Day Two

ERA System Design Review

May 10, 2005
Data Model

ERA SDR - DAY TWO

May 10, 2005
Data Replication
Logical Model
Conceptual Model
Notation
Methodology

Data Model - Agenda
Data Model - Methodology
Data Model - Methodology
Data Model - Methodology - Physical Model
new concepts.

A stereotype is a way to extend the core semantics of UML to express

Packages are stereotyped to identify their hierarchical level

Packages used to represent groups of objects and NAMA constructs

Suitable to represent hierarchical nature of the data model

UML notation used throughout the data architecture and design

Data Model – Notation
From lower level models includes dependencies derived from NARA data.

UML package notation with stereotype: <<Data Category>>

Only included those relevant to Enterprise Architecture.

Lockheed Martin Proprietary Information

Data Model - Conceptual Model
simple object model
- Dependencies derived from stereotype $<\text{subject area}>$
- UML package notation with architecture
- Those identified in NARA data
- Not always a one-to-one with architecture
- ERA from NARA data
- Only included those relevant to optimizing data categories

Subject Areas

Data Model - Conceptual Model
Key UML class diagrams contained

UML class notation

Simple objects represented using

Starting point for the logical model

Collection of lower level objects

A simple object represents a

each subject area decomposed

Simple Object Model Example

Data Model - Conceptual Model
Logical Model will be further decomposed, fully attributed and normalized as future design increments. The logical model represented by packages at the conceptual model are simple objects from the operations diagram notation - described by UML model.

Logical Model is a decomposition and a normalized object model consists of complex object model.

Data Model – Logical Model
Disposition Agreement Group

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

Provence

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

Complex Object Model Highlights

Data Model - Logical Model
Templates are extended hierarchically through inheritance

Object for relating data files to each record manifestation

The physical view of a record

Digital Adaptation

Notion of a manifestation object

Object to manage copies and versions of records

Electronic Record Version Control

Complex Object Model Highlights

Data Model - Logical Model (continued)
System Data Store

- Contains system management data such as logs, inventory data, etc.
- Several orders of magnitude smaller than the electronic archives
- Assets (excluding the records themselves)
- Records catalog, search indices and instance operational data
- A set of relational and object databases containing:

Instance Data Store

- Records prevented potential system contamination or viruses and misclassifications
- Electronic records transferred undergoing ingest processing

Ingest Working Storage

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

Electronic Archives

ERA system comprised of four persistent stores

Data Model - Data Stores
Replication of Records is managed using Active Store.

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

Replication is required among other instances is unavailable
- Each instance can operate with limited degradation
- Operational data
- Each instance has a current copy of mission, security and

Replication ensures

ERA is a distributed system and requires data replication

Data Model - Replication
Note: Replication of documentary materials does not include records

network traffic

"On update" reduces "On create" or

Replication "On create" or

Replication ensures that all current data instances have the same

Replication can be created or updated at any instance

Documentary materials data

Example of instance

Data Model - Replication
Data Model - Replication

- Concurrency is restored
- Information will occur once the propagation of new or updated
  information can continue to operate
- If access to the master is lost the master copy to each instance
  replication of user data is from a
  master
  asset is lost
  user access to services and
  available. If security data is lost, security data must be highly
  master: Example of flow down from
and system level auditing occur for a consolidated view.

Replication to master must occur at each instance event and audit logs are created including data including:

Example of flow up to master:

**Data Model - Replication**
- Variety Volume replicator for the systems
- Replication based on reading the transaction logs
- Guest Shareplex in conjunction with Oracle database
- COTS solution

LM has chosen to implement asynchonous replication for ERA

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

- Synchronous replication

**Data Model - Replication**
These Data Models LM and NARA will collaborate on the development of further performance modeling and testing. Replication models refined in Increment 1 based on Physical models defined in Increment 1. Logical models refined in Increment 1. Architecture flows down from NARA’s Data Model – Conclusions/Forward Plan.
Local Services and Control Design (LSC)

ERA SDR - DAY TWO

May 14, 2005
RID Discussion
Initial Product Selections
Physical Design
Service Design
Functional Architecture
Key Requirements
Description of Functionality

Agenda
Local Services & Control Design
Global Unique Identifier (GUID) management
- Interfaces between ERA Instances
- External interfaces
- Security services
- Service orchestration and mediation
- User interface portal
- Local services and control provides

LSAC Description
service registration and management
LM30 – Provide a common facility for collaboration tools
LM24.2 – Include asynchronous
LM21 – Use web portal architecture
LM21 – Use web portal architecture
applicances
from user networks, including security
LM14 – Include a single point of access
and assets
authorized access to all system services
LM13 – Require authenticated and
services
management service to orchestrate
LM2 – Mediation and business process

