Aligning OAIS with the Enterprise Architecture

8th European Conference on Digital Archiving, 2010
Geneva, Switzerland
Outline

• Digital Preservation as a Problem
• Context
• The Enterprise Architecture Perspective
  • Zachman Framework
  • TOGAF
• Reference Architecture
  • Shaman RA
• OAIS Reference Model
  • Modelling OAIS
• Conclusions
Generic and common requirements:

- **Integrity**: Effective preservation requires that the informational content of objects remains unchanged through its lifetime.

- **Reliability**: A copy (or representation) of any preserved object must survive over its system’s lifetime.

- **Authenticity Assurance**: A future consumer may require the accessed information to be trustworthy.
Digital Preservation as a Problem (2/2)

• Provenance: A future consumer may require information concerning the origins of the object.

• Dealing with Obsolescence: Digital objects should be able to be exploited independently of any technological context (ideally…).

• Scalability: Digital preservation systems might be required to face technological evolution through the addition of new components.

• Heterogeneity: Digital preservation system’s components should be heterogeneous due to technology disruption.
Context

(http://grito.intraneia.pt)
- National project
- **Exclusive storage** clusters (dedicated to digital preservation)
- **Extended storage** clusters (using surplus resources of computing clusters)

**SHAMAN** - Sustaining Heritage Access through Multivalent ArchiviNg
(http://shaman-ip.eu/shaman)
- European project
- Three domains of focus: memory institutions, **engineering** and **e-Science**
- Strong focus on authenticity and integrity
- Definition of frameworks and architectures for digital preservation

**Common ground:** use of data grids (massive data sets, file management, user management, networking etc.)
A reference architecture presents a way of recording a specific body of knowledge, with the purpose of making it available for further practical reuse.

According to the ANSI/IEEE Std. 1471-2000:

architecture is "the fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution"

Therefore, a reference architecture for digital preservation must provide a way to capture the knowledge in the domain, so that it can be instantiated in concrete architectures for real system implementations!
Generic focus (a model based on generic requirements and assumptions…).

- The SHAMAN DoW
- The initial work…
- …

- SOA…
- TRAC criteria…
- …
SHAMAN RA
Information Lifecycle (1/2)
The context of the business

The interfaces of the digital preservation system

The digital preservation system
### From the lifecycle context

**A taxonomy of vulnerabilities and threats to digital preservation (1/2)**

<table>
<thead>
<tr>
<th>Vulnerabilities</th>
<th>Threats</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>Software Faults</td>
<td>Software Obsolescence</td>
</tr>
<tr>
<td>Data</td>
<td>Media Faults</td>
<td>Media Obsolescence</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>Hardware Faults</td>
<td>Hardware Obsolescence</td>
</tr>
<tr>
<td></td>
<td>Communication Faults</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Network Service Failures</td>
<td></td>
</tr>
<tr>
<td>Disasters</td>
<td>Natural Disasters</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human Operational Errors</td>
<td></td>
</tr>
<tr>
<td>Attacks</td>
<td>External Attacks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Internal Attacks</td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>Organizational Failures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Economic Failures</td>
<td></td>
</tr>
<tr>
<td>Business Requirements</td>
<td>Legal Requirements</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stakeholders’ Requirements</td>
<td></td>
</tr>
</tbody>
</table>
## From the lifecycle context

A taxonomy of vulnerabilities and threats to digital preservation (2/2)

<table>
<thead>
<tr>
<th>Vulnerabilities</th>
<th>Process</th>
<th>Data</th>
<th>Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Software Faults</td>
<td>Media Faults</td>
<td>Hardware Faults</td>
</tr>
<tr>
<td></td>
<td>Software Obsolescence</td>
<td>Media Obsolescence</td>
<td>Hardware Obsolescence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Communication Faults</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Network Service Failures</td>
</tr>
<tr>
<td>Threats</td>
<td>Disasters</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Natural Disasters</td>
<td>Human Operational Errors</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Attacks</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>External Attacks</td>
<td>Internal Attacks</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Management</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Organizational Failures</td>
<td>Economic Failures</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Business Requirements</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Legal Requirements</td>
<td>Stakeholders’ Requirements</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T</th>
<th>T</th>
<th>O</th>
<th>O</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>O</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>O</td>
<td>O</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>O</td>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>

*Notes: T = Technology, O = Organizational, C = Cultural.*
Technology + Organization + Context
= Enterprise Architecture
The Zachman Framework is not a methodology for creating the implementation (an instantiation) of the object. The Zachman Framework is the ontology for describing the Enterprise. The Framework (ontology) is a STRUCTURE whereas a methodology is a PROCESS. 

