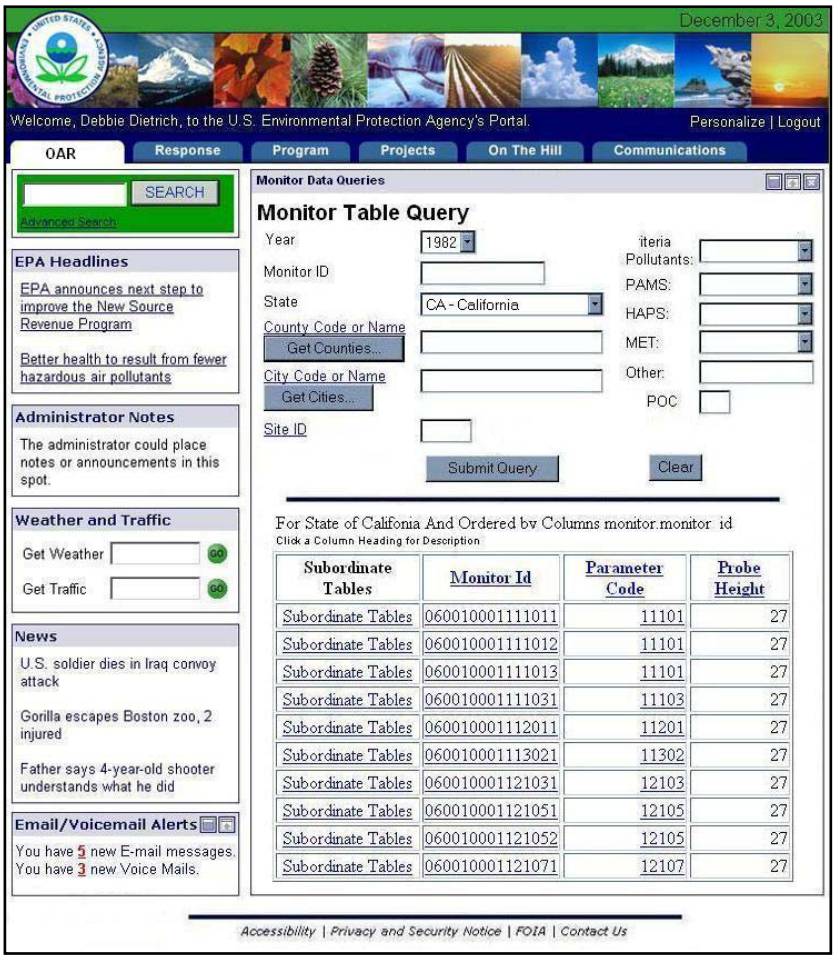

Figure E-3: EPA Portal Demonstration Screen – Data Assets Combined into Framework for Business Area Warehouses

