ISO/IEC JTC 1 N 9034

2008-04-22

ISO/IEC JTC 1
Information Technology

Document Type:	**Other Document(Defined)**
Document Title:	**Convenor's report fromt he recent ISO/IEC JTC 1 SWG-ARM meeting in London England 10-11 March**
Document Source:	**SWG-ARM Convenor**
Reference:	
Document Status:	**This document is circulated to JTC 1 National Bodies for information**
Action ID:	**Information**
Due Date:	
No. of Pages:	**24**

Secretariat, ISO/IEC JTC 1, American National Standards Institute, 25 West 43rd Street, New York, NY 10036; Telephone: 1 212 642 4932; Facsimile: 1 212 840 2298; Email: lrajchel@ansi.org

JTC1 SWG-ARM meeting
10-11 March 2008, BSI, London

14 attendees from

5 countries,

6 JTC1 and ISO committees,

and the AES.

Presentations to SWG-ARM

- **Mr Richard Clark – Creating the JPEG Historical Archive**
- **Mr Alan Shipman - Archiving electronic documents: the PDF/A approach**
- **Mr Mark Yonge - Requirements for archiving of standards development information, texts and associated material – the AES experience**
- **Mr Laurent Duplouy - Experiences of archiving and retention of electronic records in the French National Library**
- **Jean Pierre Evain - Use of structured metadata for television production and archiving**
- **Tony Hammond Health Informatics: Security Requirements for Archiving of Electronic Health Records**
- **Hendry – ISO/IEC JTC1/SC29 WG11 on Professional Archival Application Format**
- **Prof Bryan Manning – Biometrics and Medical Informatics Archiving**
- *Additionally ISO had provided a presentation on Livelink but noone from ISO or Opentext attended the meeting*

Challenges of content archiving

- Access requirement
  - long-term content availability
  - long-term content readability
- Over time, archived records may be re-classified, reflecting both changes in content and/or regulatory requirements. These changes may require that content be moved
  - to a different storage provider, and/or
  - a different storage medium.
- Solution must provide
  - robust metadata
  - engine, records management capabilities and world-class, full-text search and
  - discovery engine with the robust, scalable, and secure virtualized storage,

Features of Content Archiving for Livelink Enterprise Server

- Ensure the authenticity of archived content
- Reduce storage space requirements
- Retain detailed audit trails of activity within the archive
- Ensure content accessibility
- Define archiving rules to specify the storage provider to which certain content should be archived
- Caching capabilities ensure the fastest content retrieval response time
- Write individual files or container files to storage subsystem
- *Note: the ISO implementation of Livelink does not include specific archiving facilities*

Issues identified when creating JPEG archive - 1

- Acquiring/locating copies of earlier documents
- Cost of digitising older documents, essential for long term archiving given that originals are scattered across the world
- Variation in NB/committee policies
  - a) material retained
  - b) formats used,
  - c) access to information
- Availability of archived documentation (not covered explicitly in Directives)

Issues identified when creating JPEG archive - 2

- Versioning of documents is still a problem, come committees (eg JPEG) keep a copy of all versions but others (eg MPEG) replace copies on the register with the latest version and do not retain copies of intermediate versions
- More ad hoc material is being used in development of standards eg email, wiki, VOIP teleconferences etc; these may have different access constraints leading to archiving issues
- Increasing focus on legal issues
  - copyright,
  - patented technology in specifications
  - differences in treatment in different jurisdictions (eg software patents)
- Importance of management of document creation
  - a) metadata
  - b) formats used
  - c) defining access to information (currently and in the future)
  - d) exposure to search engines etc
- Archiving of metadata, search tools, rendering engines etc used

Issues identified in document management

- Information acquired during standards development is a key asset for JTC1
- Loss of information
  - Reduces effectiveness
  - Destroys historical archive
  - Reduces public accountability
  - Decreases confidence in process by standards developers/users

Issues identified in document management

- ISO 19005 PDF/A
  - Standardised and documented
  - ISO 32000 – Reference manual
  - Multiple parts for new facilities
  - Creation tools now available
  - Becoming widely implemented
  - Format checker
- Is this an answer for JTC1?