LSAC Key Requirements
Portal Framework

- Identity Management is centralized in ERA Management
- Language (SAML) standard
- Includes Single Sign-On that follows the Security Assertion Markup Language (SAML) standard
- For Directory Services
- Follows the Lightweight Directory Access Protocol (LDAP) standard
- Authentication, Authorization, and Authorization
- Intrusion detection and Intrusion detection systems (IDS) provide
- Router, firewalls, and Intrusion Detection Systems (IDS) provide
- Perimeter security

LSAC Design Highlights
collaboration functionally

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

Collaboration

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

Enterprise Content Management (ECM)

LSAC Design Highlights
stores
- Provides support for relational, object-oriented, and file-based data

Data Service
- Provides for the storage and management of persistent data

Other ERA component
- Each orchestration is packaged as a J2EE enterprise application
- Each orchestration is deployable independently of any
- Orchestrations follow the Business Process Execution Language
- The LM Team design uses orchestrations to encode business

Orchestration point model would yield:
- Mediation
- Includes queues, which provide a distributed messaging
- Includes mapping the logical data model to the physical data model

LSAC Design Highlights
Development of Orchestrations

LSAC Design Highlights –
Lockheed Martin Proprietary Information

ERS System Design Review - Day 2

- SOA Infrastructure includes mediation, queues, and security

- Services within ERA

- Interface partners use the same SOA Infrastructure as all of the Interfaces with External Systems

LSAC Design Highlights
GUID used to validate file integrity
- Relationships can be reconstructed from archival storage
  - In a self-describing manner
  - In Records Catalog
    - Relationships held:
      - Relating records and their lifecycle metadata
      - Relating original, POE, redacted versions
      - Collecting files into records
    - Relationships include:
      - Management includes relating assets
        - High statistical probability of uniqueness
        - 256 bits = 1 x 10^27 values
        - NIST standard algorithm
      - GUID implementation is SHA256 hash
        - GUID is a File identifier

Create and Manage GUID
LSAC Design Details
Delegated administration of authorization
Authentication, identification, and authorization
Choose integrated identity management suite
Secure, scalable, mature application framework
Choose J2EE over .NET or custom
Leverage COTS products to provide core infrastructure services
Centralized
Core services to manage business service components
ERA Management provides:
Distributed
Core services that are leveraged by business service components
LSSC provides:
Collected common infrastructure services into core packages
Implementation a Service Oriented Architecture

LSSC Design Trades
LS4C Physical Design
hosts workstations for media production
archival assets ensures long-term continuity of data
- clustering, replication, backup, and ability to restore from
- persistence of long-lived, non-record assets
- contains an instance data store
- balancing, intrusion detection, and firewalls
- provides a main network switch that contains integrated load
- configurations
- scales vertically and horizontally to larger or smaller
- clustering
- includes dedicated servers where needed for performance or
- scalability, operations, and management
- physical server in a manner that is convenient for deployment,
- allows unrelated service components to be hosted on the same
- includes partitioned servers
- includes the main business process components of the ERA System

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

Web Server VLAN

LSAC Physical Design
Application Services
LS4C abstracts the core infrastructure from the Business
management
- User authentication, identification, authorization, and session
- Relational and object data management
- Business process mediation and orchestration
- Workbench and collaboration

LS4C uses COTS frameworks for:
- Business Process and Business Rules Management
- Business Process and Business Rules Management
- Core service components that are leveraged by business services
- Security services
- Distributed enterprise-wide infrastructure services

LS4C provides:

Conclusions
RID-LMCO0136 Primary COTS Selections

RID-LMCO0133 GUIDS in the Storage Architecture

RID-LMCO0131 Creation and Monitoring of Workflows

RID-LMCO0130 COTS Workflow COTS

RID-LMCO0116 COTS use for LSAC and Management

RID-LMCO0115 Missing LMC Requirements in the SADD

LS4C RIDS
Ingest Design

ERA SDR - DAY TWO

Rick Rogers
May 10, 2005
RID Discussion

Physical Design

Ingest Modelling

Design Highlights / Trades

Service Design

Functional Architecture

Key Requirements

Description of Functionality
Validates transfer

Performs virus scans

Prepares those electronic records for storage within the ERA System

Entitles

Mechanisms to receive the electronic records from the transferring

Ingest provides ingest description
to support large ingest volumes.