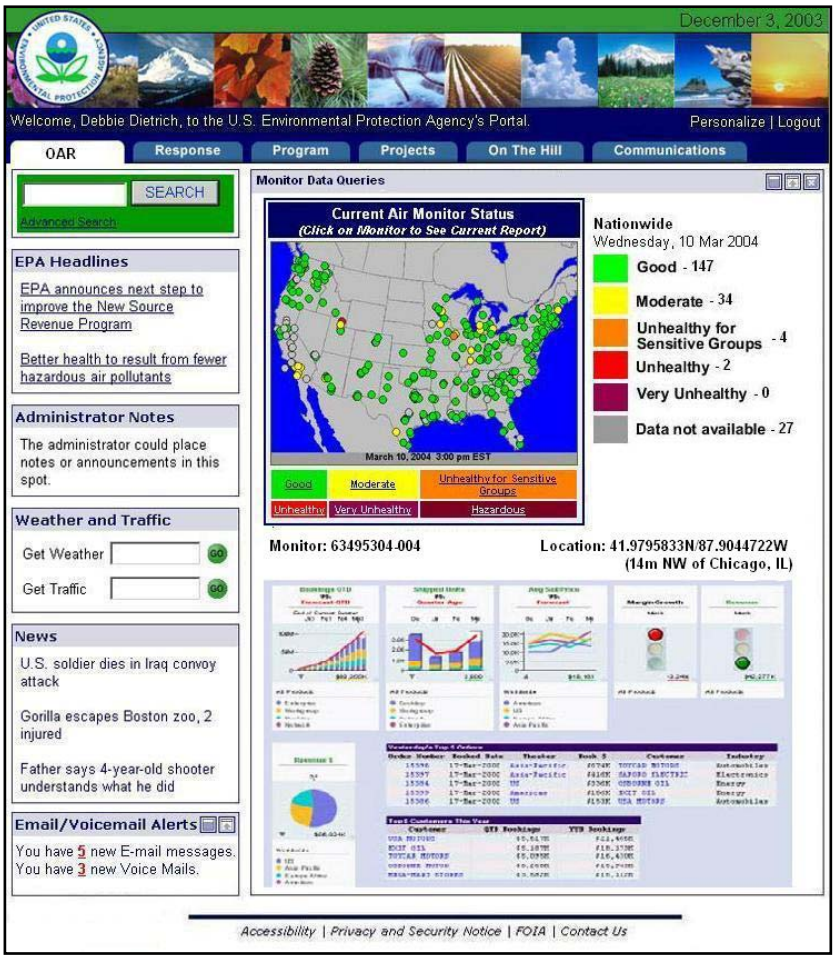

Figure E-4: EPA Portal Demonstration Screen – Integrated Geospatial Tools, Dashboard Tools