"The Zachman Framework is not a methodology for creating the implementation (an instantiation) of the object. The Zachman Framework is the ontology for describing the Enterprise. The Framework (ontology) is a STRUCTURE whereas a methodology is a PROCESS. “
TOGAF - The Open Group Framework

Welcome to TOGAF™ Version 9 "Enterprise Edition"

to the Downloads page

About TOGAF™

TOGAF is an industry standard architecture framework that may be used freely by any organization wishing to develop an architecture for use within that organization.

TOGAF has been developed and continuously evolved since the early 1990s by representatives of some of the world’s leading IT customer and vendor organizations, working in The Open Group’s Architecture Forum. Details of the Forum, and its plans for evolving TOGAF in the current year, are given on the Architecture Forum web site.

About TOGAF Version 9 Enterprise Edition

TOGAF Version 9 Enterprise Edition ("TOGAF 9") is a detailed method and set of supporting resources for developing an Enterprise Architecture. Developed and endorsed by the membership of The Open Group’s Architecture Forum, TOGAF 9 represents an industry consensus framework and method for Enterprise Architecture that is available for use internally by any organization around the world, members and non-members of The Open Group alike - subject to license conditions - see Downloading TOGAF 9.

As a comprehensive, open method for Enterprise Architecture, TOGAF 9 compliances, and can be used in conjunction with, other frameworks that are more focused on specific aspects of architecture or for vertical sectors such as Government, Defense, and Finance.

Learn more about TOGAF 9 from the following white papers:
- TOGAF 9: Introduction
- TOGAF 9: Migration Overview
- visit the TOGAF 9 information site

About TOGAF Certification

Certification from The Open Group

Why Become Certified?

For architecture service providers and tools vendors, the new certification program provides a way to demonstrate clearly how their services and products support the Enterprise Architected using TOGAF.

For individual Enterprise Architects, TOGAF certification demonstrates clearly to employers and peers their commitment to Enterprise Architecture as a discipline. In particular, it demonstrates that they possess a body of core knowledge about TOGAF as an open, industry standard framework and method for Enterprise Architecture.

The Open Group publishes the definitive register of TOGAF certified individuals, and certifies service and product offerings, and issues certificates that can be used by vendors in promotion.

How to Become Certified?

You may choose how to prepare for the TOGAF 9 certification examinations, whether that is by self-study, attending a training course, or a combination of both. To pass the examination, you must answer 110 questions in 2.5 hours.

Visit the register of Accredited TOGAF Certification Training Providers.
TOGAF overview

ADM

Requirements Management

Architecture Content Framework

ADM Guidelines & Techniques

Enterprise Continuum

Reference Models

Architecture Capability Framework
The SHAMAN Reference Architecture

- Part 1 Framework
  - depends on
  - supports

- Part 2 Process
  - draws on
  - depends on

- Part 3 Foundations
- Part 4 Glossary
The SHAMAN Reference Architecture Part 1 – Framework, which describes the architectural framework and respective viewpoints;

The SHAMAN Reference Architecture Part 2 – Process, which describes the process for the development of preservation architectures derived from the Reference Architecture;

The SHAMAN Reference Architecture Part 3 – Foundations, which describes the foundations of this work and provides references for the instantiation of concrete architectures;

The SHAMAN Reference Architecture Part 4 – Glossary, which contains definitions for the main terms used in this Reference Architecture.
# Viewpoint Framework

<table>
<thead>
<tr>
<th>Stakeholders</th>
<th>Decision Making</th>
<th>System Building and Operation</th>
<th>Viewpoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designated Community</td>
<td>Regulator</td>
<td>Requirements and Conformance</td>
<td>Preservation Strategic Planning</td>
</tr>
<tr>
<td>Auditor</td>
<td>Preservation Manager</td>
<td>Business Governance</td>
<td></td>
</tr>
<tr>
<td>Organization Manager</td>
<td>Technology Manager</td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Designer</td>
<td>Technology Provider</td>
<td>System Building and Support</td>
<td></td>
</tr>
<tr>
<td>Technology Operator</td>
<td>Preservation Operator</td>
<td>Acting and Operation</td>
<td></td>
</tr>
<tr>
<td>Producer</td>
<td>Consumer</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Structural View
Moving beyond

...moving from an informal way of expressing (OAIS Reference Model Figure F-1: Composite of Functional Entities)...