LM31.1 - Fully automated ingest process, determined level and access restrictions are segregated until their security access storage so that transferred records are managed transfers.

LM3.7 - Segregated ingest working

LM1.8 - Ingest tools that are used to

Ingest Key Requirements
Key Features:

- Ingest Process
  - Automated
  - Fully Segregated

- Storage
  - Ingest Working

- Entitles
  - Tools for Transferring

Ingest Functional Architecture
Lockheed Martin Proprietary Information

**ERA System Design Review - Day 2**

**Legend:**
- Common Infrastructure
- Business Application
- Supporting Business Application Services
- Workbench
- Ingest includes

Ingest within SOA
and validated

- Transfers remain in Ingest Working Storage until they are screened
- Segregated Ingest Storage
discovered later

artifacts so that transfers can be placed into Archival Storage and

- Ingest automatically captures descriptive data from all available
Automated Descrptive Data

Agreements

- Requests are automatically validated against Transfer
Automated Transfer Management

- Developing archival metadata deferred to Records Management

- Performing automatically

- Virus scanning, initial security review, and transfer validation
Automated Ingest

Ingest Design Highlights
Conducting research on automating extraction of descriptive data

Conducting targeted entity extraction

Conducted implementations client-side tools to facilitate packaging

Considered high-volume ingest

Records Management

Moved archival description from ingest (as originally proposed) to

Ingest

Design

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

Ingest Physical Design
which maximizes system security

Segregates the ingest of records into ERA from all other functions

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

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

Automated Descriptive Data

Ensures ingested records are validated, virus scanned, and security

Supports high-volume ingest

Automated Ingest

Ingest Conclusions
Records Management Design

ERA SDR - DAY TWO

May 10, 2005

Rick Rogers
Records Management Agenda
Records Management Description

Records Management provides

- Management functionality for FOIA and Privacy Act requests
- Selected management functionality for non-electronic records
- Capability to implement disposition instructions processes
- Management functionality for access review and redaction
  - Management of records life cycle data, descriptions, and arrangements
  - Management of instructions, appraisals, transfer agreements, templates, authority
  - Management of electronic records and other assets within the ERA System
- Services necessary to manage the archival properties and attributes of the electronic records and other assets within the ERA System
managing FOIA and Privacy Act Request
LM2.10, and LM2.10 – Common approach of record life cycle
review and perform redaction in context
LM17, LM18 – Facility to perform access
records life cycle data and transactions
LM5 – Centralized facility to manage all
approach to managing assets
LM1, LM3, LM4, LM7, LM9 – Common
Records Management Key Requirements
Architecture

Records Management Functional

Key Features:
- Centralized
- Access Review
- Management
- Approach
- Asset
- Disposition
- Instructions
- Centralized
- Redaction

ERA System Boundary
Preservation and Service Level Plans – Authority Sources leveraged throughout the system, such as in Authority Sources

- Supports both hierarchical and network topologies.

Authority Sources

- Responds to these questions and provides a guide to every asset within ERA

Interview-Style User-Interfaces

- Provides a guide to every asset within ERA

Centralized Records Catalog

- Ensures consistency, integrity, and authenticity of the Records as they are managed within the system

Centralized Records Life Cycle Management

Highlights

Records Management Design
Dependence on specific hardware and software technologies

- Includes an approach for ensuring the archives are free from

Persistent Archives

- Includes an approach to identifying assets in a permanent and

Persistent Identifiers

- Includes an event database for event-driven instructions

Deterministic Disposition Instructions

- Includes a template editor framework

- Supports Inheritance and context hierarchies

- Leverages templates throughout ERM throughout archival and system

Highlights

Records Management Design
- Allows business rules to be centrally defined and managed.
- Decided to include a formal business rules layer.
- Services within LSAC while including specialized methods.
- Allows the various Manage Asset Services to call upon common
- Decided to use a facade pattern for managing assets.
- Evolvability and persistence.
- Inaccessible/formats, which would degrade the system's
- COTS RMA Products.
- RMA products often capture key data and metadata in proprietary.
- ERAs' requirements are specialized, and not a natural fit for most.
- Management Application (RMA).
- Considered and decided against using a COTS Records
- Life Cycle Data management.
- These services are closely related to Records Catalog and Records
- Originally proposed (to Records Management
- Moved Access Control and Redaction from Dissemination (as

Records Management Design Trades
Records Management Physical Design

the common COTS-based infrastructure in LSAC

Services developed as J2EE Web Services, with support from

the Local Services & Control Design charts

System/Business Applications VLAN, which is described in

Records Management services are implemented on the

Records Management Conference