IMPLEMENTING INTERPARES: PROVIDING EVIDENCE OF THE AUTHENTICITY OF DIGITAL RECORDS

Adam Jansen, PhD
State Archivist
Hawai‘i State Archives

18 Feb 2020

Challenges of Digital Records

- Records are increasing created digitally
- Most systems designed to manage data, not records
- Records moved from one system to another
- Continuing research into systems-centric support of authenticity is warranted

- How to provide documentation of continued authenticity throughout the preservation process?

Goal of the Study

Present a model of the technological features of preservation systems drawn from the case studies that support the authenticity of digital records.

Achieve this goal by analyzing how traditional archival concepts of authenticity are implemented using technology.

Questions needing to be answered

What are the technological features of preservation systems that influence the assessment, documentation, and maintenance of the authenticity of digital records as they move across space and through time?

1. *What does 'authentic digital records' mean in this context?*
2. *What technological features support the assessment of the authenticity of records as they are transferred to a preservation system?*
3. *What technological features support the documentation of the authenticity of records ingested into a preservation system?*
4. *What technological features of a preservation system support the maintenance of the authenticity of records over the long term?*

What is a record?

- Diplomatic Analysis Template (InterPARES 3)
  - Must possess stable content and fixed form, and be affixed to a stable medium
  - Must participate in an action and be a natural byproduct of that action
  - Possess an archival bond with other records
  - Creation must involve at least three persons (author, addressee and writer)
  - Must possess an identifiable context

What does authenticity mean in this context?

- "one that can be proven to a) be what it purports to be, b) have been created or sent by the agent purported to have created or sent it, and c) have been created or sent when purported." (ISO15489)
- “[t]he quality of being genuine, not a counterfeit, and free from tampering, and is typically inferred from internal and external evidence, including its physical characteristics, structure, content, and context” (SAA Glossary)
- “To assess the authenticity of an electronic record, the preserver must be able to establish its *identity* and demonstrate its *integrity*.” (Authenticity Task Force, InterPARES, 2001)

Case Study Selection

- Selection Included:
  - Include two archives and one library with a special digital collection,
  - Drawn from three countries on two different continents
  - Include one institution at the national level, one at province level and one metropolitan,
  - Different technological infrastructures -- Include two institutions using an open source solution and one that developed their own custom in-house solution

How do technological features support authenticity?

- Technological Features discovered through
  - Interviews
  - Observation
  - Documentation
  - Code Analysis

- Related the Technological Features to the Benchmark and Baseline requirements developed to InterPARES

FINDINGS

Mapping preservation activities

- Example from Metropolitan Case Study :

“This ability to forward migrate the native file formats into formats that are deemed viable for preservation over a longer-term provides support for *Benchmark Requirement A.4: Protective Procedures Media and Technology*”

- Description of Benchmark Requirement A.4:

“Procedures to counteract media fragility and technological obsolescence include: planning upgrades to the organization’s technology base; ensuring the ability to retrieve, access, and use stored records when components of the electronic system are changed; refreshing the records by regularly moving them from one storage medium to another; and migrating records from an obsolescent technology to a new technology.”

Example: Summary of Technological Features observed in Transfer Function

	Case Study One: Metropolitan			Case Study Two: Province			Case Study Three: National	
Sec.	Technological Feature	Supports	Sec.	Technological Feature	Supports	Sec.	Technological Feature	Supports
4.2	Capture and associate metadata within classification scheme	A.1	5.2.1	Authority defined accessed roles	A.2	6.2.1	Users must have Active Directory Accounts	A.2
4.2	Define security and access privileges for each user	A.2	5.2.2	Comparison of SIP to agreed contents	3.2 PAIS	6.2.1	Active Directory Accounts must have transfer privileges	B.1.b

CREATING THE MODEL

Development of *Technical Features Supporting the Authenticity of Digital Records* Model (TechSAR Model)

- Based on analysis of case study implementations
- Supplement existing preservation models by adding in technical descriptions and examples
- Separated into three functional areas of:
  - Transfer
  - Ingest
  - Maintenance
- Technological features from case studies grouped by functional areas
  - Repeating features across case studies
  - Unique implementations that provide strong support of authenticity
- Expressed visually through UML Activity Diagrams
- Described textually though Use Case Template created for Agile (Cockburn, 2001)