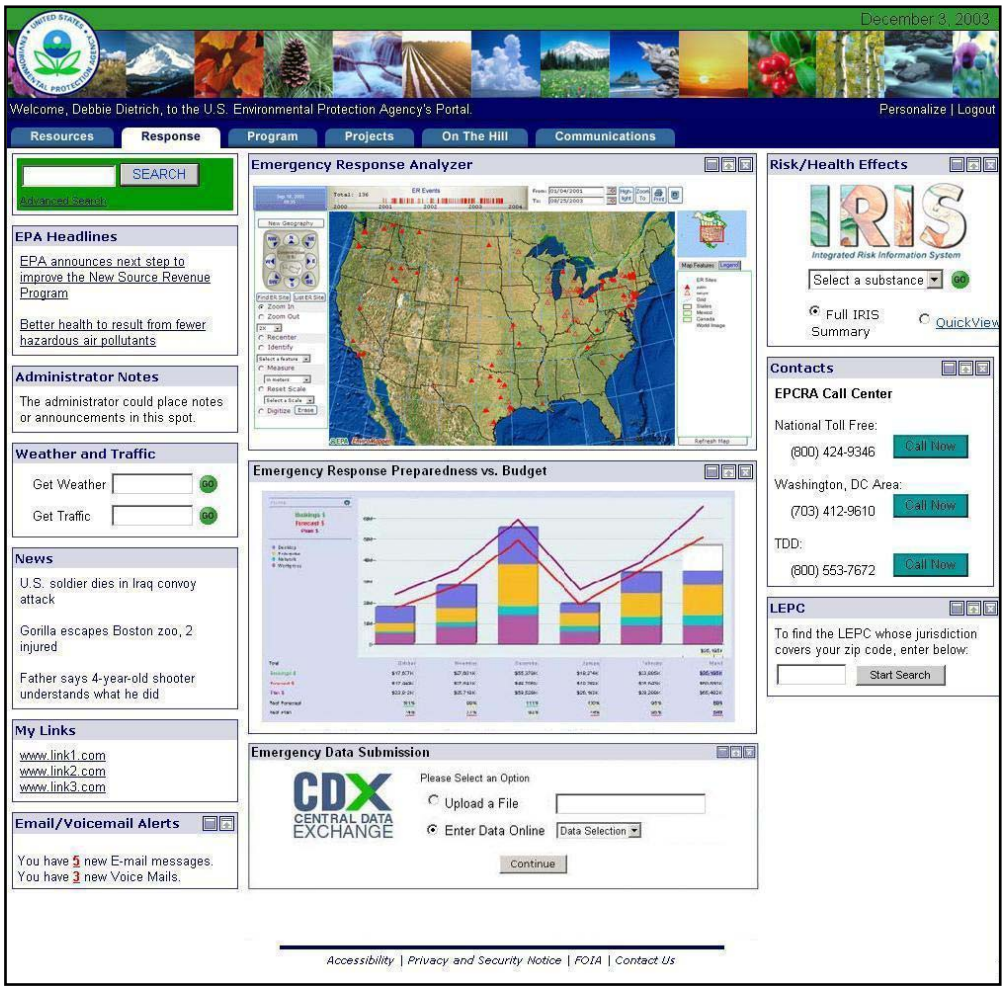

Figure E-5: EPA Portal Demonstration Screen – One Stop for Exchanging Data, Integrated Access to Data Sources and Applications, Dynamic and Static Reporting Capabilities, Meets E-Gov/eSignature Requirements, Secured Access

This page is intentionally left blank.

Attachment F: EA Maturity Assessments

This page is intentionally left blank.

Government Accountability Office

GAO first rated agencies in 2001, when EPA self-scored itself at 3 out of a possible 5 on Version 1.0 of the GAO scoring criteria. After changing its criteria (Version 1.1), GAO dropped EPA to a 1 in 2003.

A comparison of the two GAO maturity measurements is shown on Table F-1, along with EPA's current self-assessment for both versions of the GAO criteria. EPA now rates its EA program as maturity Stage 4 on GAO Version 1.0 and Stage 3 on Version 1.1:

- **Self-assessment using GAO Version 1.0 Criteria:** The Agency now has a baseline architecture ("as-is"), target architecture ("to-be"), and sequencing plan, which achieves Stage 4.
- **Self-assessment using GAO Version 1.1 Criteria:** With the expansion of its core EA team size and with a reassessment of support resources in programs throughout the Agency, EPA's EA resources are adequate to achieve Stage 2. In addition, EPA has a written and approved organization policy (the Interim EA Policy). This achieves Stage 3.

In FY 2005, EPA will approve detailed procedures for EA maintenance (formal management of EA change), create metrics for measuring EA benefits (measure and report EA return on investment), and institute EA compliance measurement and reporting, thus achieving Stage 4 under Version 1.1.

GAO Management Maturity Framework Version 1.0

Achieved 2004
Expected 2005
Expected 2006

Stage	Element	Satisfied 2001	2003 GAO Re-assessment	2004 (Self Assessed)
Stage 3: Developing architecture products	Written/approved policy exists for EA development.	Yes	Yes	Yes
	EA products are under configuration management.	Yes	Yes	Yes
	EA products described or will describe enterprise's business--and the data, applications, and technology that support it.	Yes	Yes	Yes
	EA products desribe or will describe "as is" environment, "to be" environment, and sequencing plan.	Yes	Yes	Yes
	EA scope is enterprise-focused	Yes	Yes	Yes
Stage 4: Completing architecture products	**Written/approved policy exists for information technology investment compliance with EA.**	**Yes**	**No**	**Yes**
	EA products are under configuration management.		Yes	
	EA products described enterprise's business--and the data, applications, and technology that support it.	Yes	Yes	Yes
	EA products describe "as is" environment, "to be" environment, and sequencing plan.	**No**	**No**	**Yes**
	EA scope is enterprise-focused.	Yes	Yes	Yes
Stage 5: Leveraging the EA for managing change	Written/approved policy exists for EA maintenance.	Yes	No	2005
	Either EA steering committee, investment review board, or agency head has approved EA.	Yes	Yes	Yes
	Metrics exist for measuring EA benefits.	No	No	2005

