A Capability Model for Digital Preservation

Analyzing Concerns, Drivers, Constraints, Capabilities and Maturities

Christoph Becker
Vienna University of Technology
Vienna, Austria
www.ifs.tuwien.ac.at/~becker

Gonçalo Antunes,
José Barateiro,
Ricardo Vieira
INESC-ID Information Systems Group, Lisbon, Portugal
Digital Longevity

- Numerous reference models, frameworks and concepts
  - OAIS and trust: TRAC, RAC (ISO 16363), NESTOR...
  - Records Management: MoReq, ISO 15489...
  - Risk: DRAMBORA...
  - Planning: PLATO, PLATTER
  - Economics: BRTF, LIFE...

- Yet, we still lack a holistic view
  - Maturity of the field is unclear and evolving
  - Integration into Information Systems and Information Technology fields is unclear
  - How does Digital Preservation relate to, e.g., IT Governance?
  - How can we assess and improve organizational capabilities?
  - How can we deal with non-repository scenarios?
About systems, requirements and preservation

Scenarios of systems and their perceived relevance of digital preservation requirements

The Digital Preservation System (DPS):
DP as *functional requirements*

The Systems of Systems (SoS): Business system delegates DP responsibility to a DPS

The “Digital Preservation Ready” System (DPR):
Longevity as a *non-functional requirement*!
Background

• Enterprise Architecture (EA)
  – Holistic system architecture approach to information systems and technology in organizations

• IT Governance
  – decision making and communication within IT-supported organizations
  – leadership, organisational structures and processes
  – ensure that the IT sustains the organisation’s objectives
  – COBIT: Control Objectives for IT

• The goals of Reference Architectures
  – Process
  – Stakeholder concerns
  – Independent of business domain and organization
Make DP ubiquitous in systems

• Consider clearly defined Goals...
  – Business process
  – Stakeholder concerns

• Align business and technology according to the best references in Enterprise Architecture (EA)...
  – TOGAF Architecture Development Method

• Follow IT Governance best practices...
  – COBIT (goal-oriented, process-oriented, control-based)
  – Maturity Model based on CMM

• Define a Reference Architecture for reuse in the EA processes where DP is a relevant concern
Consolidation of DP Reference Models
Contextualization of a DP Architecture
Modeling of DP Capabilities
Creation of a DP Architecture Vision

SHAMAN-RA

CONCRETE ARCHITECTURES

- H. Architecture Change Management
- G. Implementation Governance
- F. Migration Planning
- E. Opportunities and Solutions
- D. Technology Architecture
- C. Information Systems Architecture
- B. Business Architecture
- Requirements Management
A Capability-based Reference Architecture

Domain Knowledge
- SHAMAN-RA v1.0
- OAIS
- TRAC/RAC
- TDR 2002
- NESTOR
- Planets Planning method
- Planets Functional Model
- PREMIS
- BRTF Sustainability Report
- DRAMBORA
- PARSE.Insight
- ...

Standards and Best-Practices
- OMG UML
- OMG BMM
- OMG SBVR
- OMG OSM
- ISO 27000: Security
- IEEE Std. 1471-2000
- Zachman Framework
- COBIT
- DoDAF
- ...

Stakeholders
Concerns
Influencers
Goals
Capabilities
Metamodel of key elements

- Stakeholder
- Concern
- Key Question
- Influencer
- Constraint
- Driver
- Goal
- Desired Result
- Capability
- Key Performance Indicator
- Course of Action
- Maturity

Relationships:
- Stakeholder has Concern
- Concern expressed by Key Question
- Concern relates to Influencer
- Constraint can be Driver
- Constraint limits deployment of Capability
- Capability delivers Maturity
- Key Performance Indicator quantified by"
Digital Preservation Capabilities

A capability is an “ability that an organization, person, or system possesses. Capabilities are typically expressed in general and high-level terms and typically require a combination of organization, people, processes, and technology to achieve”

A capability can control, inform, include, or depend on another capability
Digital Preservation Capabilities

A **capability** is an “ability that an organization, person, or system possesses. Capabilities are typically expressed in general and high-level terms and typically require a combination of organization, people, processes, and technology to achieve”