	Case Study One: Metropolitan			Case Study Two: Province			Case Study Three: National	
Sec.	Technological Feature	Supports	Sec.	Technological Feature	Supports	Sec.	Technological Feature	Supports
4.2	Capture and associate metadata within classification scheme	A.1	5.2.1	Authority defined accessed roles	A.2	6.2.1	Users must have Active Directory Accounts	A.2
4.2	Define security and access privileges for each user	A.2	5.2.2	Comparison of SIP to agreed contents	3.2 PAIS	6.2.1	Active Directory Accounts must have transfer privileges	B.1.b

	Case Study One: Metropolitan			Case Study Two: Province			Case Study Three: National	
Sec.	Technological Feature	Supports	Sec.	Technological Feature	Supports	Sec.	Technological Feature	Supports
4.2	Capture and associate metadata within classification scheme	A.1	5.2.1	Authority defined access roles	A.2	6.2.1	Users must have Active Directory Accounts	A.2
4.2	Define security and access privileges for each user	A.2	5.2.2	Comparison of SIP to agreed contents	3.2 PAIS	6.2.1	Active Directory Accounts must have transfer privileges	B.1.b

	Case Study One: Metropolitan			Case Study Two: Province			Case Study Three: National	
Sec.	Technological Feature	Supports	Sec.	Technological Feature	Supports	Sec.	Technological Feature	Supports
4.2	Capture and associate metadata within classification scheme	A.1	5.2.1	Authority defined access roles	A.2	6.2.1	Users must have Active Directory Accounts	A.2
4.2	Define security and access privileges for each user	A.2	5.2.2	Comparison of SIP to agreed contents	3.2 PAIS	6.2.1	Active Directory Accounts must have transfer privileges	B.1.b

	Case Study One: Metropolitan			Case Study Two: Province			Case Study Three: National	
Sec.	Technological Feature	Supports	Sec.	Technological Feature	Supports	Sec.	Technological Feature	Supports
4.2	Capture and associate metadata within classification scheme	A.1	5.2.1	Authority defined access roles	A.2	6.2.1	Users must have Active Directory Accounts	A.2
4.2	Define security and access privileges for each user	A.2	5.2.2	Comparison of SIP to agreed contents	3.2 PAIS	6.2.1	Active Directory Accounts must have transfer privileges	B.1.b

7.6.1 Directory Service of Known Users

Use Case Number	directoryService
Goal in Context	Validate access credentials and permissions within the system.
Scope	Transfer, Maintenance
Preconditions	• Roles have been defined in the Directory Service. • The Entity has been entered into the Directory Service. • The Entity has been assigned to a Role.
Success End Condition	The Entity is validated, and its assigned role/permissions are returned.
Trigger	Node passed entity credentials verification by Archives.
Description	This node is responsible for assigning and enforcing responsibility for individuals and computer resources to execute the create, modify, annotate, relocate, and destroy actions on records. When passed a set of username/password account credentials, this node will validate those credentials against those in the directory store. If the credentials are valid, the node will return the role/permissions that entity has been assigned. Every Entity within the Directory Service is required to have a Name, a unique identifier (username), Title, group assignment (such as Department/Division/Branch of operations), and contact information.
Authenticity Support	Benchmark Requirement A.2: Access Privileges, by assigning responsibility to create, modify, annotate, relocate or destroy records depending on the role by the appropriate authority; and Baseline Requirement B.1.b: Security controls, by verifying that the entity has been assigned the responsibility to execute the attempted action.

Example	The following is an example of a directory entry (created by the author) expressed in JSON for an individual with a Department Records Officer (DRO) role assigned to them:
	
	{
	"id": 8,
	"email": "john.doe@example.com",
	"firstName": "John",
	"middleName": null,
	"lastName": "Doe",
	"phoneNumber": "4145551234",
	"passwordHash":
	"06e0e6637d27b2622ab52022db713ce2",
	"role": {
	"name": "DRO"
	},
	"organization": {
	"id": 4,
	"name": "Accounting, Department of"
	},
	"status": {
	"name": "Enabled"
	}
	}