Issues identified by AES

- Preservation of material: what format?
- Need regular transfer of archive to current media
- Interchange of material between organisations
- Need policy for appropriate document deletion
- Some dynamic material cannot exist in paper form (eg databases, dictionaries etc)
- Rules for authoring can help
- Metadata needed to describe archived material
- Policies for storing emails, websites etc are needed and *must* be followed by all

AES recommendations:

- Plan to be “readable” in 20 years
- Avoid unnecessary complexity
- Use standardised data file types
- Migrate to stable new technologies
- Archive older formats
- Keep paper copies

Information management in production

The content model is very similar across domains and applications

However, the richer the metadata set the more difficult to share

Need to focus on business objects, workflows, processes, use cases and best practices to help...

...towards a common data model, service oriented architectures and the development of reusable web services

Standards challenge: shift to open source?

Information management in production

- **Asset management vital**
- **Service oriented architecture**
  - Workflow independence
  - Identify common interfaces
  - Reusable, adaptable web services
  - Open source rather than standardisation
- **Metadata is key**
- **Need new formats**
- **Connecting & Managing media workflows and digital assets is important in both standardisation and broadcasting**

French National Library

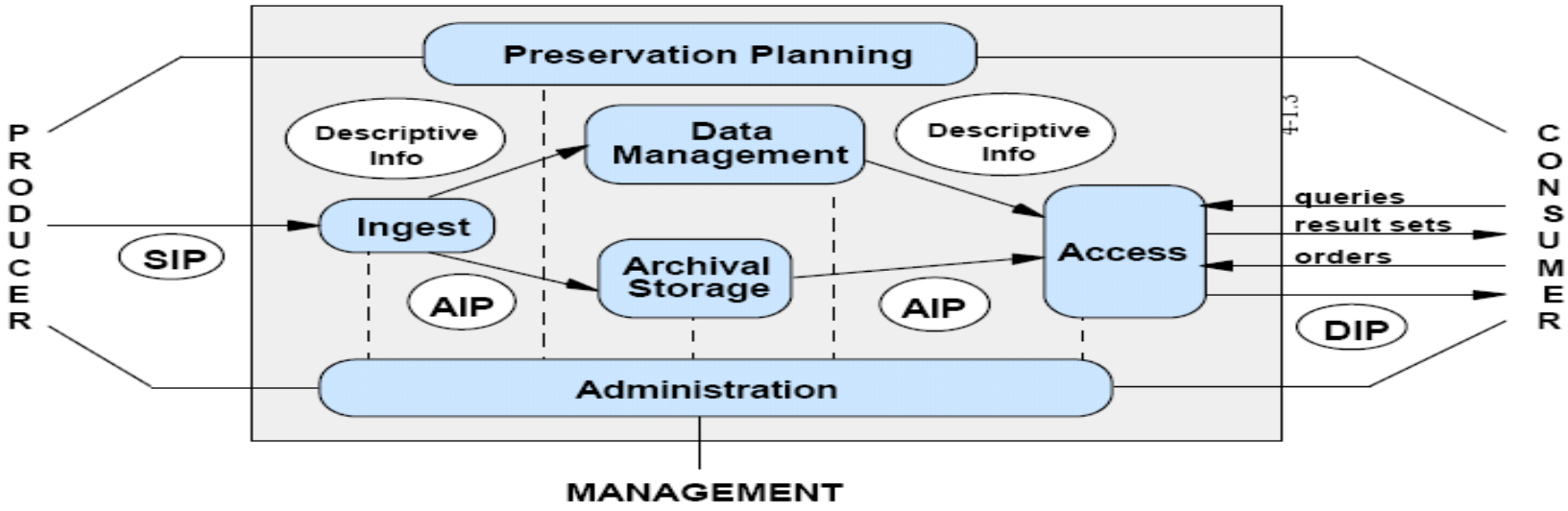
- Issues
  - input in native format
  - Disappearance of analogue material with mass digitisation
  - Obsolescence of hardware may impede restitution
- Variety of material
  - Recording media/formats
  - File formats
  - Ways of processing
  - Lifetime of material

