ERA System Design Review

Day Two

May 10, 2005

ERA SDR – DAY TWO

Data Model

[redacted]

May 10, 2005

Data Model – Agenda

Methodology

Notation

Conceptual Model

Logical Model

Data Replication

Data Model – Methodology

Data model addresses persisted data that is global to ERA

Inputs from ERA Domain model and Service design

Hierarchy of modeling steps, each adding more detail

- – Conceptual model
  - High level data specification from the business perspective
  - Described in the data architecture section
- – Logical Model
  - Detailed data specification from the system design perspective
  - Partially described in the SADD data design section, refined at PDR / CDR
- – Physical Model
  - Specification in terms of implementation (relational database, XML schema, etc)
- – Implementation
  - Actual physical implementation of the data model

Data Model – Methodology

[redacted]

Data Model – Methodology – Conceptual Model

[redacted]

Data Model – Methodology – Logical Model

[redacted]

Data Model – Methodology – Physical Model

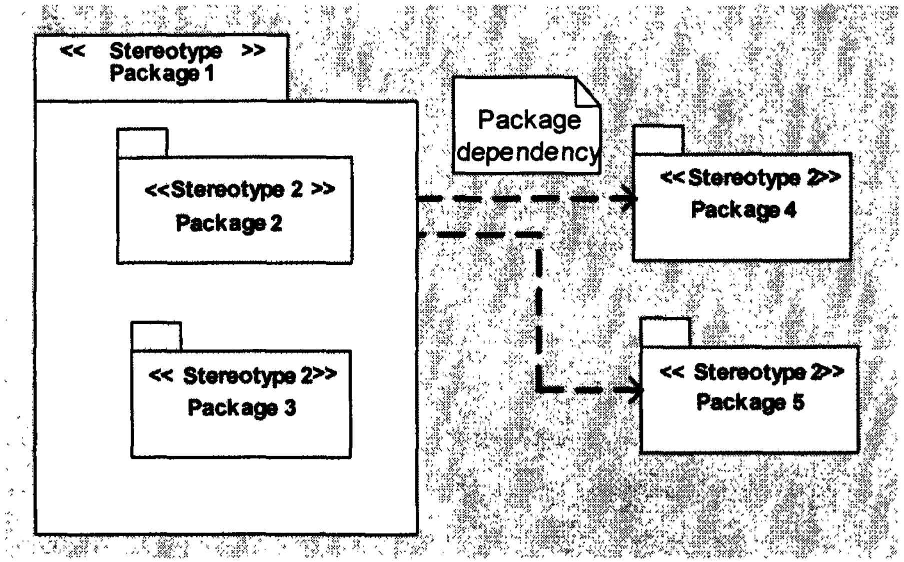
[redacted]

Data Model – Notation

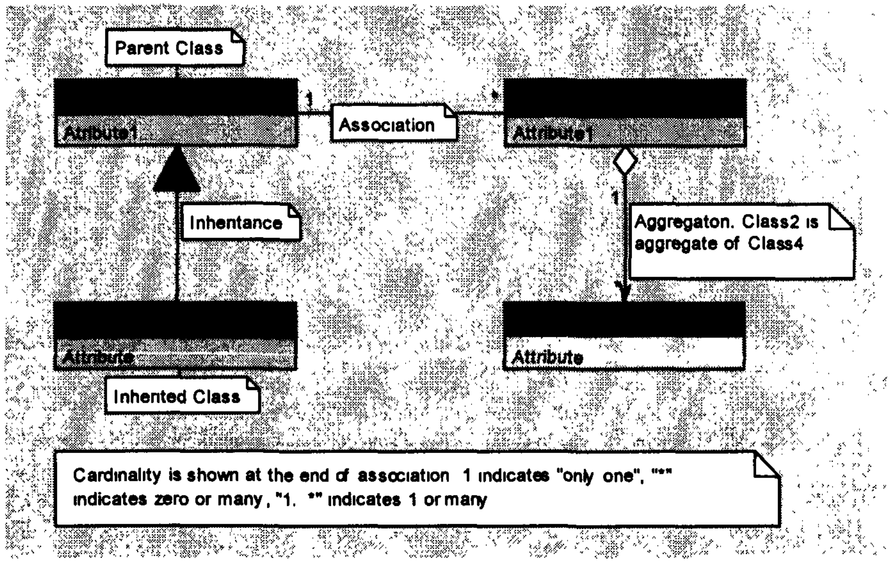
UML Notation used throughout the data architecture and design

- Suited to represent hierarchical nature of the data model
- Packages used to represent groups of objects and NARA constructs
- Packages are stereotyped to identify their hierarchical level
  - A stereotype is a way to extend the core semantics of UML to express new concepts.

Package diagram



Class diagram

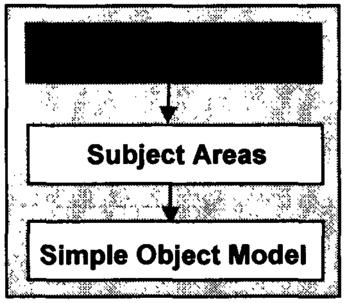


Aggregaton. Class2 is aggregate of Class4

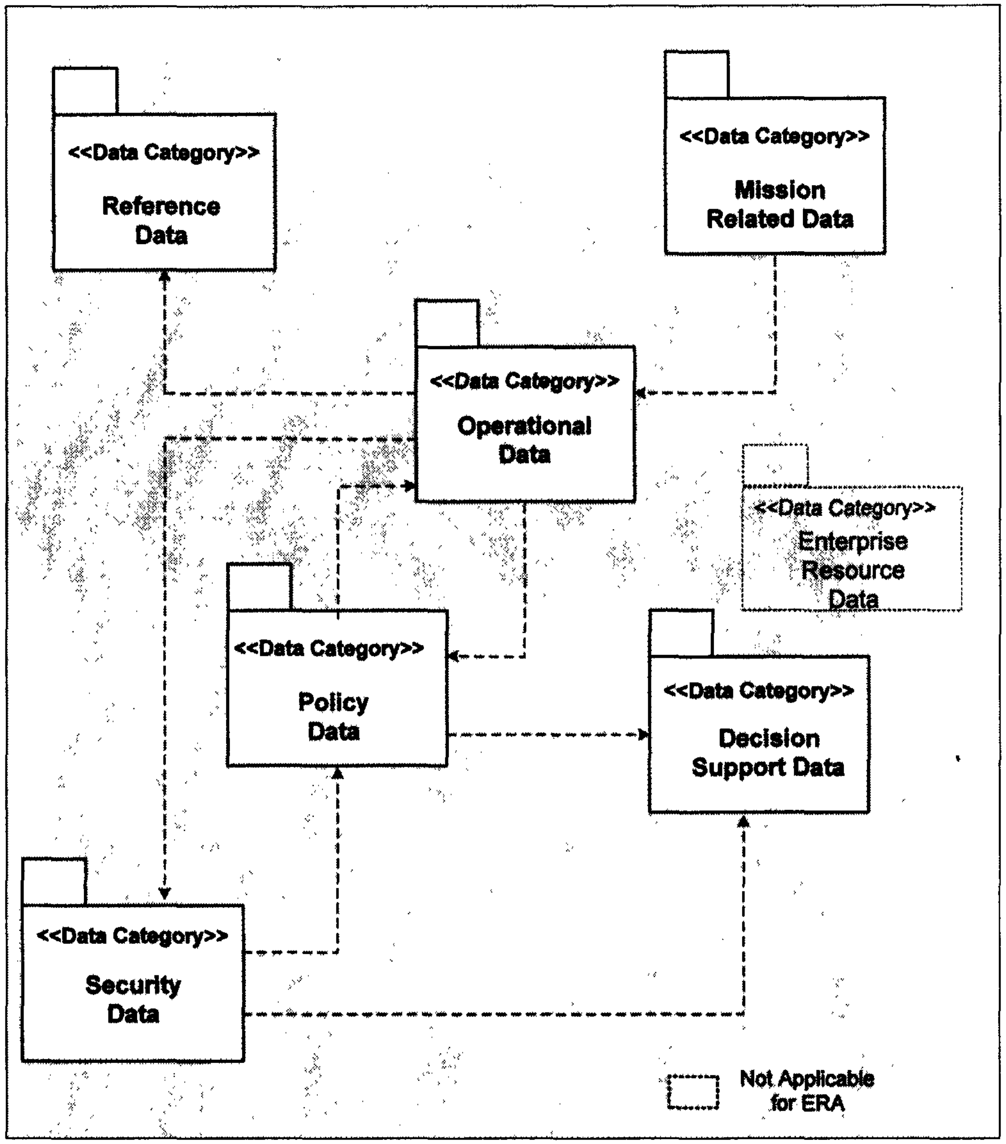
Cardinality is shown at the end of association 1 indicates "only one", "*" indicates zero or many, "1. *" indicates 1 or many

Data Model – Conceptual Model

Data Categories

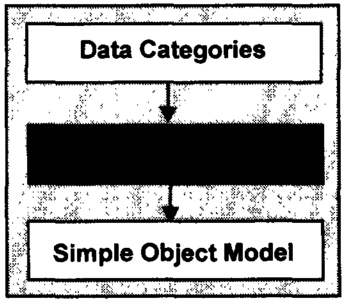

[redacted]

- Only included those relevant to ERA from NARA Data Architecture
- UML package notation with <<Data Category>> stereotype
- Includes dependencies derived from lower level models

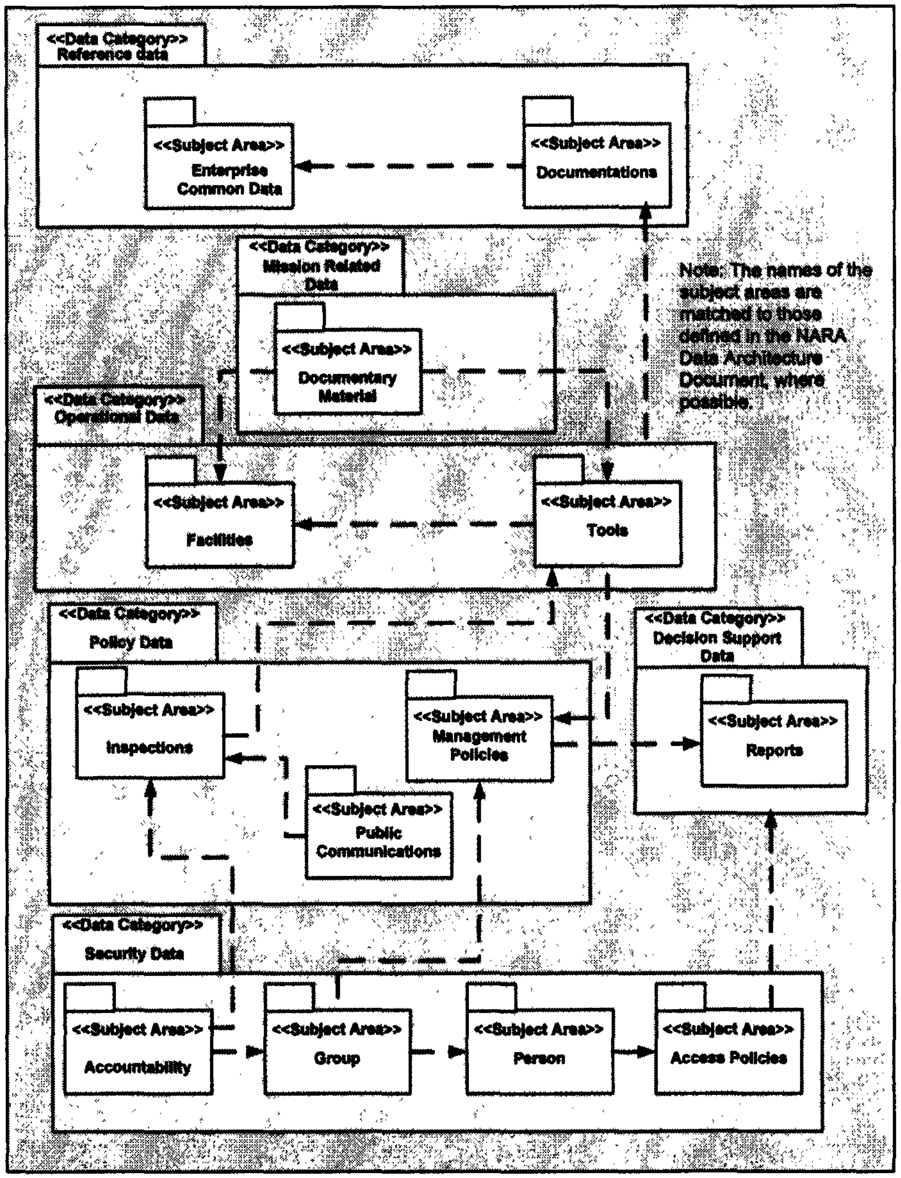


Data Model – Conceptual Model

Subject Areas



- Only included those relevant to ERA from NARA Data Architecture
- Not always a one-to-one with those identified in NARA data architecture
- UML package notation with <<subject area>> stereotype
- Dependencies derived from simple object model