**Preserve Contents** is the ability to maintain content authentic and understandable to the defined user community over time and assure its provenance.
### Core Preservation Capabilities

**Preservation Planning**  
Monitor, steer and control the preservation operation of content

**Preservation Operation**  
Control the deployment and execution of preservation plans.
Core Preservation Capabilities

Preservation Planning

- Monitor, steer and control the preservation operation of content
- Controls
- Influencers and Decision making
- Options diagnosis
- Specification and delivery
- Monitoring
- Inform

Preservation Operation

- Control the deployment and execution of preservation plans.
### Core Preservation Capabilities

**Preservation Planning**

- Monitor, steer and control the preservation operation of content
- • Influencers and Decision making
  • Options diagnosis
  • Specification and delivery
  • Monitoring

**Preservation Operation**

- Control the deployment and execution of preservation plans.
- • Analyze content
  • Execute preservation actions
  • Ensure adequate provenance trail
  • Handle preservation metadata
  • Conduct Quality Assurance
  • Provide reports and statistics
Core Preservation Capabilities

<table>
<thead>
<tr>
<th>Preservation Planning</th>
<th>Preservation Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitor, steer and control the preservation operation of content</td>
<td>Control the deployment and execution of preservation plans.</td>
</tr>
</tbody>
</table>
| • Influencers and Decision making  
• Options diagnosis  
• Specification and delivery  
• Monitoring | • Analyze content  
• Execute preservation actions  
• Ensure adequate provenance trail  
• Handle preservation metadata  
• Conduct Quality Assurance  
• Provide reports and statistics |
| “Migrate this set of images (in TIFF-5) to JP2 using ImageMagick 6.3 with parameters a,b,c” | • Analyze original  
• Migrate, analyze output  
• Conduct quality assurance  
• Provenance, metadata, Reporting |
Preservation Planning Capabilities

• Planning Operational Preservation:
  – make drivers and goals operational and assess options against these criteria to deliver efficient decisions and operational plans
  1. Influencers and Decision Making
  2. Options Diagnosis
  3. Specification and Delivery

• Monitoring
  1. Internal Monitoring
  2. External Monitoring
Preservation Operation Capabilities

• Analysis
• Action
• Quality Assurance
• Preservation Metadata
• Plan Deployment
• Reporting and Statistics
Example: Measuring Reporting and Statistics

• Timeliness
• Currentness
• Completeness
• Relevance
• Correctness
• Understandability
Analysing drivers and constraints

- Stakeholders concerned
- Concerns addressed
- Goals impacted
- Capabilities affected
- Metrics applicable
A Capability Maturity Model for Preservation Operation

The CMM has been shown to be a powerful instrument for assessment and improvement in Software Engineering