GAO Management Maturity Framework Version 1.1

Stage	Element	Satisfied 2003	2004 (Self Assessed)
Stage 2: Building the EA management fountation.	**Adequate resources exist.**	**No**	**Yes**
	Committee or group representing the enterprise is reposnible for directing, overseeing, or approving EA.	Yes	Yes
	Program office responsible for EA development and maintenance exists.	Yes	Yes
	etc.	Yes	Yes
	etc.	Yes	Yes
Stage 3: Developing architecture products	Written/approved organization policy exists for EA development.	Yes	Yes
	EA products are under configuration management.	Yes	Yes
	EA products describe or will describe both "as-is" and "to-be" environments, as well as a sequencing plan.	Yes	Yes
	Both "as-is" and "to-be" environments are described or will be described in terms given in Stage 2.	Yes	Yes
	These descriptions address or will address security	Yes	Yes
	Progress against EA plans is measured and reported.	Yes	Yes
Stage 4: Completing architecture products	**Written/approved policy exists for EA maintenance.**	**No**	**Yes**
	EA products and management processes undergo independent verification and validation.	No	Yes
	EA products describe both "as-is" and "to-be" environments, as well as a sequencing plan.	**No**	**Yes**
	Both "as-is" and "to-be" environments are described in terms given in Stage 2.	Yes	Yes
	These descriptions address security.	Yes	Yes
	Organization CIO has approved current version of EA.	Yes	Yes
	Committee or group representing the enterprise or the investment review board has approved current version of EA.	Yes	Yes
	Quality of EA products is measured and reported.	**No**	**2005**
	Both "as-is" and "to-be" environments are described or will be described in terms given in Stage 2.	Yes	Yes
Stage 5: Leveraging the EA for managing change	Written and approved organization policy exists for IT investment compliance with EA.	Yes	Yes
	Process exists to formally manage EA change.	**No**	**2005**
	EA is integral component of IT investment management process.	Yes	Yes
	EA products are periodically updated.	Yes	Yes
	IT investments comply with EA.	Yes	Yes
	Organization head has approved current version of EA.	**No**	**2005**
	Return on EA investment is measured and reported.	**No**	**2006**
	Compliance with EA is measured and reported.	**No**	**2005**

Table F-1: Comparison of GAO Rankings and Current and Projected Self-Assessment

The OMB Enterprise Architecture Maturity Model: In June 2004, OMB issued the Enterprise Architecture Assessment Framework (EAAF) Version 1.0. Using this set of criteria—quite different from GAO's—OMB rated EPA at 3.13.

Over the coming year, EPA will institute improvements to its program to raise its score to at least 4.0 on the OMB scale, as illustrated in Table F-2.

Assessment Category	Actions	Current Score	Projected FY 2005 Score	Key FY 2005 EA Enabling Activities
CHANGE	Facilitating and managing change to any aspect of the enterprise			
	A. Architectural Approach	3	4	Key the sequencing plan to capital investments
	B. Strategic Direction	3	4	Implement a comprehensive communications plan
INTEGRATION	Realizing the business rules are consistent across the organization, the data and its use are certain, interfaces and information flow are standardized, and the connectivity and interoperability are managed across the enterprise			
	A. Interoperability	3	4	
	B. Data	4	5	Implement a comprehensive data strategy
	C. Business Logic	3	3	Move to pattern-driven business architecture
	D. Interface	4	4	Integrate performance measures across applications programs
CONVERGENCE	Striving toward a standard IT product portfolio as contained in the Technical Reference Model (TRM)			
	A. Components	4	5	
	B. Technical Platform	4	5	Create the ability to examine redundance across products
	C. Performance	3	4	Link performance measures to technical and service layers
	D. Security	3	4	Install security stds at all levels of components/services/tech.
BUSINESS ALIGNMENT	Ensuring the practices of the enterprise are aligned with strategic management intent			
	A. Strategic Goals	3	4	Improve resource allocations through performance objectives
	B. Business Target	2	3	Link the EPA business vision to technology architecture

Key	
0	No evidence presented
1	EA is initial, informal, and ad-hoc
2	Formal but basic, follows some best practices
3	EA is beginning to be operationalized across the enterprise (i.e. part of transition, CPIC, budget)
4	EA is operationalized and provides performance impact to business operations
5	IT planning is optimized through the EA

Table F-2: Current and Projected EPA EAAF Scores

This page is intentionally left blank.