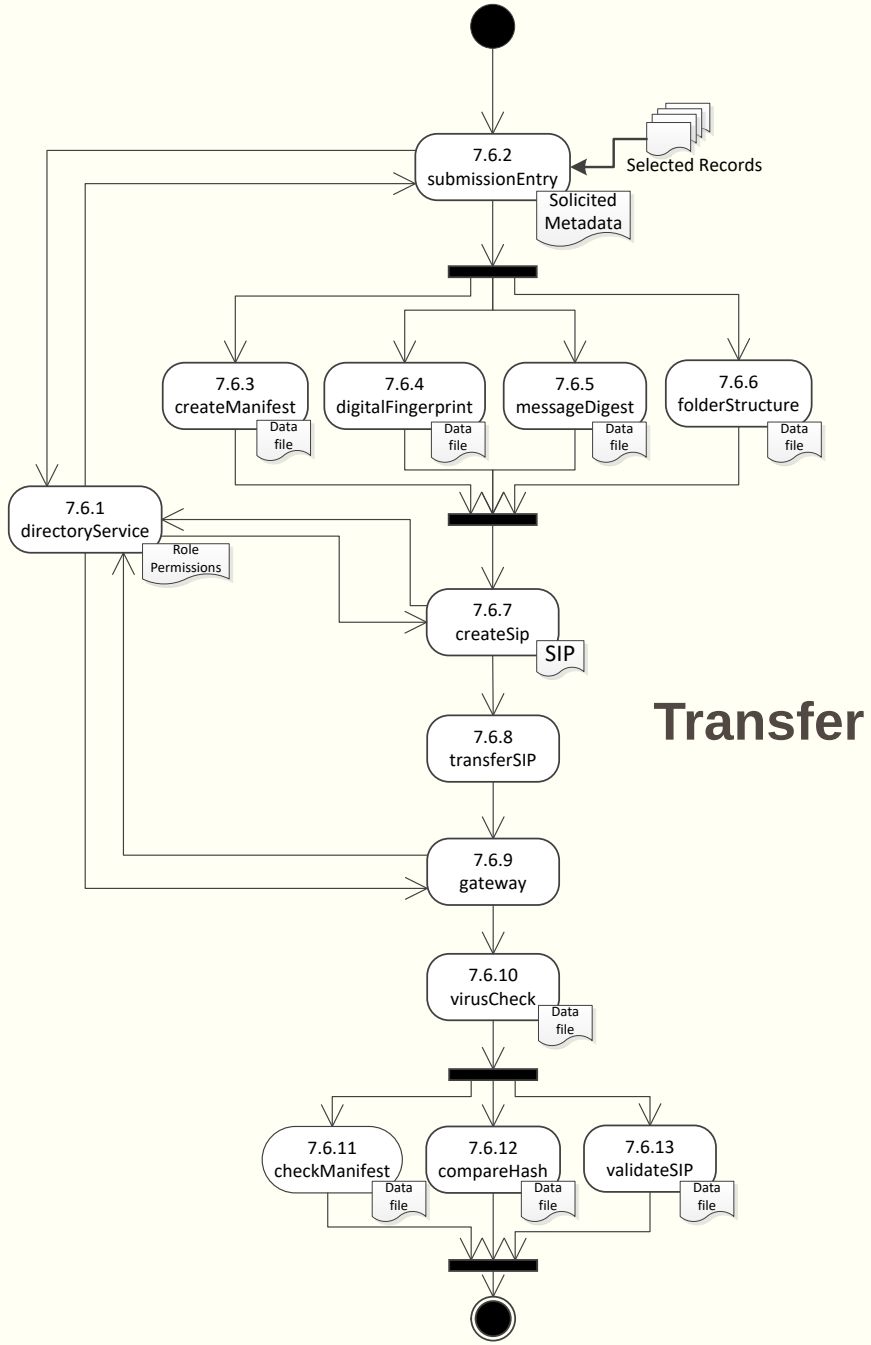
Activity Diagrams

4.2
Define security and access privileges for each user

5.2.1