...to a more appropriately formal, traceable and objectively represented meta-model...
The process

1. Preservation Strategic Planning
2. Business Governance
3. Acting and Operation
4. System Building and Support
5. Architecture Realization

Requirements and Conformance

4.1. Data
4.2. Applications
4.3. Technology

4. Data
4. Applications
4. Technology

3. Acting and Operation
2. Business Governance
1. Preservation Strategic Planning
5. Architecture Realization

4. System Building and Support
Modelling examples (1/4)

UML

- **OAIS**
  - Preservation Planning
  - Data Management
    - Database
    - Access
  - Archival Storage
    - Media
    - AIP (from Archival Storage)
- **Descriptive Information**
  - SIP «flow»
  - AIP «flow»
- **Producer**
  - Ingest «flow»
- **Consumer**
  - DIP «flow»
- **Management**
Modelling examples (2/4)
BPMN
Modelling examples (1/4)

UML

act Ingest Activity

Producer | Ingest | Administration | Data Management | Archival storage
---|---|---|---|---
Submit SIP | Receive submission | | | |
Modelling examples (4/4)
BPMN

BPMN Ingest

Receive SIP from Producer

Check SIP errors

Quality assurance

Generate AIP

Generate descriptive info

Send storage request

Send database update request

Data base update confirmation

Receive database update request

Update database

Send database update confirmation

Receive audit report request

Send audit report

Receive audit report

Generate audit report

Receive audit report

Send audit report

Receive report request

Generate report

Receive report

Report received

Send report

Receive report

Generate report

Receive audit report

Send audit report

Receive report

Generate report

Receive audit report

Send audit report

Receive report

Generate report

Receive audit report

Send audit report

Receive report

Generate report

Receive audit report

Send audit report

Receive report

Generate report

Receive audit report

Send audit report

Receive report

Generate report

Receive audit report

Send audit report
Deployment example

- Policy: Description
- Process Modeling Tool
- Service Orchestration
- ESB: SOAP/REST over HTTP
- Services: ex: Enterprise Architect, Eclipse BPMN, XML Editor, Text Editor...
- Processes: Specification
- Process Execution Language Generator
- Service wrapper: ex: JBOSS, jBPM, Apache ODE, ...
- Data Grid: iRODS, ...
- Database: Oracle, MySQL, ...
- Legacy Digital Library System: DSPACE, Kopal, ...
- Processes: ex: BPMN, AGWL, UML Activity Diagrams, Petri net, DAG...
- Service: ex: BPEL, C-GWL, jPDL, ...
- Ex: Text, MS Word, PDF, XML, ...
- Ex: Specification
- JEXL, BAM, AGWL, XML, ...
- Processes: «flow»
- Service Execution Language: Specification
- Service:
- Ex: Search & Browse Integration Service, ...
- Ex: JBOSS, jBPM, Apache ODE, ...
- ESB: SOAP/REST over HTTP
- Service wrapper
- Data Grid
- Database
- Legacy Digital Library System

Instituto de Engenharia de Sistemas e Computadores Investigação e Desenvolvimento em Lisboa
Conclusions

• Digital preservation is a very complex problem!!! Therefore:
  
  • We surveyed the main requirements to digital preservation and classified the threats and vulnerabilities that might endanger preservation using a taxonomy of threats and vulnerabilities.
  
  • We propose the alignment of OAIS with the Enterprise Architecture
  
  • We propose a process “inspired” by TOGAF to develop create preservation architectures.
José Barateiro – jbarateiro@lnec.pt
Gonçalo Antunes – goncalo.antunes@ist.utl.pt
José Borbinha – jlb@ist.utl.pt