Workflows in Digital Preservation
Overview

- Workflows
- Processes in Archives
- Workflow Modeling for Digital Archives
- Workflows in Digital Archives
  - Example: Open Archival Information System (OAIS) and “Producer-Archive Interface Methodology Abstract Standard”
- The State of Workflow
Workflow Roots

- Office Automation Systems (mid 1970s)
- Document Management Systems (mid 1980s)
- Email Applications
- Database Management
Definition: Workflow

“A workflow is a specific representation of a process, which is designed in such a way that the formal coordination mechanisms between activities, applications, and process participants can be controlled by an information system, the so-called workflow management system.”

What does this definition mean?

1. A formal process definition exists
2. This process definition is the result of process design
3. A workflow management system coordinates the steps in the process and assigns the participants as defined by the process definition
4. The participants can be (IT-) applications or humans.
Processes in Archives

- If defined, mainly for arrangement
- High degree of implicit knowledge
- Local flavours
- No standards
- Very adaptive
- Ever changing
Why Workflows?

- Massive amounts of similar data cannot be processed “by hand” (automation)
  - Quality management
- Authentic data needs an audit trail, documenting its handling (retraceability)
- Implicit knowledge becomes explicit (sharing knowledge)
  - Centralized implementation
  - Iterative development
Barriers for Workflows in Digital Archives

- Small number of digital archives
- Little interest
- Little research
- Implicit knowledge
- Small market
  - No “out of the box” solutions
  - DSpace implements a simple workflow
    - Fixed
    - Only for Ingest (data and metadata checks)
Rectulance against Workflow Management

- Some quotes:
  - “It’s just another hype”
  - “We’re low on resources already!”
  - “Will I still be able to do … ?”
  - “Another 'big brother' tool”
  - “This is too technical”
  - “We've always done it this way”
Workflows for Digital Preservation

Disadvantages:
- Initial costs
- Complexity
  - Efforts for definition
  - Set-up
- More control
- Too many tools and standards

Advantages:
- Documentation
- Automation
  - High throughput
  - Audit trail
- More control
- Sharing knowledge
- Connecting Applications
ikeep
Specific Features of Digital Preservation
Workflow Modeling

- Formal Semantics
- Flexibility
- Decomposition
- Simplicity
- Stability
- Retraceability
Formal Semantics

- Unambiguous design and enactment of workflows
- Textual definitions of semantics will change their meaning in time
- Mathematical backing needed
- Additional advantages:
  - Automatic model checks for certain properties of a workflow (e.g. deadlocks)
  - Simplified implementation
Flexibility

- Every workflow will change with time
- New technology will bring new possibilities
  - New possibilities should not require a new modeling language
  - Eases long term data management (of the workflows)
De decomposition

- Workflows should not be monolithic
- Vertical decomposition in modules that can be composed into some new workflow
- Horizontal decomposition to enable more implementational details at lower levels
- Reuse of tested building blocks
- Parallel development
Simplicity

- The principles should be easy to understand by current and future users (e.g. submitters, personnel and researchers)
- Better user acceptance because everybody can understand how it works
- Simple principles are harder to misinterpret in the long term
Stability

- The notion of workflows is relatively new
- Every change in semantics must be preserved (exponential complexity in data management)
- Workflow modeling standards have to show that they are here to stay
- Active research, development and use keeps standards alive
- Backwards compatibility is a must
Retraceability

- The audit trail should contain enough data to “relive” a workflow (including all routing decisions and error-handling)
- The workflow definition is still needed
- Basic requirement for authentic digital objects
iikeep  Workflows and the OAIS RM
What is the OAIS RM?

- The Open Archival Information System Reference Model (OAIS RM) is a unique standard for digital archives (ISO 14721:2003)
- Framework for
  - Terminology and concepts
  - Architecture of digital archives
    - Defines roles and responsibilities
    - Defines functional entities and logical data models
Implementation of OAIS Workflows

- OAIS RM defines multiple high-level workflows
- Scattered in the standard, as it only describes the functions and their interactions
  - Mainly data flows
- Act as guidelines
  - Implementation by the user
  - No definition of audit data
Example Ingest Workflow 1/2

- Defined by the OAIS RM
  1. Producer delivers Submission Information Package (SIP)
  2. Ingest checks SIP against contract (validation)
  3. Ingest creates Archival Information Package AIP(s) and Package Description(s)
  4. Ingest stores Package Description(s) (metadata) in Data Management
  5. Ingest stores AIP(s) in Archival Storage
  6. Ingest notifies the producer, that the data has been stored and that the may delete his data

ERPANET Workshop on Workflow / 13. October 2004 / Stephan Heuscher / p. 21
ikeep

OAIS Ingest Workflow

- High level
- Simple steps
- Circles and arrows
Example Ingest Workflow 2/2

- Refined step 2 (validation) by “Producer-Archive Interface Methodology Abstract Standard” (CCSDS 651.0-B-1)

- Steps:
  - Apply the validations
    - Systematic validation (after each transfer session)
    - In-depth validation (only on a coherent set of data)
  - If the validation has failed, the producer is notified.
**Validation Sub-Workflow**

- Same here
What is missing?

- Data flow
- Interfaces
- Audit Data
- Roles and resources
- Semantics
The State of Workflow in Preservation 1/2


The State of Workflow in Preservation 2/2

- No common “workflow basis” (à la relational algebra for relational databases)
- No single standard for workflow definition (à la structured query language (SQL) for relational databases)
- Are we too early?
Thank You

Contact:

- Stephan Heuscher
  stephan.heuscher@ikeep.com
  m +41 79 277 23 84
  t +41 31 998 42 80
  f +41 31 998 42 81

- ikeep Ltd.
  Digital Archives Services
  Morgenstrasse 129
  CH-3018 Bern
  Switzerland
Workflow Lifecycle 1/2

- Planning
  - Define process, organization, and data models
  - Define goals
- Implementation
  - Transform models for execution
  - Create interfaces to systems and users
Workflow Lifecycle 2/2

- Enactment
  - Coordinate control and data flow, resource assignments and application invocation
  - Monitor workflow enactment
  - Notify users
- Evaluation
  - Evaluate audit trail data