Authority defined access roles

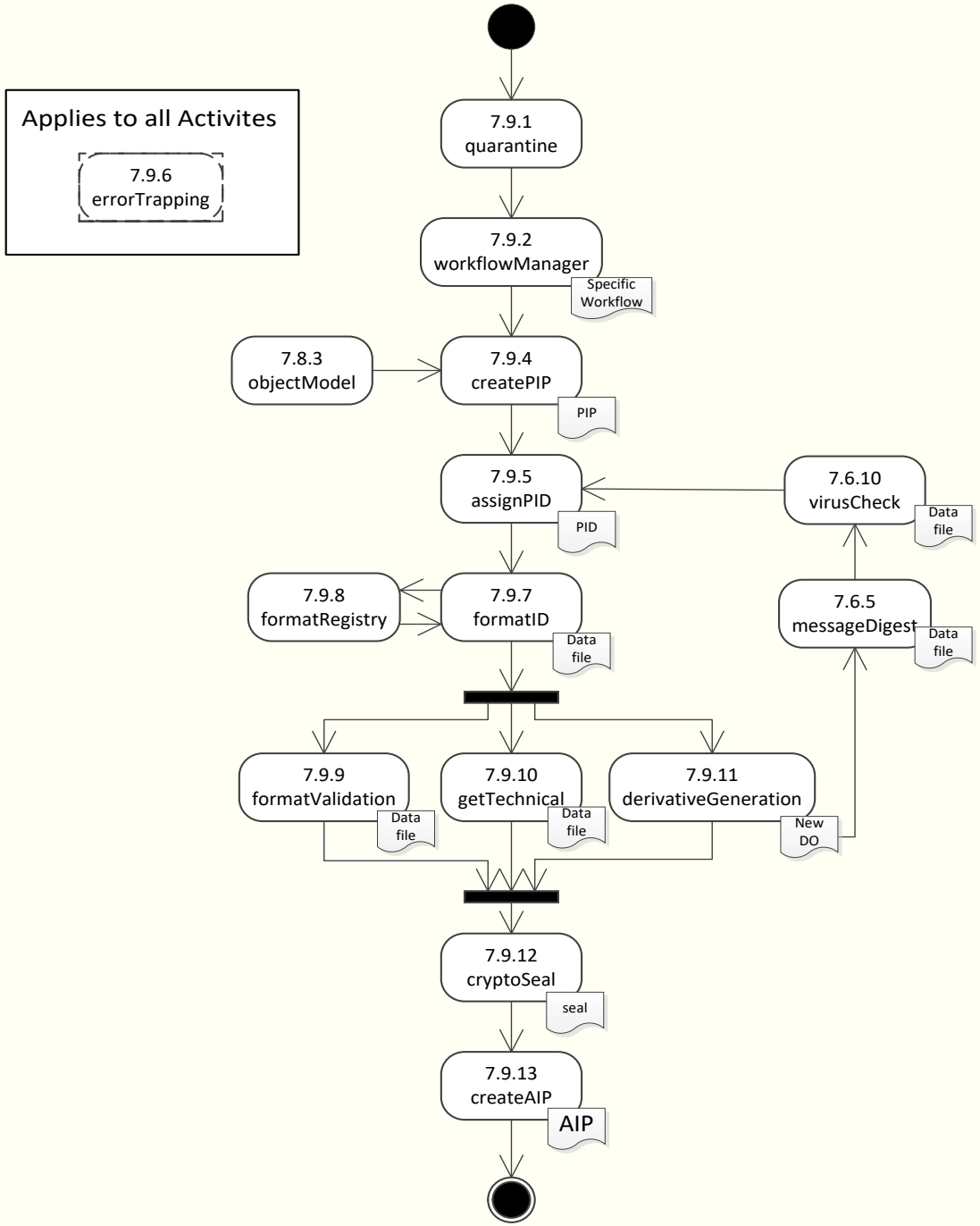
5.2.2
Validate access credentials prior to transfer