<<Data Category>> Reference data
<<Subject Area>> Enterprise Common Data
<<Subject Area>> Documentations
<<Data Category>> Mission Related Data
<<Subject Area>> Documentary Material
<<Data Category>> Operational Data
<<Subject Area>> Facilities
<<Subject Area>> Tools
<<Data Category>> Policy Data
<<Subject Area>> Inspections
<<Subject Area>> Management Policies
<<Subject Area>> Public Communications
<<Data Category>> Decision Support Data
<<Subject Area>> Reports
<<Data Category>> Security Data
<<Subject Area>> Accountability
<<Subject Area>> Group
<<Subject Area>> Person
<<Subject Area>> Access Policies

Note: The names of the subject areas are matched to those defined in the NARA Data Architecture Document, where possible.

Data Model – Conceptual Model

Subject Area Highlights

[redacted]

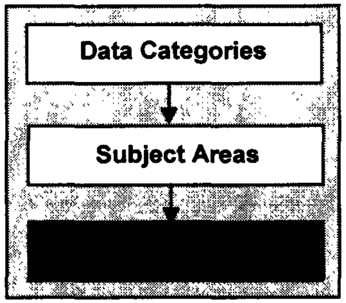
Data Model – Conceptual Model

Subject Area Highlights

[illegible]	[illegible]	[illegible]
Security Data	Access	Access Policies [illegible]
	Authorization	[illegible]
	Authentication	[illegible]
	[illegible]	[illegible]
Reference Data	[illegible]	[illegible]

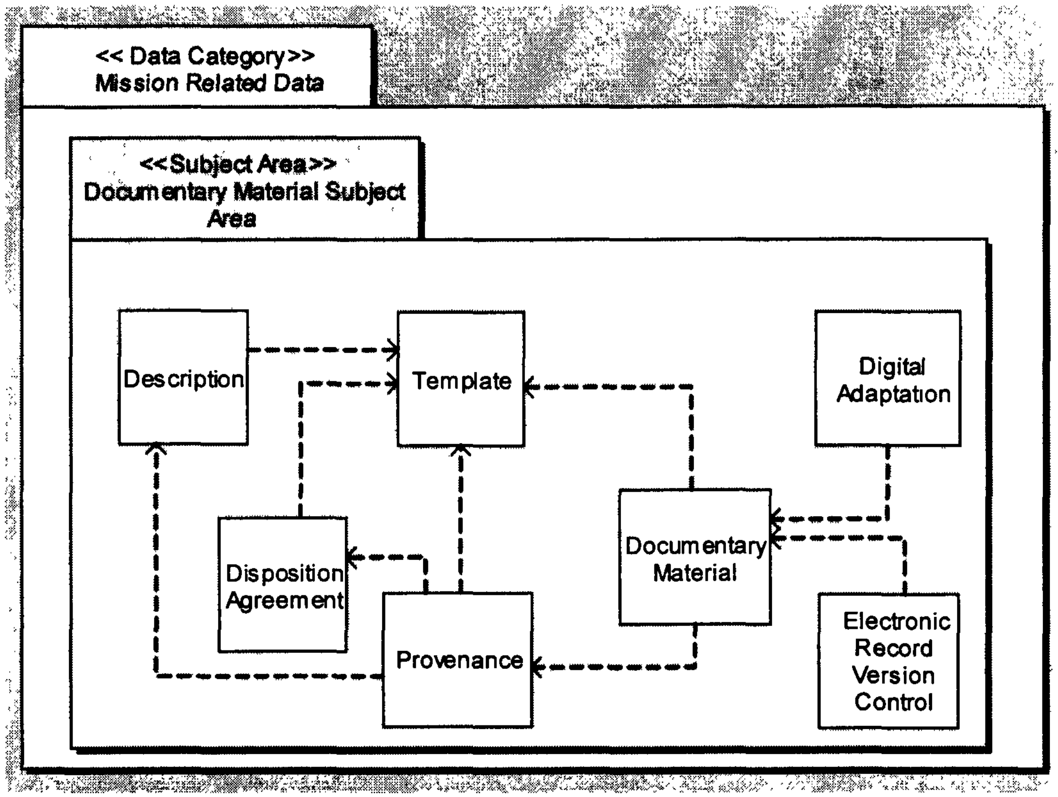
Data Model – Conceptual Model

Simple Object Model Example



[redacted]

- Each subject area decomposed into one or more simple objects
- A simple object represents a collection of lower level objects
- Starting point for the logical model
- Simple objects represented using UML class notation
- Key UML class diagrams contained in the SADD



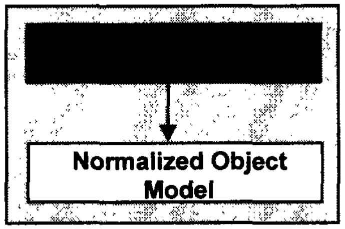
Data Model – Conceptual Model

Documentary Material Subject Area: Simple Object Model

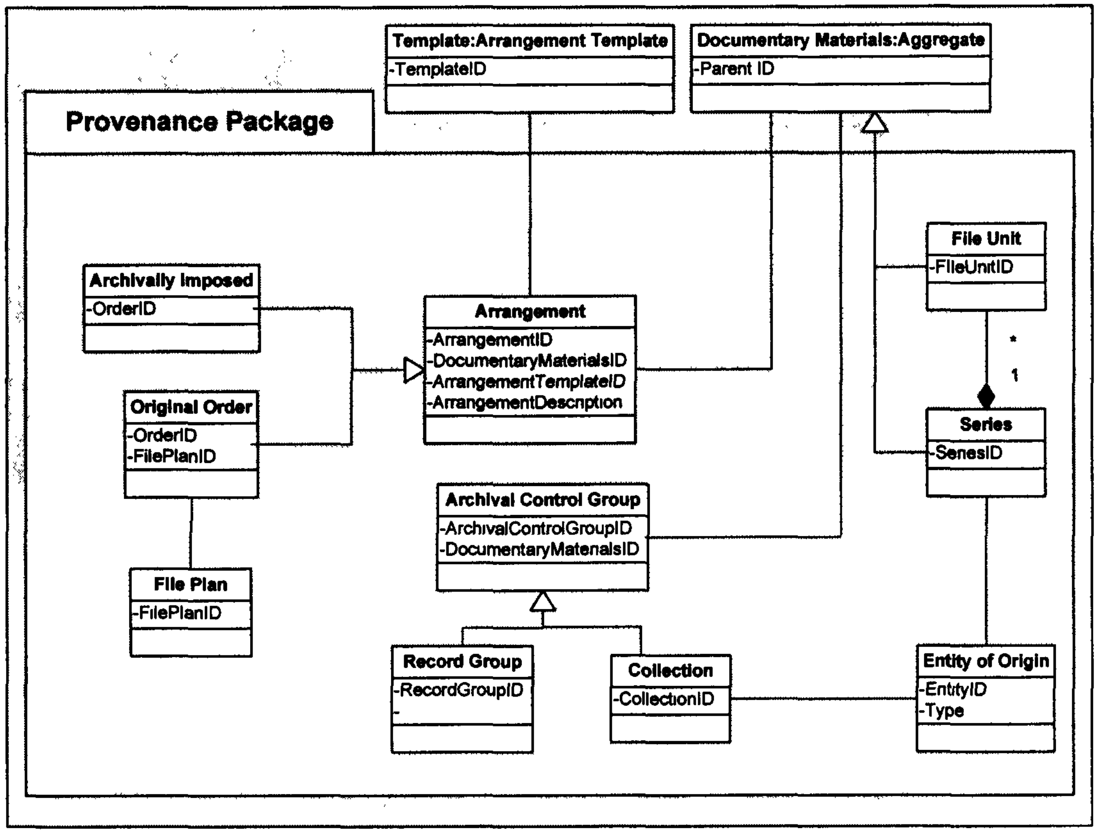
Documentary Material	Any documentary [illegible] Record, Document, [illegible]
Provenance	[illegible]
Disposition Authority	[illegible]
Description	[illegible]
[illegible]	[illegible]

Data Model – Logical Model

Logical Model



- Consists of complex object model and a normalized object model
- Logical Model is a decomposition of the conceptual simple object model
  - Described by UML Class diagram notation
  - Classes contain no operations
  - Simple objects from the conceptual model are represented by packages at the logical model



- Logical model will be further decomposed, fully attributed and normalized at future design increments

Data Model – Logical Model

Complex Object Model Highlights

Documentary materials

- Objects and relationships that describe artifacts, records and documents
- Retains the idea of an "Aggregate" object that may contain other aggregations of documentary materials

Provenance

- Objects that describe the provenance and arrangement of records
- Aggregates in many forms including Series, File Unit and Record Group

Disposition Agreement

- Objects and their relationships relating to disposition agreements, record schedules, transfer agreements, and appraisal reports

Data Model – Logical Model (continued)

Complex Object Model Highlights

Electronic Record Version Control

- Objects to manage copies and versions of records
- Notion of a Manifestation object

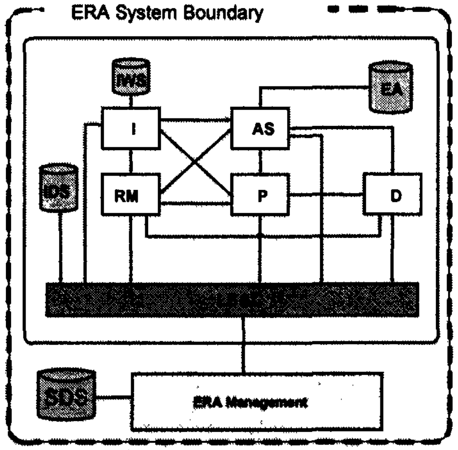
Digital Adaptation

- The physical view of a record
- Objects for relating data files to each record manifestation

Template

- Templates are extended hierarchically through inheritance

Data Model - Data Stores



ERA system comprised of four persistent stores

Electronic Archives

- Contains electronic records and other assets
- Serves as the repository for an instance
- Safe-Store repository for another instance

Ingest Working Storage

- Contains electronic records transfers undergoing ingest processing
- Prevents potential system contamination of viruses and miss-classified records

Instance Data Store

- A set of relational and object databases containing:
  - Records catalog, search indices and instance operational data
  - Assets (excluding the records themselves)
- Several orders of magnitude smaller than the electronic archives