<table>
<thead>
<tr>
<th>Awareness and Communication</th>
<th>Policies, Plans and Procedures</th>
<th>Tools and Automation</th>
<th>Skills and Expertise</th>
<th>Responsibility and Accountability</th>
<th>Goal Setting and Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Initial / ad-hoc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Repeatable, but Intuitive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Defined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Managed and Measurable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Optimized</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Awareness and Communication</td>
<td>Policies, Plans and Procedures</td>
<td>Tools and Automation</td>
<td>Skills and Expertise</td>
<td>Responsibility and Accountability</td>
</tr>
<tr>
<td>---</td>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Need for operations recognized</td>
<td>Some uncontrolled, undocumented operations</td>
<td>Some tools employed ad-hoc</td>
<td>No awareness of skills/expertise</td>
<td>None assigned</td>
</tr>
<tr>
<td>2</td>
<td>Awareness of role of operations, no formal reporting, individual communication</td>
<td>Operational procedures emerge, informal, intuitive. Individual procedures, ad-hoc QA</td>
<td>Automated tools used incoherently based on need and availability</td>
<td>Hands-on experience, informal peer training</td>
<td>Responsibility emerges, undocumented, no accountability</td>
</tr>
<tr>
<td>3</td>
<td>Understanding of role of operations, reporting guidelines not enforced</td>
<td>Defined process, driven by components and services.</td>
<td>Plans deployed according to spec, but mostly manual</td>
<td>Formal training plan, but training still individual initiative</td>
<td>Responsibility assigned, accountability not always provided</td>
</tr>
<tr>
<td>4</td>
<td>…</td>
<td>automated system controls automated operations</td>
<td>skills and expertise defined for all roles</td>
<td>Responsibility and accountability defined for all roles</td>
<td>Metrics aligned with goals, compliance enforced</td>
</tr>
<tr>
<td>5</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>RACI plan fully traceable to ops</td>
</tr>
<tr>
<td>Awareness and Communication</td>
<td>Policies, Plans and Procedures</td>
<td>Tools and Automation</td>
<td>Skills and Expertise</td>
<td>Responsibility and Accountability</td>
<td>Goal Setting and Measurement</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>----------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td><strong>1</strong> Need for operations recognized</td>
<td>Some uncontrolled, undocumented operations</td>
<td>Some tools employed ad-hoc</td>
<td>No awareness of skills / expertise</td>
<td>None assigned</td>
<td>Operations react to incidents, not tracked</td>
</tr>
<tr>
<td><strong>2</strong> Awareness of role of operations, no formal reporting, individual communication</td>
<td>Operational procedures emerge, informal, intuitive. Individual procedures, ad-hoc QA</td>
<td>Automated tools used incoherently based on need and availability</td>
<td>Hands-on experience, informal peer training</td>
<td>Responsibility emerges, undocumented, no accountability</td>
<td>Individual awareness of short-term goals, no consistent definition</td>
</tr>
<tr>
<td><strong>3</strong> Understanding of role of operations, reporting guidelines not enforced</td>
<td>Defined process, driven by components and services.</td>
<td>Plans deployed according to spec, but mostly manual</td>
<td>Formal training plan, but training still individual initiative</td>
<td>Responsibility assigned, accountability not always provided</td>
<td>Operational goals specified, no formal metrics or alignment to goals</td>
</tr>
<tr>
<td><strong>4</strong> ...</td>
<td>...</td>
<td>automated system controls automated operations</td>
<td>skills and expertise defined for all roles</td>
<td>Responsibility and accountability defined for all roles</td>
<td>Metrics aligned with goals, compliance enforced</td>
</tr>
<tr>
<td><strong>5</strong> ...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>RACI plan fully traceable to ops</td>
<td>...</td>
</tr>
<tr>
<td>Awareness and Communication</td>
<td>Policies, Plans and Procedures</td>
<td>Tools and Automation</td>
<td>Skills and Expertise</td>
<td>Responsibility and Accountability</td>
<td>Goal Setting and Measurement</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>----------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>1 Need for operations recognized</td>
<td>Some uncontrolled, undocumented operations</td>
<td>Some tools employed ad-hoc</td>
<td>No awareness of skills / expertise</td>
<td>None assigned</td>
<td>Operations react to incidents, not tracked</td>
</tr>
<tr>
<td>2 Awareness of role of operations, no formal reporting, individual communication</td>
<td>Operational procedures emerge, informal, intuitive. Individual procedures, ad-hoc QA</td>
<td>Automated tools used incoherently based on need and availability</td>
<td>Hands-on experience, informal peer training</td>
<td>Responsibility emerges, undocumented, no accountability</td>
<td>Individual awareness of short-term goals, no consistent definition</td>
</tr>
<tr>