6.2.1
Users must have Active Directory Accounts



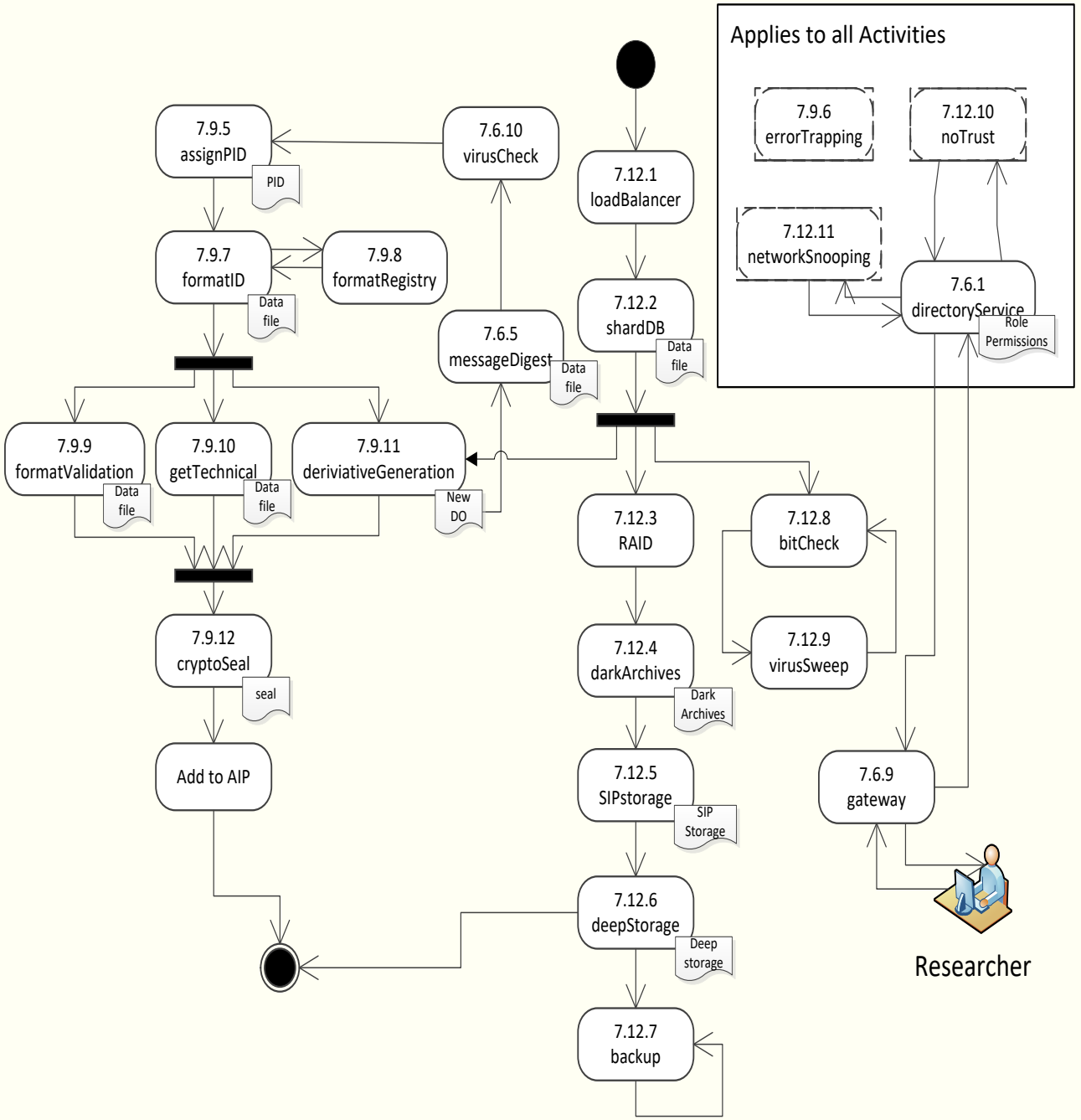
Transfer

Activity Diagrams



Ingest

Activity Diagrams



Maintenance

IMPLEMENTING THE MODEL

Hawaii State Digital Archives

- Initial sample of over 15million digital records, 350TB
- Strong focus on capturing or producing evidence of the authenticity of the preserved records
- Built upon an existing open source preservation application
  - Use the TechSAR model to develop the functional requirements
  - Kept the workflow manager
  - Writing our own microservices
- Started with using the Diplomatic Analysis Template as foundation
  - First must prove that it IS a record
  - Capture metadata the provides evidence of the

Implementation

- As each TechSAR activity is designed as a microservice
  - Refer to the Baseline or Benchmark requirement that is supported by the activity
  - Decompose its operation into documentable metadata
    - Capture, create or obtain necessary evidence of how the activity supports the authenticity of that record
  - Write that metadata into the Archival Information package that is being preserved

- At completion of development, circle back to confirm coverage of TechSAR model
  - Look for gaps in existing model, areas where model can be expanded

Examples

Activity: Copy SIP to second storage array

Supports: *Benchmark Requirement A.7: Identification of Authoritative Record,* by maintaining a copy of the SIP as it was transferred by the submitter. Should any questions arise as to the legitimacy of the processed record stored in the preservation system, or the process through which that record was ingested and maintained, the original SIP may be referenced for comparison.

Evidence

```
<PREMIS:event>
    <PREMIS:eventIdentifier>
        <PREMIS:eventIdentifierType>UUID</PREMIS:eventIdentifierType>
        <PREMIS:eventIdentifierValue>9111e6b5-ff23-331d-81a1-9355656b7ab5
        </PREMIS:eventIdentifierValue>
    </PREMIS:eventIdentifier>
    <PREMIS:eventType>Storage Migration</PREMIS:eventType>
    <PREMIS:eventDateTime>2017-10-25T20:37:54Z</PREMIS:eventDateTime>
    <PREMIS:contentLocation>
        <PREMIS:contentLocationType>Tape Library</PREMIS:contentLocationType>
        <PREMIS:contentLocationValue>G12</PREMIS:contentLocationValue>
    </PREMIS:contentLocation>
</PREMIS:event>
```