System Data Store

- Contains system management data such as logs, inventory data, etc

Data Model – Replication

ERA is a distributed system and requires data replication

Replication ensures

- Each instance has a current copy of mission, security and operational data
- Each instance can operate with limited degradation, when access to other instances is unavailable

Replication is required

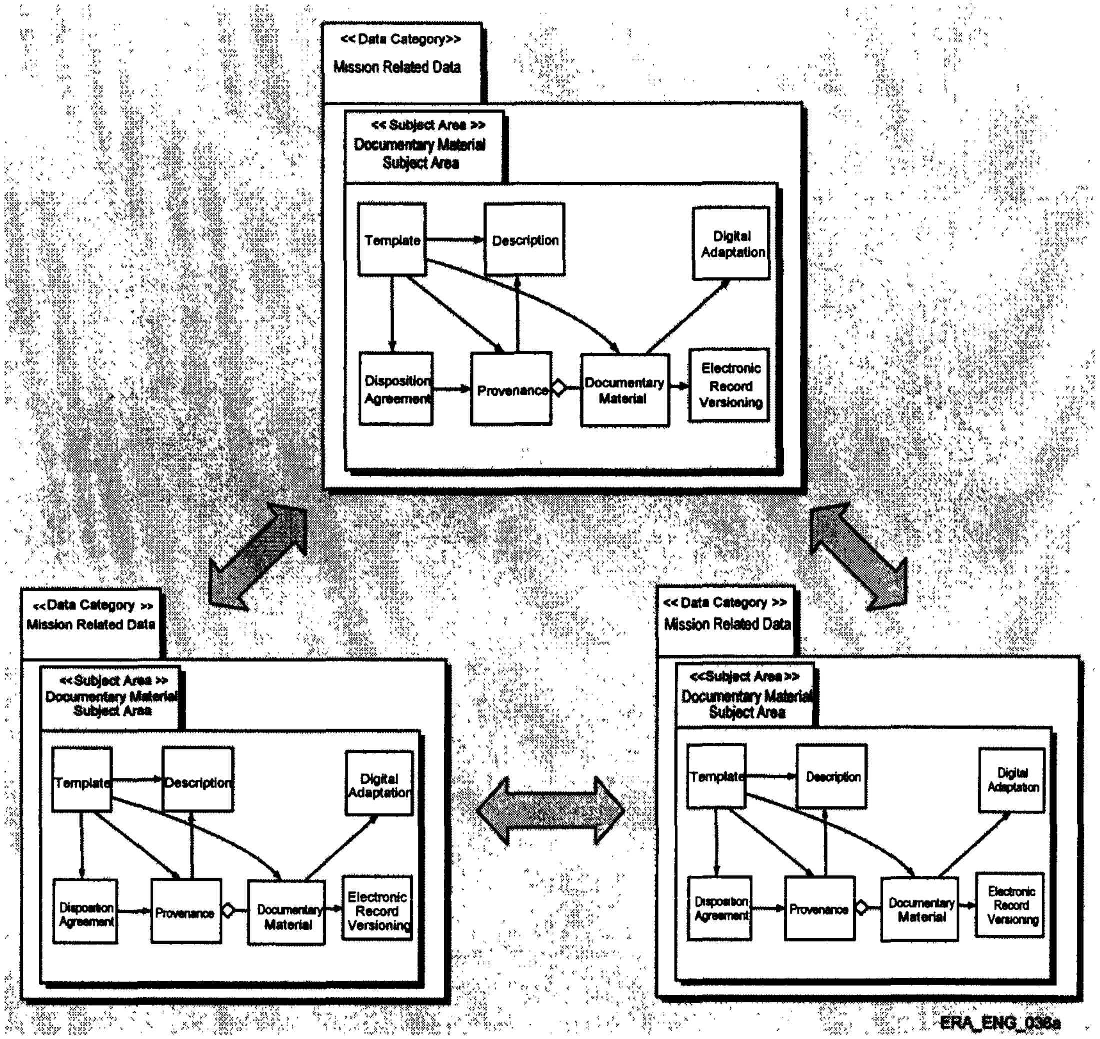
- Among the instances within a federation of classification
- Up / Down from the central management suite

Replication of Records is managed using Active Safe-Store

Data Model – Replication

Example of instance replication within a federation:

- Documentary materials data can be created or updated at any instance
- Replication ensures that all instances have the same current data
- Replication “On create” or “On update” reduces network traffic

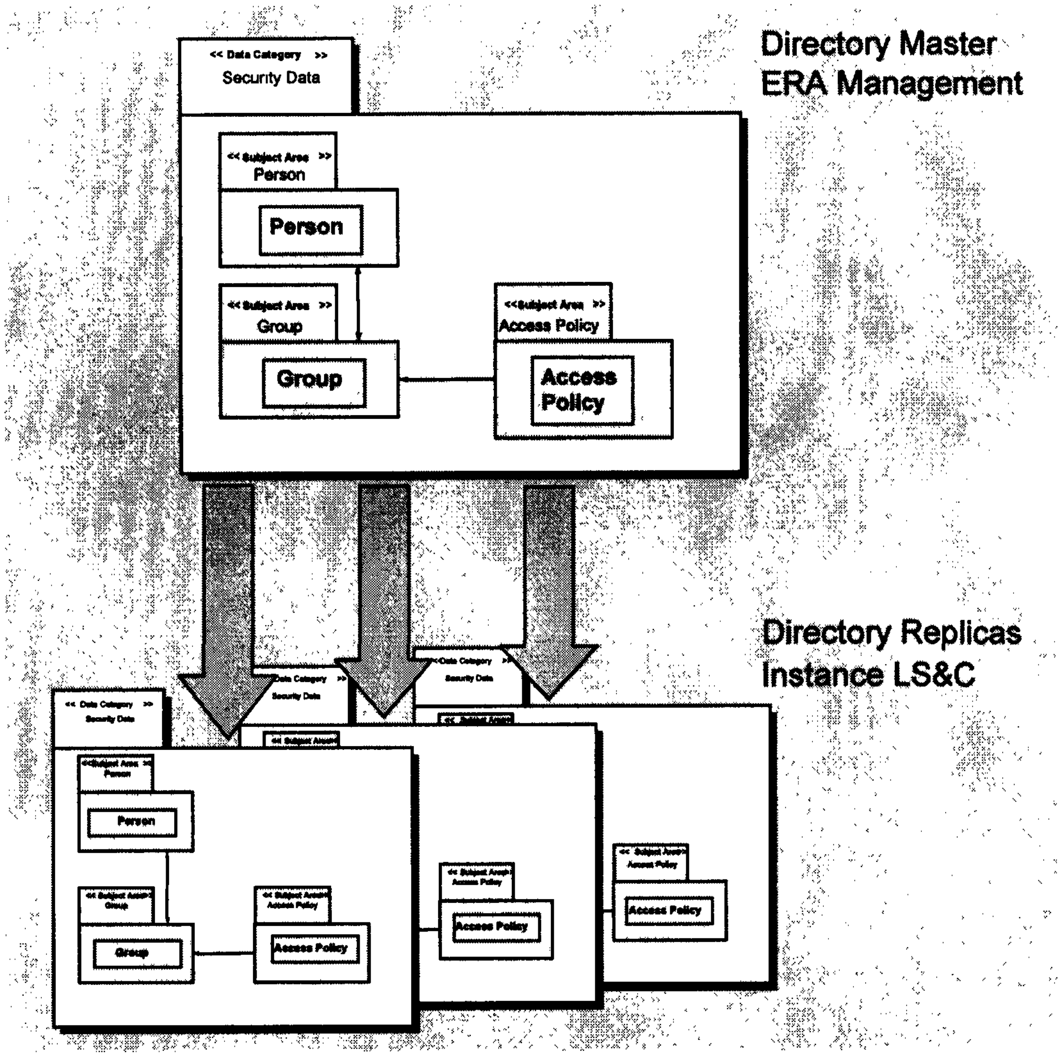


Note: Replication of Documentary Materials does not include records

Data Model – Replication

Example of flow down from master:

- Security data must be highly available. If security data is lost, user access to services and assets is lost
- Replication of user data is from a master copy to each instance
- If access to the master is lost the instance can continue to operate fully
- Propagation of new or updated information will occur once the connectivity is restored

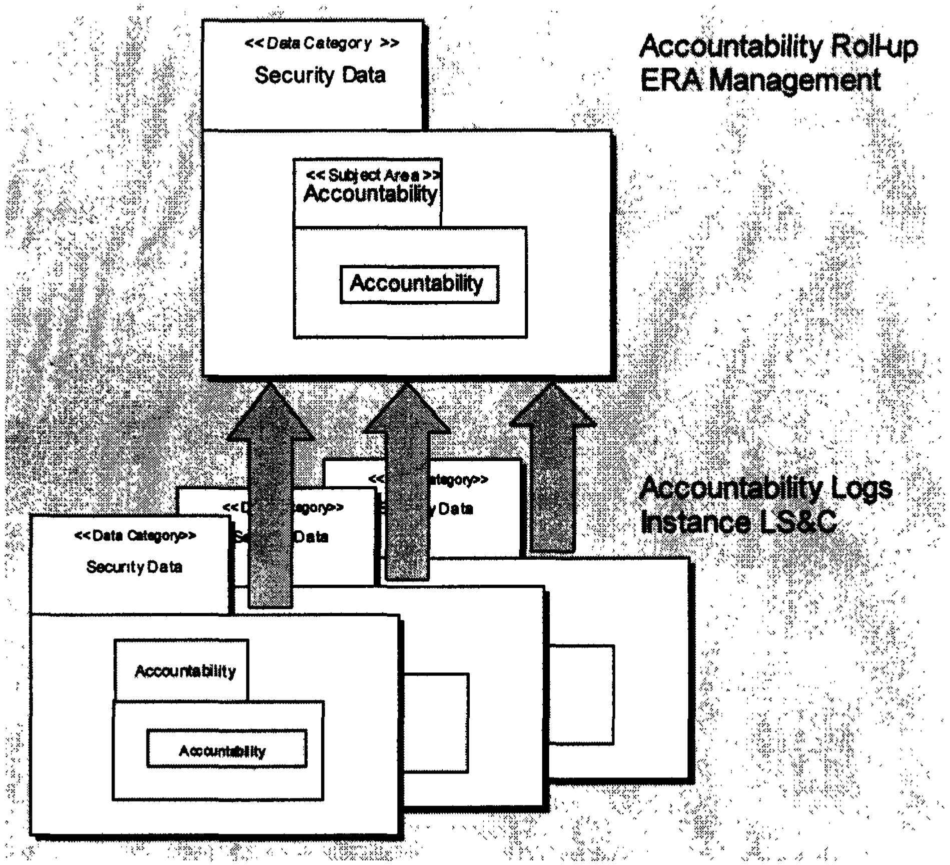


Data Model – Replication

Example of flow up to master:

- Accountability data including event and audit logs is created at each instance

- Replication to master must occur for a consolidated view and system level auditing



Data Model – Replication

Replication Models

- Synchronous replication
  - Dual commit to all data stores before releasing control
  - Has latency issues due to finite speed of light (> 1 ms)
  - Fails when target is unavailable
- Asynchronous replication
  - Releases control, then must assure:
    - All transactions are made to all data stores
    - "In order" writes
  - Avoids latency issues
  - Buffers transactions when target is unavailable

LM has chosen to implement asynchronous replication for ERA

- COTS solution
  - Quest SharePlex in conjunction with Oracle database
    - Replication based on reading the transaction logs
  - Verity Volume replicator for file systems