<td>3 Understanding of role of operations, reporting guidelines not enforced</td>
<td>Defined process, driven by components and services.</td>
<td>Plans deployed according to spec, but mostly manual</td>
<td>Formal training plan, but training still individual initiative</td>
<td>Responsibility assigned, accountability not always provided</td>
<td>Operational goals specified, no formal metrics or alignment to goals</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>automated system controls automated operations</td>
<td>skills and expertise defined for all roles</td>
<td>Responsibility and accountability defined for all roles</td>
<td>Metrics aligned with goals, compliance enforced</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>RACI plan fully traceable to ops</td>
<td></td>
</tr>
<tr>
<td>Awareness and Communication</td>
<td>Policies, Plans and Procedures</td>
<td>Tools and Automation</td>
<td>Skills and Expertise</td>
<td>Responsibility and Accountability</td>
<td>Goal Setting and Measurement</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Need for operations recognized</td>
<td>Some uncontrolled, undocumented operations</td>
<td>Some tools employed ad-hoc</td>
<td>No awareness of skills / expertise</td>
<td>None assigned</td>
<td>Operations react to incidents, not tracked</td>
</tr>
<tr>
<td>Awareness of role of operations, no formal reporting, individual communication</td>
<td>Operational procedures emerge, informal, intuitive. Individual procedures, ad-hoc QA</td>
<td>Automated tools used incoherently based on need and availability</td>
<td>Hands-on experience, informal peer training</td>
<td>Responsibility emerges, undocumented, no accountability</td>
<td>Individual awareness of short-term goals, no consistent definition</td>
</tr>
<tr>
<td>Understanding of role of operations, reporting guidelines not enforced</td>
<td>Defined process, driven by components and services.</td>
<td>Plans deployed according to spec, but mostly manual</td>
<td>Formal training plan, but training still individual initiative</td>
<td>Responsibility assigned, accountability not always provided</td>
<td>Operational goals specified, no formal metrics or alignment to goals</td>
</tr>
<tr>
<td>Automated system controls automated operations</td>
<td>Automated system controls automated operations</td>
<td>Automated system controls automated operations</td>
<td>Automated system controls automated operations</td>
<td>Responsibility and accountability defined for all roles</td>
<td>Metrics aligned with goals, compliance enforced</td>
</tr>
<tr>
<td>RACI plan fully traceable to ops</td>
<td>RACI plan fully traceable to ops</td>
<td>RACI plan fully traceable to ops</td>
<td>RACI plan fully traceable to ops</td>
<td>RACI plan fully traceable to ops</td>
<td>RACI plan fully traceable to ops</td>
</tr>
<tr>
<td>Awareness and Communication</td>
<td>Policies, Plans and Procedures</td>
<td>Tools and Automation</td>
<td>Skills and Expertise</td>
<td>Responsibility and Accountability</td>
<td>Goal Setting and Measurement</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>----------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>1叶</td>
<td>Need for operations recognized</td>
<td>Some uncontrolled, undocumented operations</td>
<td>Some tools employed ad-hoc</td>
<td>No awareness of skills/expertise</td>
<td>None assigned</td>
</tr>
<tr>
<td>2</td>
<td>Awareness of role of operations, no formal reporting, individual communication</td>
<td>Operational procedures emerge, informal, intuitive. Individual procedures, ad-hoc QA</td>
<td>Automated tools used incoherently based on need and availability</td>
<td>Hands-on experience, informal peer training</td>
<td>Responsibility emerges, undocumented, no accountability</td>
</tr>
<tr>
<td>3</td>
<td>Understanding of role of operations, reporting guidelines not enforced</td>
<td>Defined process, driven by components and services.</td>
<td>Plans deployed according to spec, but mostly manual</td>
<td>Formal training plan, but training still individual initiative</td>
<td>Responsibility assigned, accountability not always provided</td>
</tr>
<tr>
<td>4</td>
<td>...</td>
<td>...</td>
<td>automated system controls automated operations</td>
<td>skills and expertise defined for all roles</td>
<td>Responsibility and accountability defined for all roles</td>
</tr>
<tr>
<td>5</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>RACI plan fully traceable to ops</td>
</tr>
</tbody>
</table>
Conclusion

• Capability-based Reference Architecture
  – Supports deployment of preservation capability to any scenario
  – Enables customized, technology-independent architecture
  – Facilitates organizational change across technology generations

• Contextual Enterprise Architecture approach
  – Improves Separation of Concerns
  – Enables traceability

• Capability maturity model
  – Provides Decision support mechanism
  – Supports prioritizing improvements
  – Guides gap analysis
  – Facilitates management buy-in
Thank you!

• Questions?

Related reading: Control Objectives for DP: Digital Preservation as an Integrated Part of IT Governance published in the proceedings of the 74th Annual Meeting of the American Society for Information Science and Technology (ASIST)

www.ifs.tuwien.ac.at/~becker