Example

Activity: Produce cryptographic time-stamp seal a record upon ingestion

Supports: *Benchmark Requirement A.6: Authentication of Record* by establishing and implementing a mechanism to independently determine the genuineness of the record provided to the researcher

Evidence:

<Time-stampRequest>
<MessageImprint>
<messageDigestAlgorithm> SHA-256 </messageDigestAlgorithm>
<messageDigest>rDMpt/JGmkrvGYziezSNO0V8vPMBRcqtQuYr20wjATc=</messageDigest>
<OID>2.16.840.1.101.3.4.2.1</OID>
</MessageImprint>
<Nonce>1V5dRoF17J5LsxHkdrxX1k7G8SmzG1iqP2qO9bl95pbm2wJAu9Hspw==</Nonce>
</Time-stampRequest>

<Time-stampResponse>
<Time-stampToken>
<Bytes>ZIIKMwYJKoZIhvcNAQcCoIIKJDCCCiACAQMxDzANBglghkgBZQMEAgEFADCCATcGCyqGSIb3DQEJEAEEoIIBJgSCASIwggEeAgEBBgorB
…7139EBn+aoLqct3Nff69Bas1NK8S08FAMkqLj/0YV3phx9rxgukG5zlQ6qVMwoCNzZlJmISX9bsWmcjXgmgGao=</Bytes>
<GeneralizedTime>20181016122152.424Z</GeneralizedTime>
</Time-stampToken>
</Time-stampResponse>
</Time-stamp>

Example

Activity: Extract technical and descriptive metadata

Supports: *Benchmark Requirements A.1.a: Identity of the record,* by expressing explicitly those attributes of the record that are embedded in the digital object, such as chronological date, name of the author, and title

Example

<tool name="Exiftool" version="9.13">
<rawOutput>
ExifToolVersion 9.13
FileName Wildlife.wmv
FileSize 25 MB
FileModifyDate 201:05:07 10:29:25-10:00
FileAccessDate 2013:05:07 10:29:41-10:00
FileInodeChangeDate 2013:05:07 10:29:41-10:00
<VideoPixelAspectRatio>1</VideoPixelAspectRatio>
<VideoAlphaMode>None</VideoAlphaMode>
<AudioSampleRate>44,100</AudioSampleRate>
<AudioSampleType>16-bit integer</AudioSampleType>
<AudioChannelType>Stereo</AudioChannelType>
<VideoFrameRate>29.970030</VideoFrameRate>
<Description>Footage: Small World Productions, Inc; Tourism New Zealand | Producer: Gary F. Spradling | Music: Steve Ball</Description>
<Rights>© 2008 Microsoft Corporation</Rights>
<Title>Wildlife in HD</Title>

Detailed Use Cases

Preconditions: Digital objects have passed QUALITY_ASSURANCE; secondary storage location has sufficient storage availability for duplicate SIP copy; network connectivity; Multi-malware scanning engines configured.

Trigger: QUALITY_ASSURANCE has confirmed all verifiable formats are valid for the SIP

Successful Outcome: All digital objects successfully scanned for malware, infected files cleaned, un-cleanable files moved into error folder

Steps

Records Receiver confirms that the virus definition is up to date. If the malware definition is not up to date, then Ex. 1; else, continue.

Malware scanning tool scans all digital objects

If the malware scanner detects dangerous items, then malware scanner attempts to clean infected digital objects. If files cannot be cleaned, then Ex. 2; else, MALWARE_CHECK end; continue VALIDATION (INGEST2.0).

Records Receiver moves a copy of the entire SIP to a secondary storage location for disaster recovery.

Detailed Use Case (Continued)

Exceptions

Ex. 1 – If the virus definition is not on the current version, then update virus definition. If the virus definition cannot update, then suspend the process and notify *System Administrator*; else, continue MALWARE_CHECK.

Ex. 2 – If the virus scanning tool cannot clean infected files, then the infected files are deleted from the system, and the *Records Creator* is notified of infection and destruction of the zombie digital objects.

Event Details

Name and version of service invoking MALWARE_CHECK

Date of Scan

Signature version of all engines performing scans

PID of digital objects undergoing the scan

Results of scan

QUESTIONS?

Adam Jansen, PhD
State Archivist
Hawai‘i State Archives
Adam.Jansen@Hawaii.Gov

https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0384577