Data Model – Conclusions/Forward Plan

Architecture flows down from NARA's Data Architecture

Logical models refined in Increment 1

Physical models defined in Increment 1

Replication models refined in Increment 1 based on further performance modeling and testing

LM and NARA will collaborate on the development of these Data Models

Data Model – RIDs

RID-LMC00117 Data Replication

BREAK

ERA SDR – DAY TWO

Local Services and Control Design (LS&C)

[redacted]

May 11, 2005

Local Services & Control Design Agenda

Description of Functionality

Key Requirements

Functional Architecture

Service Design

Physical Design

Initial Product Selections

RID Discussion

LS&C Description

Local Services and Control provides

- A user interface portal
- Service orchestration and mediation
- Security services
- External interfaces
- Interfaces between ERA Instances
- Global Unique Identifier (GUID) management

LS&C Key Requirements

LM2 – Mediation and business process management service to orchestrate services

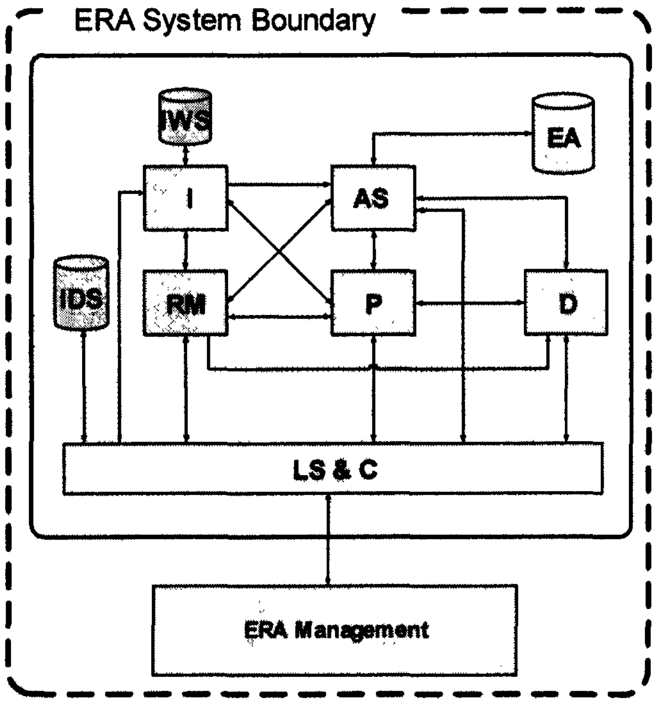
LM13 – Require authenticated and authorized access to all system services and assets

LM14 – Include a single point of access from user networks, including security appliances

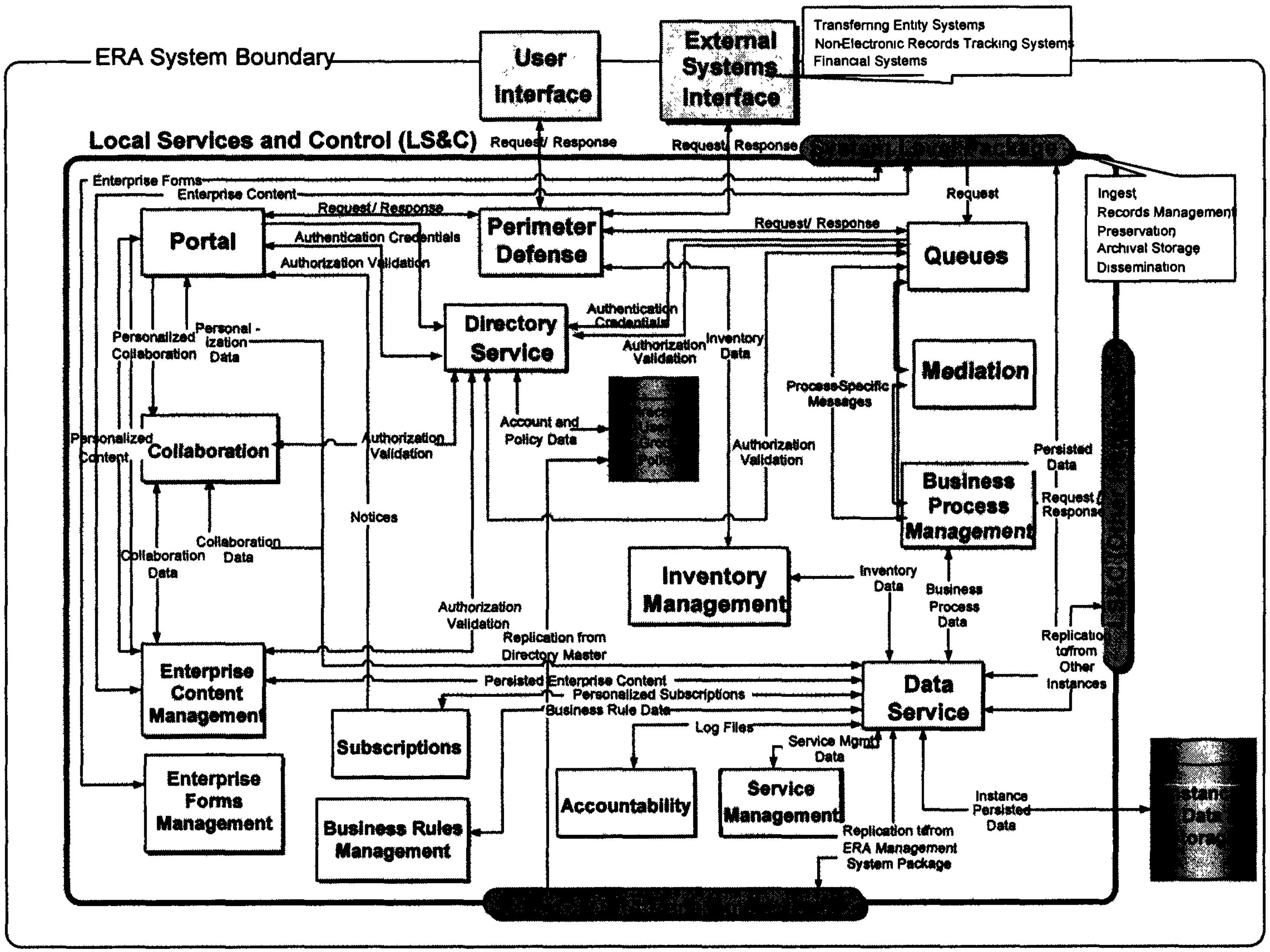
LM21 – Use web portal architecture

LM24.2 – Include asynchronous collaboration tools

LM30 – Provide a common facility for service registration and management



LS&C Functional Architecture

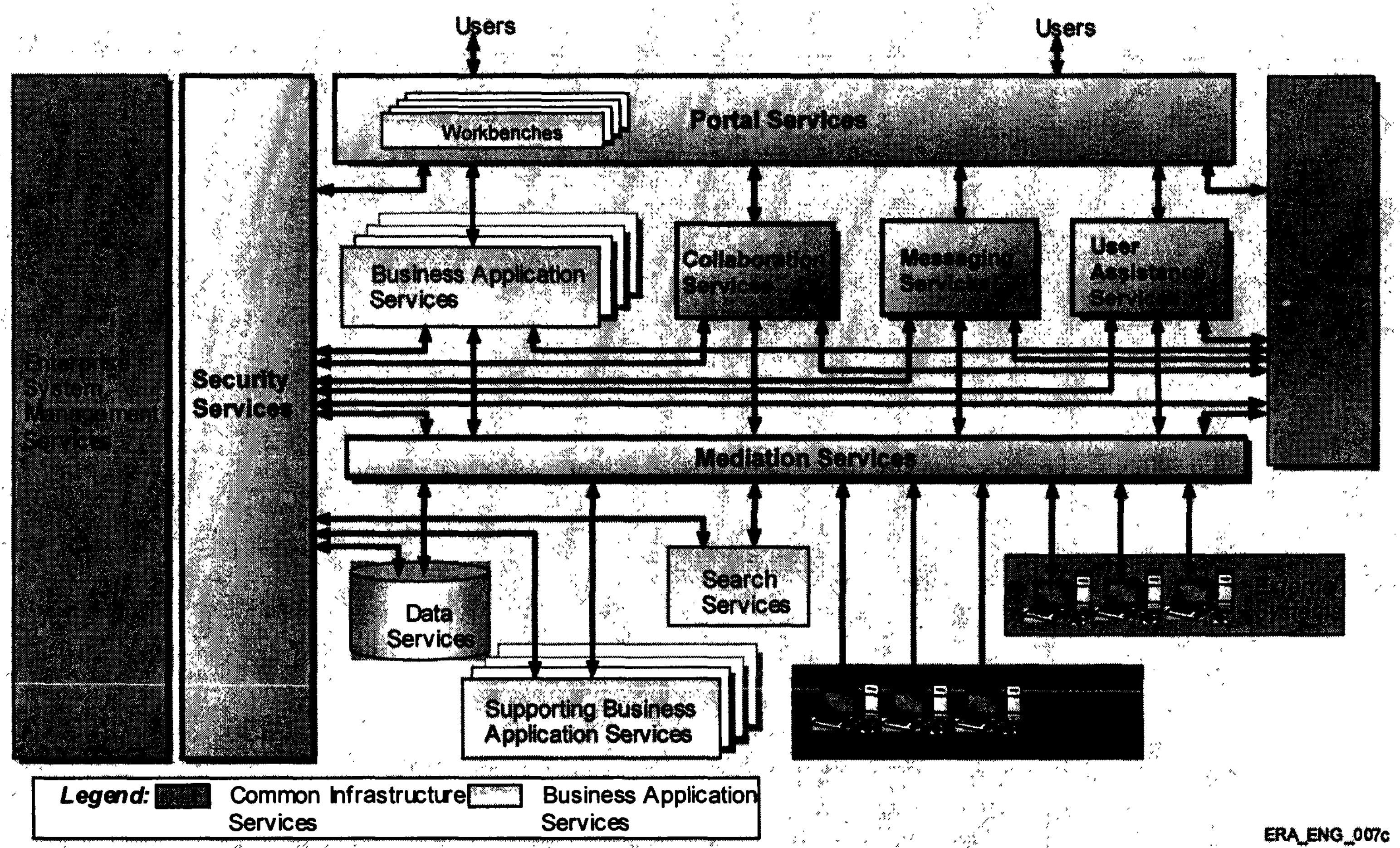


ERA_ENG_017b

LS&C within SOA

Local Services & Control includes

- Distributed Common Infrastructure Services


Users
Users
Portal Services
Workbenches
Business Application Services
Collaboration Services
Messaging Services
User Assistance Services
Enterprise System Management Services
Security Services
Mediation Services
Data Services
Search Services
Supporting Business Application Services
Legend: ▨ Common Infrastructure Services □ Business Application Services

ERA_ENG_007c

LS&C Services

Perimeter Defense	Provides network [illegible]
Directory Service	[illegible]
Accountability	[illegible]
Enterprise Content Management	[illegible]
Enterprise Forms Management	[illegible]
Collaboration	[illegible]

LS&C Services

[illegible]	[illegible]
Subscription Management	Provides for [illegible] Including [illegible]
Mediation	[illegible]
Business Process Management	[illegible]
Service Management	[illegible]
[illegible]	[illegible]

LS&C Services

Manage Media Inventory

Manage Accounts

[illegible]

LS&C Design Highlights

Perimeter Security

- Web servers are isolated in a DMZ, ensuring that external users have no direct access to system resources
- Router, firewalls, and Intrusion Detection Systems (IDS) provide intrusion deterrence and intrusion detection

Authentication, Identification, and Authorization

- Follows the Lightweight Directory Access Protocol (LDAP) standard for Directory Services
- Includes Single Sign-On that follows the Security Assertion Markup Language (SAML) standard
- Identity Management is centralized in ERA Management

Portal Framework

- Provides a configurable user workbench capability
- Facilitates ease of maintenance with server-side deployment
- Promotes a consistent user-interface across user classes and lines of business

LS&C Design Highlights

Enterprise Content Management (ECM)

- Provides coordination, workflow, and configuration management file- and object-based content
- Follows the J2EE standard, and integrates with the Portal
- Review and approval workflows may be tailored for each organization, or for each type of asset

Collaboration

- Tools include document libraries, threaded discussions, events calendar, and an SME/Point-of-Contact registry
- Team spaces provide an easy to administer collection of collaboration tools focused on a single team
- Integrated with the Portal and the Directory Service so that one set of "groups" provides common access control
- The collaboration framework is extensible to add additional collaboration functionality

LS&C Design Highlights

Mediation

- Includes mapping the logical data model to the physical data model
- Includes queues, which provide a distributed messaging infrastructure, and avoids the plethora of interfaces that a point-to-point model would yield.

Orchestration

- The LM Team design uses orchestrations to encode business processes
- Orchestrations follow the Business Process Execution Language (BPEL) standard
- Each Orchestration is packaged as a J2EE enterprise application, thus making each Orchestration deployable independent of any other ERA component

Data Service

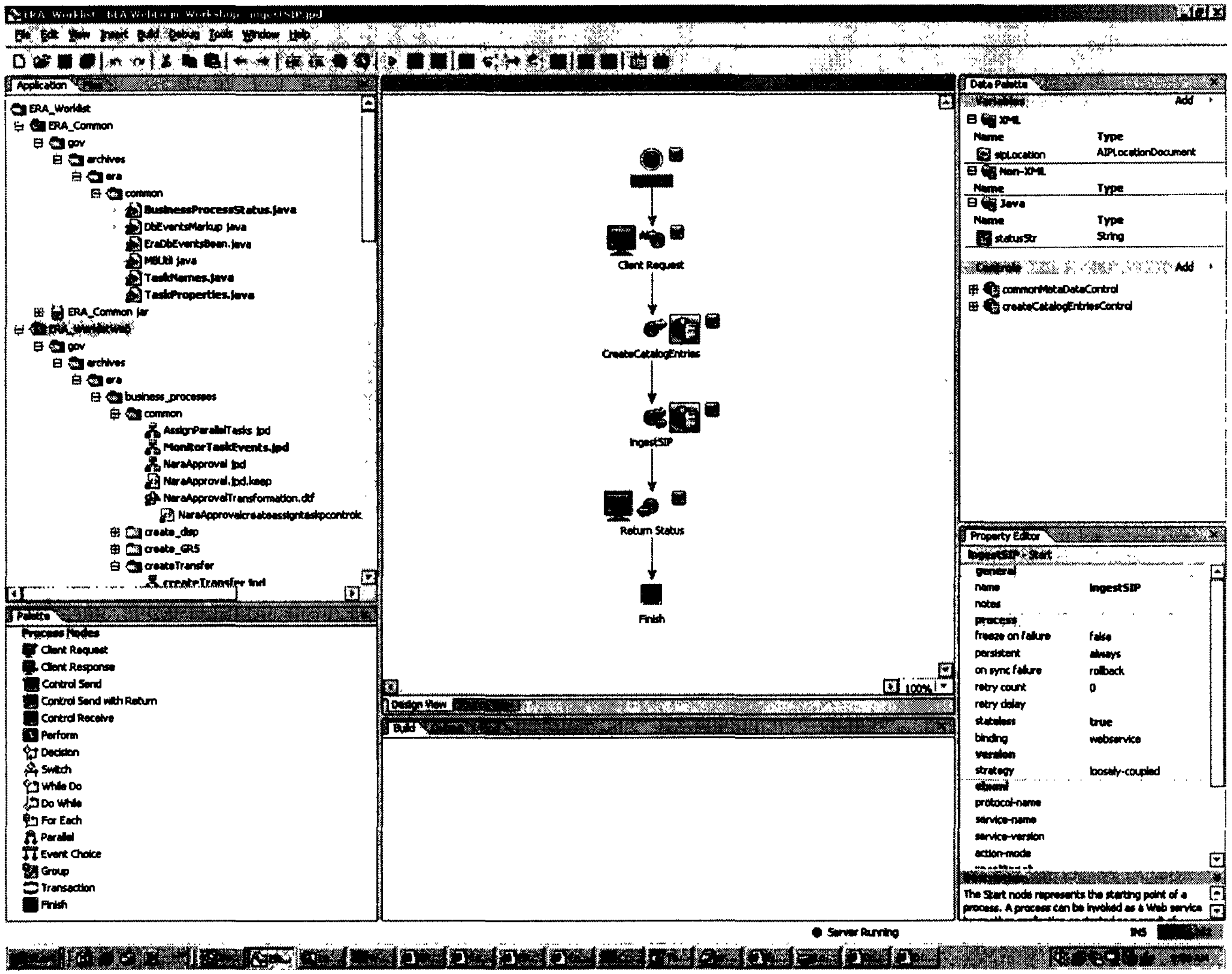
- Provides for the storage and management of persistent data
- Provides support for relational, object-oriented, and file-based data stores

LS&C Design Highlights – Development of Orchestrations

Standards compliant, graphical Integrated Development Environment (IDE)

Integrated with Software CM

Delegated authorization for creation and modification



ingestSIP - Start

general

name	ingestSIP
notes	

process

freeze on failure	false
persistent	always
on sync failure	rollback
retry count	0
retry delay	
stateless	true
binding	webservice
version	
strategy	loosely-coupled

element

protocol-name

service-name

service-version

action-mode

The Start node represents the starting point of a process. A process can be invoked as a Web service

LS&C Design Highlights

Interfaces with External Systems

- Interface partners use the same SOA infrastructure as all of the services within ERA
- SOA infrastructure includes mediation, queues, and security

LS&C Design Details – Create and Manage GUID

GUID is a *FILE* identifier

GUID implementation is SHA256 hash

- NIST standard algorithm
- 256 bits => 1 x 10^{77} values
- High statistical probability of uniqueness

Management includes relating assets

- Relationships include:
  - Collecting files into records
  - Relating original, POF, redacted versions
  - Relating records and their lifecycle metadata
- Relationships held:
  - In Records Catalog
  - In a self-describing manner
  - Relationships can be reconstructed from archival storage
- GUID used to validate file integrity

LS&C Design Trades

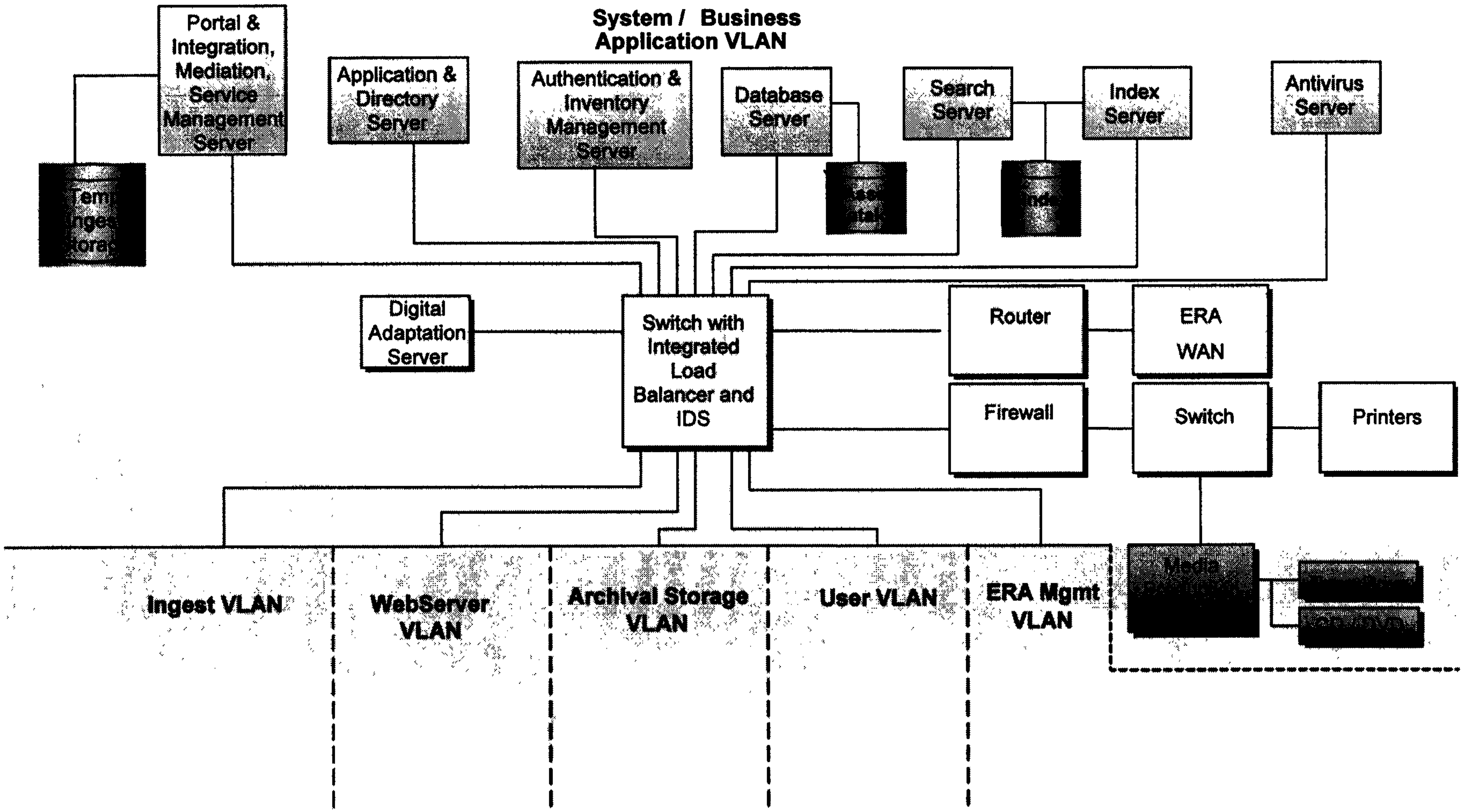
Implement a Service Oriented Architecture

- Collected common infrastructure services into core packages
- LS&C provides:
  - Core services that are leveraged by business service components
  - Distributed
- ERA Management provides:
  - Core services to manage business service components
  - Centralized

Leverage COTS products to provide core infrastructure services

- Chose J2EE over .NET or custom
  - Secure, scalable, mature application framework
- Chose Integrated identity management suite
  - Authentication, Identification, and authorization
  - Delegated administration of authorization

LS&C Physical Design



System / Business Application VLAN

Portal & Integration, Mediation, Service Management Server

Tem Inges Storag

Application & Directory Server

Authentication & Inventory Management Server

Database Server

Search Server

Index Server

Antivirus Server

Digital Adaptation Server

Switch with Integrated Load Balancer and IDS

Router

ERA WAN

Firewall

Switch

Printers

Media

Ingest VLAN

WebServer VLAN

Archival Storage VLAN

User VLAN

ERA Mgmt VLAN

LS&C Physical Design

System/Business Application VLAN

- Includes the main business process components of the ERA System
- Includes partitioned servers
  - Allows unrelated service components to be hosted on the same physical server in a manner that is convenient for deployment, scalability, operations, and management
  - Includes dedicated servers where needed for performance or clustering
  - Scales vertically and horizontally to larger or smaller configurations
- Provides a main network switch that contains integrated load balancing, intrusion detection, and firewalls
- Contains an Instance Data Store
  - Persistence of long-lived, non-record assets
  - Clustering, replication, backup, and ability to restore from archival assets ensures long-term continuance of data
- Hosts workstations for Media Production

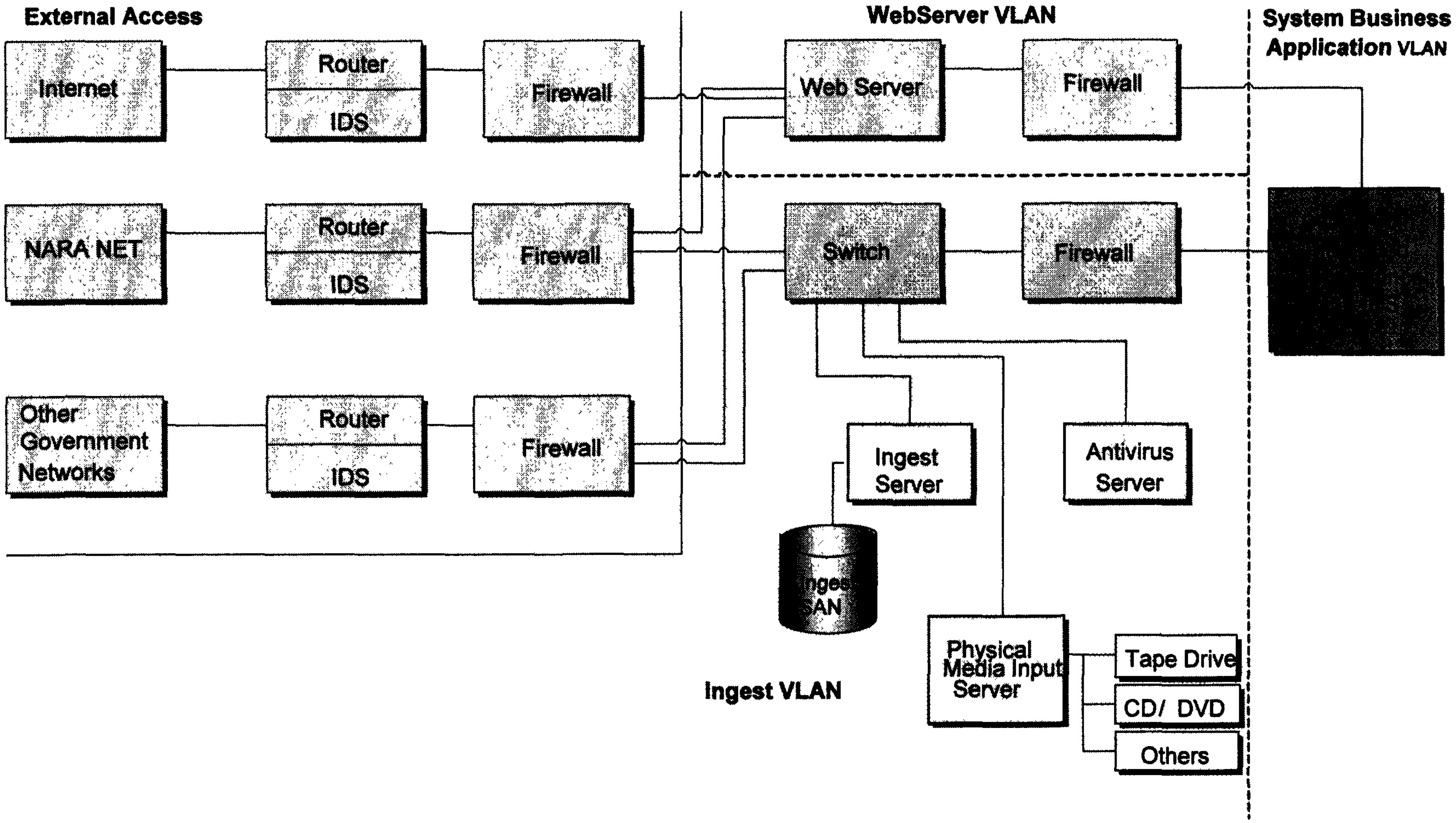
System/Business Application VLAN COTS Products

Application Server	BEA WebLogic Application Server	Sun [illegible]
Directory Server	Sun JES Directory [illegible] Identity Manager	[illegible]
Strong Authentication	[illegible]	[illegible]
Inventory Management Server	[illegible]	[illegible]
Database Server	Ora[illegible]	[illegible]
Search Server	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

System/Business Application VLAN COTS Products

[redacted]

LS&C Physical Design



LS&C Physical Design

Web Server VLAN

- Provides a a Demilitarized Zone (DMZ), isolated by firewalls and Intrusion Detection Systems (IDS) to minimize the danger of External threats and attacks to the rest of the ERA System
- Includes a WebServer, which proxies all user requests to their applicable portal-based workbenches and to their included business functionality
- No external user has the capability to communicate directly with any of the back-end business functionality or with Archival Storage

Web Server VLAN COTS Products

[redacted]

LS&C – Conclusions

LS&C provides:

- Distributed enterprise-wide infrastructure services
- Security services
- Core service components that are leveraged by business services
- Business Process and Business Rules Management

LS&C uses COTS frameworks for:

- Workbenches and collaboration
- Business process mediation and orchestration
- Relational and object data management
- User authentication, identification, authorization, and session management

LS&C abstracts the core infrastructure from the Business Application services

LS&C RIDs

RID-LMC00115 Missing LMC Requirements in the SADD

RID-LMC00116 COTS use for LS&C and Management

RID-LMC00130 Workflow COTS

RID-LMC00131 Creation and Monitoring of Workflows

RID-LMC00133 GUIDs in the Storage Architecture

RID-LMC00136 Primary COTS Selections

LUNCH

ERA SDR – DAY TWO

Ingest Design

Rick Rogers
May 10, 2005

Ingest – Agenda

Description of Functionality

Key Requirements

Functional Architecture

Service Design

Design Highlights / Trades

Ingest Modeling

Physical Design

RID Discussion

Ingest Description

Ingest provides

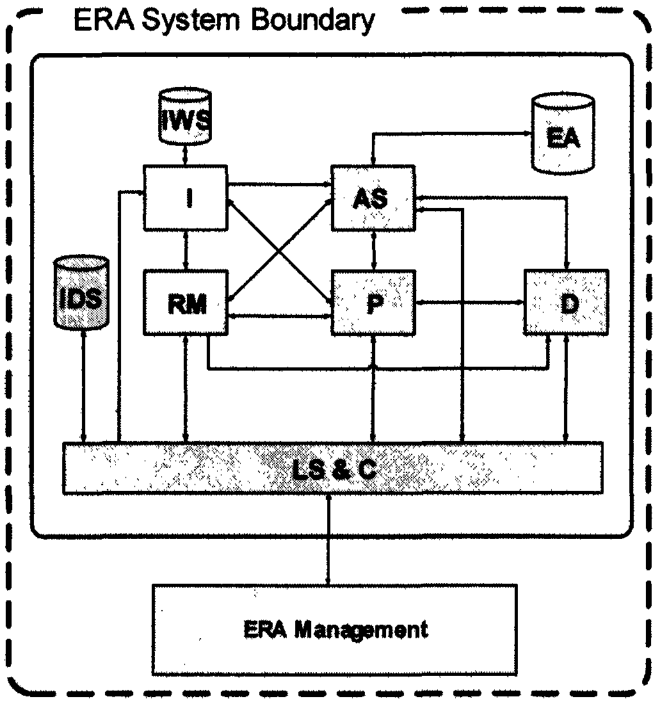
- Mechanisms to receive the electronic records from the transferring entities
- Prepares those electronic records for storage within the ERA System
- Performs virus scans
- Validates transfers

Ingest Key Requirements

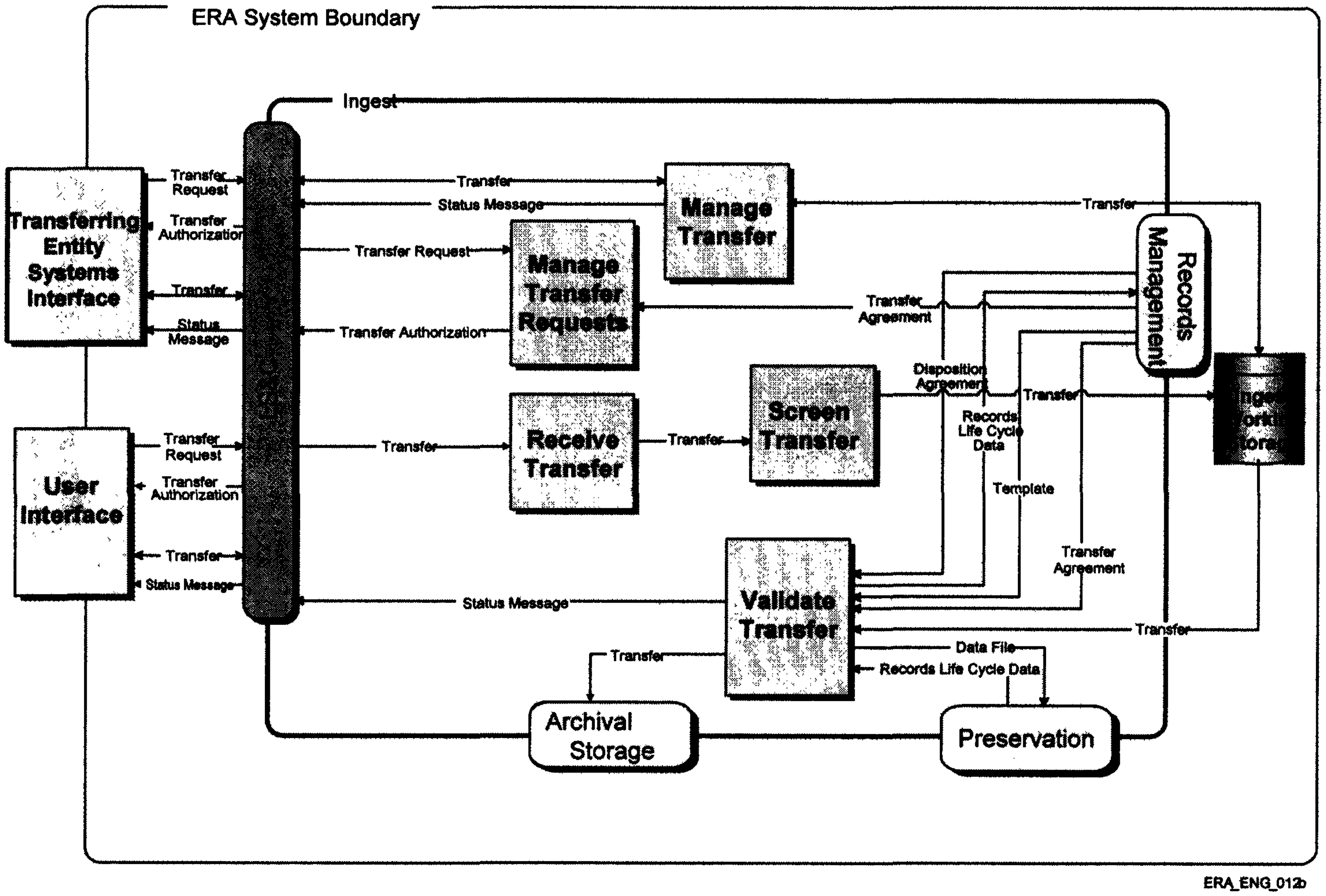
LM1.8 – Ingest tools that are used to manage transfers

LM13.7 – Segregated Ingest Working Storage so that transferred records are segregated until their security access level and access restrictions are determined

LM31.1 – Fully automated ingest process, to support large ingest volumes



Ingest Functional Architecture



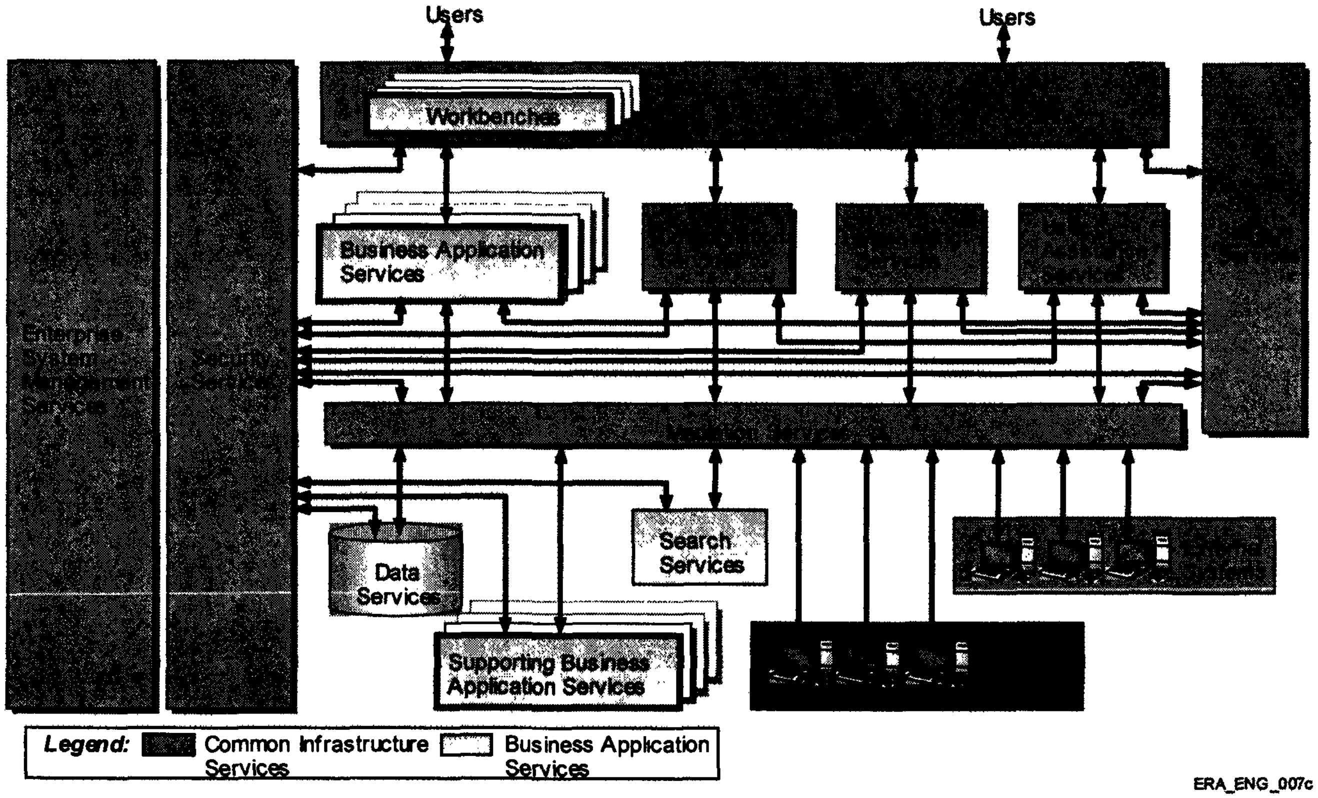
Key features:

- Tools for Transferring Entities
- Segregated Ingest Working Storage
- Fully Automated Ingest Process

Ingest within SOA

Ingest includes

- Workbenches
- Business Application Services
- Supporting Business Application Services


Users
Users
Workbenches
Business Application Services
Enterprise System Management Services
Security Services
Data Services
Search Services
Supporting Business Application Services
Legend: ▩ Common Infrastructure Services □ Business Application Services

ERA_ENG_007c

Ingest Services

Transferring Entity Workbench	Provides [illegible]
Manage Transfer Request	
Manage Transfer	
Receive Transfer	
Screen Transfer	
[illegible]	

Ingest Design Highlights

Automated Ingest

- Virus scanning, initial security review, and transfer validation performed automatically
- Developing archival metadata deferred to Records Management

Automated Transfer Management

- Transfer Requests are automatically validated against Transfer Agreements

Automated Descriptive Data

- Ingest automatically captures descriptive data from all available artifacts so that transfers can be placed into Archival Storage and discovered later

Segregated Ingest Storage

- Transfers remain in Ingest Working Storage until they are screened and validated

Ingest Design Trades

Moved archival description from Ingest (as originally proposed) to Records Management

- Facilitates high-volume ingest

Considered implementing client-side tools to facilitate packaging of transfers

- Provisionally deferred this in favor of web portal-based tools, but will revisit decision during Increment 1

Conducting research on automating extraction of descriptive data (including targeted entity extraction)

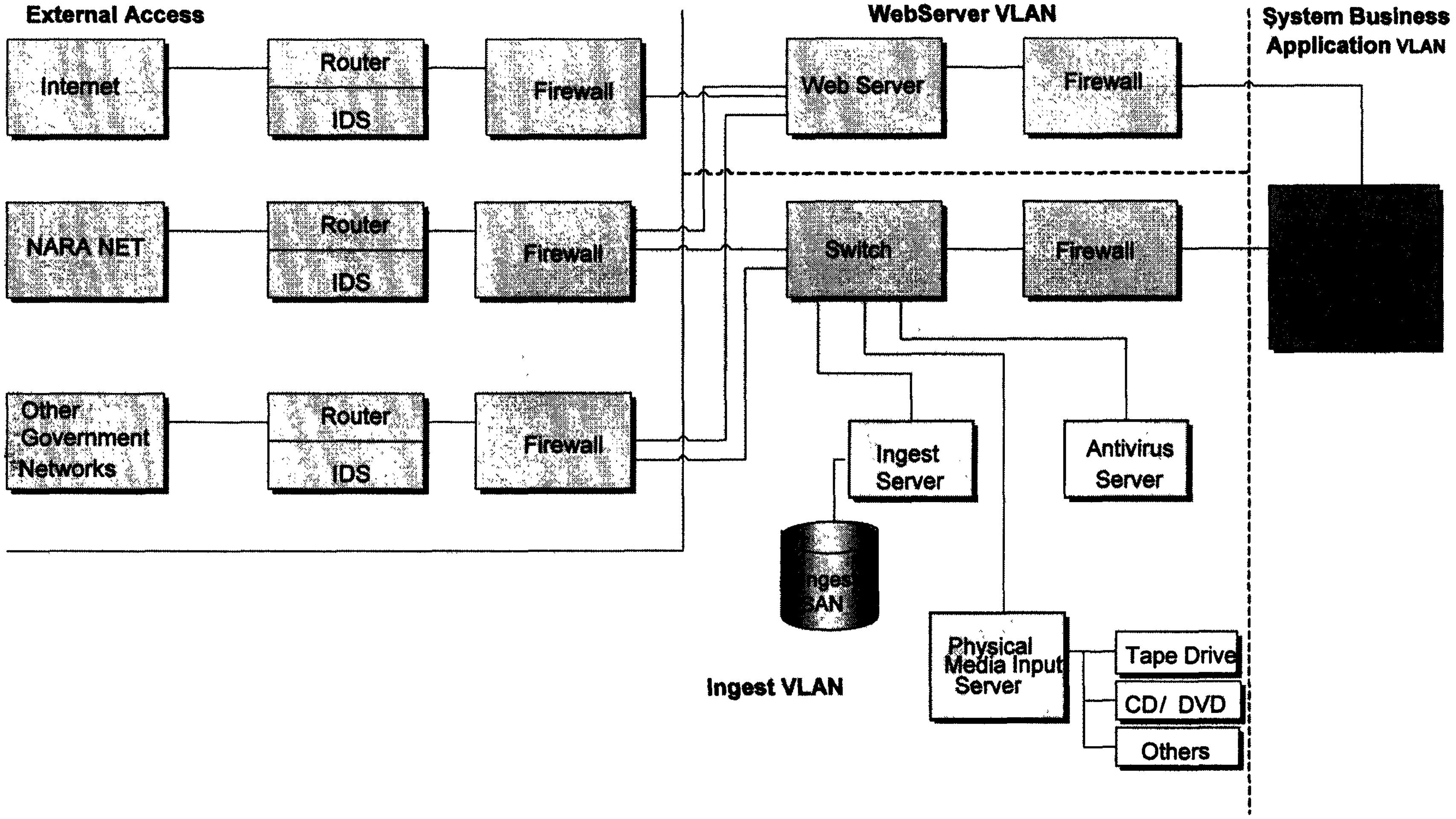
Ingest Design Modeling

[redacted]

Ingest Design Modeling

[redacted]

Ingest Physical Design



Ingest Physical Design

Ingest VLAN

- Supports Ingest services and interfaces with the System/Business Applications VLAN
- Segregates sensitive records (i.e. Title 13 and classified data) during the ingest
- Provides a Demilitarized Zone (DMZ) to ease recovering from a security violation from misclassification
- For unclassified instances, interfaces through a perimeter security layer to the Internet, NARANET and GSA Connections to receive transfers
- For classified instances, users must physically be located within the SCIF using hardware directly connected to the Instance to perform the Ingest functions
- Extensible to support JWICS and SIPRNET connections

Ingest VLAN COTS Products

Ingest Conclusions

Automated Ingest

- Supports high-volume ingest
- Ensures ingested records are validated, virus scanned, and security checked

Automated Descriptive Data

- Captures descriptive data from all available sources
- Allows ingested records to be discovered later
- Provides time for archivists to author Archival Description

Segregated Ingest Working Storage

- Contains potential virus-infected or misclassified records
- Segregates restricted records (such as Title 13)

Segregates the Ingest of records into ERA from all other functions which maximizes system security

Ingest RIDs

RID-LMC00129 Transfer Agreements

BREAK

ERA SDR – DAY TWO

Records Management Design

Rick Rogers
May 10, 2005

Records Management: Agenda

Description of Functionality

Key Requirements

Functional Architecture

Service Design

Design Highlights / Trades

Physical Design

RID Discussion

Records Management Description

Records Management provides

- Services necessary to manage the archival properties and attributes of the electronic records and other assets within the ERA System
- Capability to create and manage new versions of those assets
- Management functionality for disposition agreements, disposition instructions, appraisal, transfer agreements, templates, authority sources, records life cycle data, descriptions, and arrangements
- Management functionality for access review and redaction processes
- Capability to implement disposition instructions
- Selected management functionality for non-electronic records
- Management functionality for FOIA and Privacy Act requests

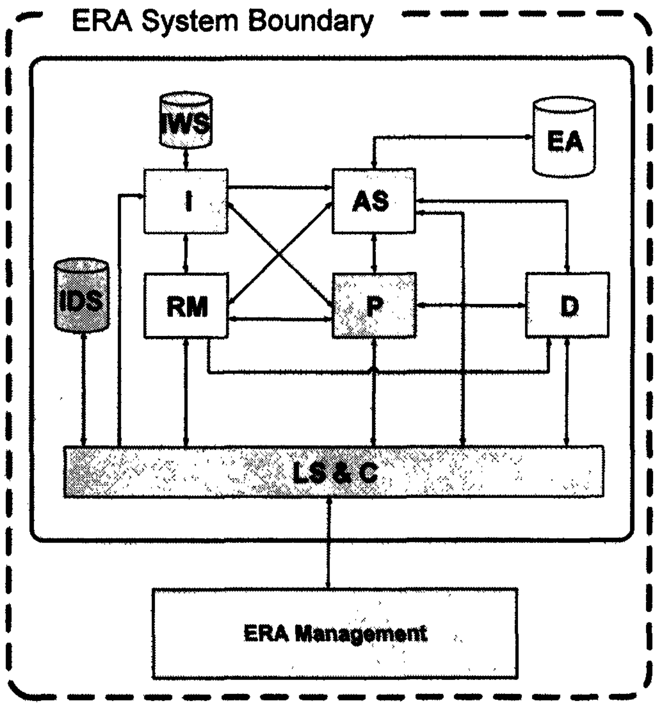
Records Management Key Requirements

LM1, LM3, LM4, LM7, LM9 – Common approach to managing assets

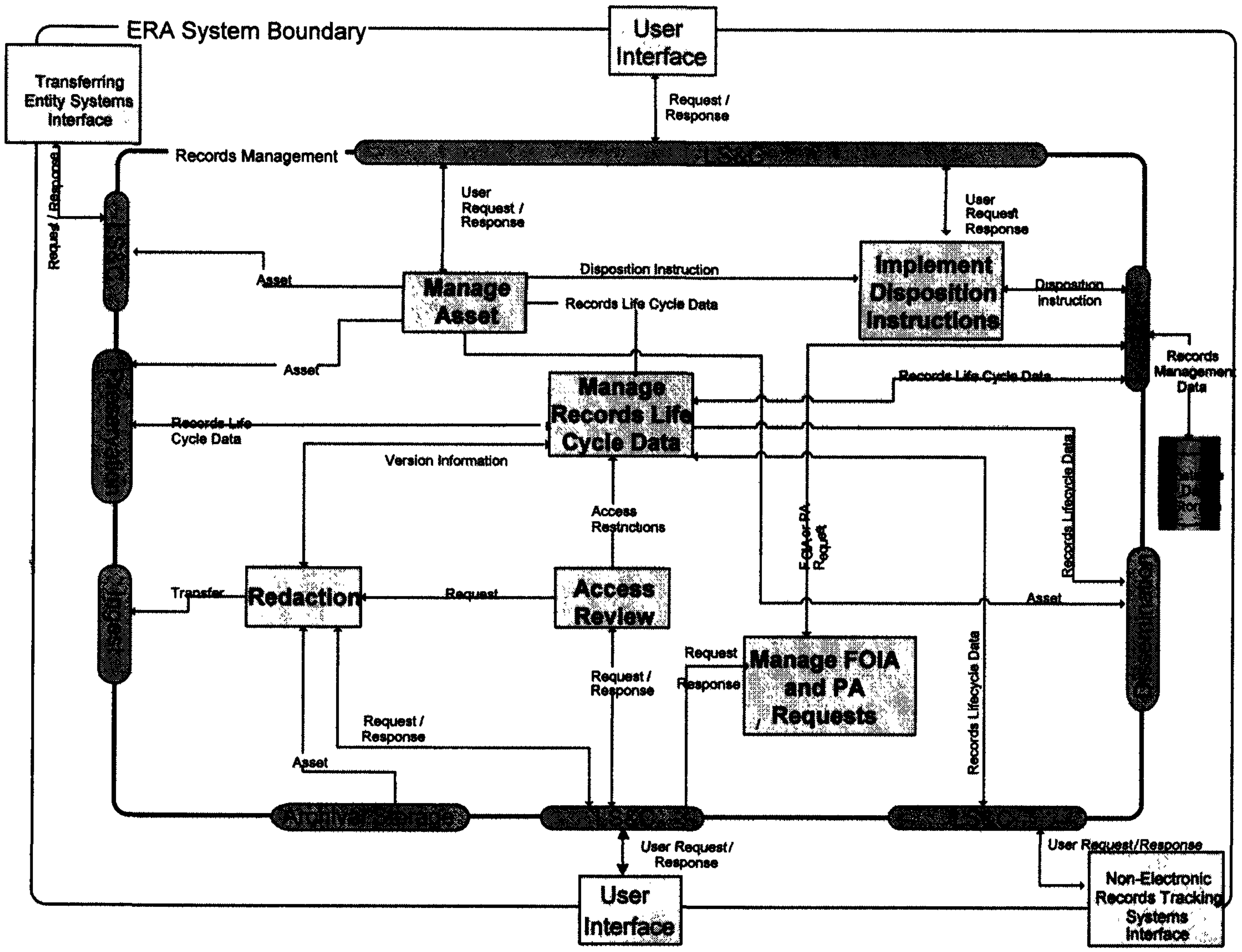
LM5 – Centralized facility to manage all records life cycle data and transactions

LM17, LM18 – Facility to perform access review and perform redaction in context of record life cycle

LM2.10, and LM2.10 – Common approach managing FOIA and Privacy Act Requests



Records Management Functional Architecture



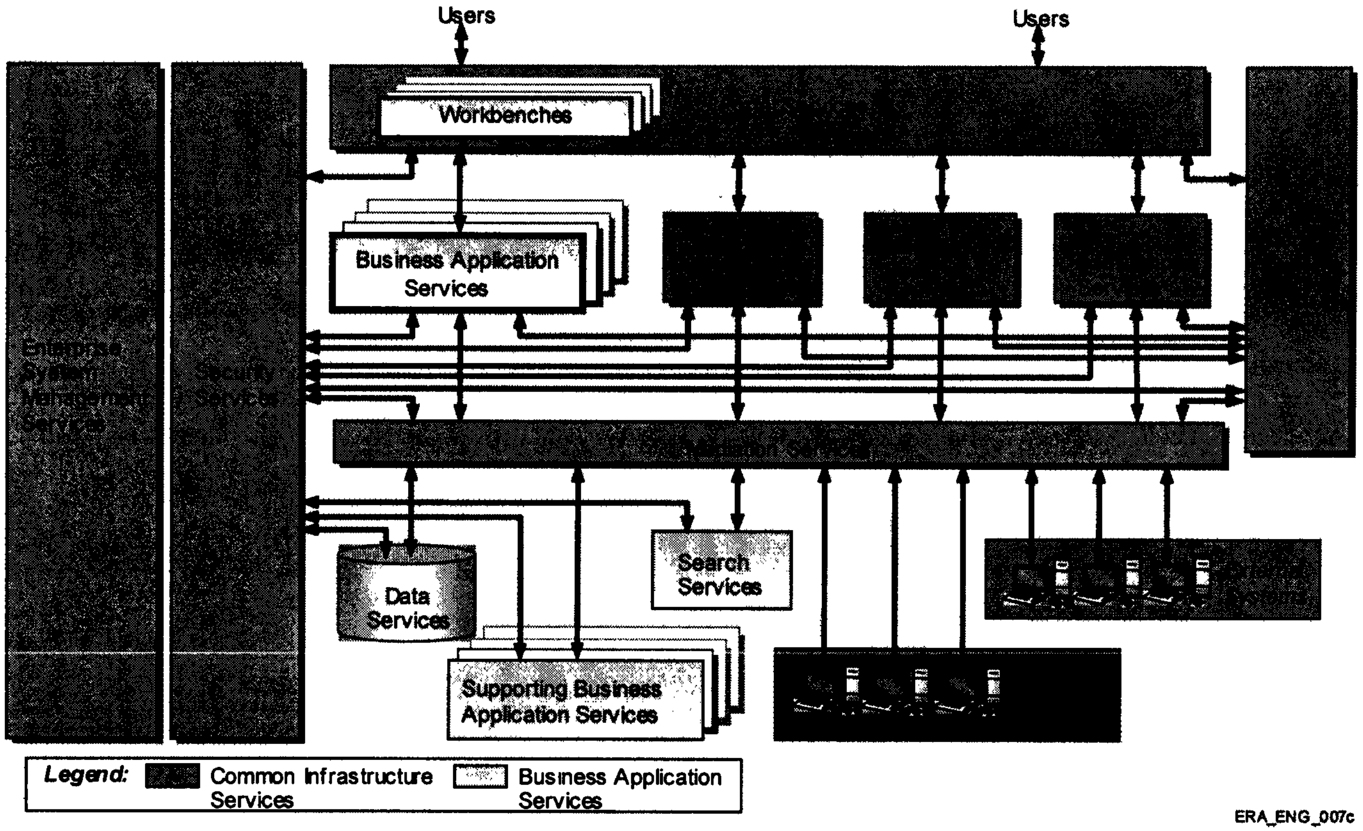
Key Features:

- Common Asset Management Approach
- Centralized Life Cycle Management
- Access Review & Redaction
- Centralized Disposition Instructions

Records Management within SOA

Records Management includes

- Workbenches
- Business Application Services


Users
Users
Workbenches
Business Application Services
Enterprise System Management Services
Security Services
Mediation Services
Data Services
Search Services
Supporting Business Application Services
Legend:
Common Infrastructure Services
Business Application Services

ERA_ENG_007c

Records Management Services

Appraiser	Provides [illegible]
Records Processor	[illegible]
Access Reviewer	[illegible]
Access Review	[illegible]
Redaction	[illegible]
[illegible]	[illegible]

Records Management Services

Transferring Entity Systems

Non-Electronic [illegible]

Records Management Design Highlights

Centralized Records Life Cycle Management
- Ensures consistency, integrity, and authenticity of the records as they are managed within the system

Centralized Records Catalog
- Provides a guide to every asset within ERA

Interview-Style User-Interfaces
- Presents a series of interview questions to the user, and uses responses to these questions to suggest responses, impute responses, and skip irrelevant responses

Authority Sources
- Supports both hierarchical and network topologies.
- Authority Sources leveraged throughout the system, such as in Preservation and Service Level Plans

Records Management Design Highlights

Templates

- Leverages templates throughout ERA for both archival and system processes
- Supports inheritance and context hierarchies
- Includes a template editor framework

Deterministic Disposition Instructions

- Includes an approach to defining disposition instructions in a way that can be automatically implemented by the system
- Includes an event database for event driven instructions

Persistent Identifiers

- Includes an approach to identifying assets in a permanent and transparent way

Persistent Archives

- Includes an approach for ensuring the archives are free from dependence on specific hardware and software technologies

Records Management Design Trades

Moved Access Review and Redaction from Dissemination (as originally proposed) to Records Management

- –These services are closely related to Records Catalog and Records Life Cycle Data management

Considered and decided against using a COTS Records Management Application (RMA)

- –ERA's requirements are specialized, and not a natural fit for most COTS RMA products
- –RMA products often capture key data and metadata in proprietary and inaccessible formats, which would degrade the system's evolvability and persistence

Decided to use a façade pattern for managing assets

- –Allows the various Manage Asset services to call upon common services within LS&C while including specialized methods

Decided to include a formal business rules layer

- –Allows business rules to be centrally defined and managed

Records Management Physical Design

Records Management services are implemented on the System/Business Applications VLAN, which is described in the Local Services & Control Design charts

Services developed as J2EE Web Services, with support from the common COTS-based infrastructure in LS&C

Records Management Conclusions

Centralize records catalog and life cycle management
- Maintains chain of custody and provenance
- Ensures consistency and integrity of records
- Provides a guide to every asset within ERA
- Driven by NARA business rules and policies

Leverage Templates
- Promote automation and consistency in archival processes
- Provide a key aspect of the "self-describing" archives

Implement Deterministic Disposition Instructions
- Automate disposition processes
- Ensure consistent and timely disposition processes

Provide Redaction Framework
- Encapsulates proprietary details of COTS products
- Allows new redaction engines to be added over time

Records Management RIDs

RID-LMC00125 Redaction of Data Types

RID-LMC00128 Redaction COTS

RID-LMC00138 Automatic Description of Descriptive Data

RID-LMC00139 Access Review

RID-LMC00140 Redaction Data Flow

END DAY TWO