French National Library

- Information on producer and consumer aspects
- Scale of archive: by 2010 over 1000Tb
- Importance of policies
  - Ingest policy
  - Archival policy
  - Access policy
- Risk assessment important

French National Library - OAIS

- Open Archival Information System (OAIS)
- **ISO 14721:2003 currently being revised**
- Functional entities can be modular allowing maintenance of parts of system



Content Information Type	Reference	Provenance	Context	Fixity
Space Science Data	• Object identifier • Journal reference • Mission, instrument, title, attribute set	• Instrument description • Processing history • Sensor description • Instrument • Instrument mode • Decommutation map • Software interface specification	• Calibration history • Related data sets • Mission • Funding history	• CRC • Checksum • Reed-Solomon coding

Digital Library Collections	• Bibliographic description
	• Persistent identifier
		• For scanned collections:
		  • metadata about the digitisation process
		  • pointer to master version
		• For born-digital publications:
		  • pointer to the digital original
		• Metadata about the preservation process:
		  • pointers to earlier versions of the collection item
		  • change history
			• Pointers to related documents in original environment at the time of publication	• Digital signature
				• Checksum
				• Authenticity indicator

Software Package	• Name • Author/Originator • Version number • Serial number	• Revision history • License holder • Registration • Copyright	• Help file • User guide • Related software • Language	• Certificate • Checksum • Encryption • CRC

French National Library - OAIS

- BNF Goals:
  - To handle the responsibility of the archiving of data
  - To provide an archival storage service
  - To reach a minimal critical mass to reduce cost (hardware, software and humans means)
  - To allow the mutualisation of archival storage between several institutions
- To do this
  - Need investment: to cover
    - Cost of the acquisition or replacement of the information, and maintenance
    - Cost of the hardware
  - Preserve for a period long enough so that must handle
    - Technological changes (new medias, new formats)
    - Evolution of the community of users on the information maintained in the archive.
  - Period that can last forever.

PA-AF Professional Archive

- Multimedia Application Format (MAF)
  - Build up by combination of several MPEG's standards + additional non-MPEG's standards
  - MAF specification describes how the combination of several MPEG's standards can work together to provide solution to specific application domain
- Benefits:
  - Data Aggregations
    - Capable of storing all related items in one archive container
    - Allow access to individual item in the archive container
  - Context description by metadata
    - Better understanding of archived items
    - Better searching facility
  - Access management
    - Rights description for each archived item
    - Who can access? What kind of action? Under what conditions?

PA-AF Professional Archive

- Goals
  - Richer packaging with context information
  - Free-choice of compression algorithm
  - Free-choice of protection algorithm
  - Cross platform operation including multiple bytes character set support

Electronic Health Records

- **Needs to address linking electronic health record data from disparate systems/ organisations**
- **Implementing current proposed standard into the NHS systems would mean a full redesign of the whole architecture**
- **2 Technical Specifications under development,**
  - **Archiving**
  - **Guidelines for Long Term Preservation of Electronic Health Records**

Biometrics

- Issues of identity
- Tying health records to, and identifying the subject
- Keeping information (may be used for long term genealogy studies as well as short term treatment)
- Allowing access to information (or relevant) part of information

Recommendations:

- **Recommendation 1:**

  To request information from ITTF on their strategy for providing archiving facilities for the Livelink services provided to committees
- **Recommendation 2:**

  To circulate the presentations for background information together with overview report to ISO/IEC JTC1
- **Recommendation 3:**

  To develop a requirements document based on the presentations and TC/SC inputs to date for discussion at the next meeting and an overview of the current archival-related standards identified in presentations etc
- **Recommendation 4:**

  To inform ITSIG of the discussions in ISO/IEC JTC 1 SWG-ARM and make them aware of the issues identified

Convenor’s Remarks

- The first day had a number of informative presentations covering a wide area, many aspects of which are in JTC1’s domain
- While there was not a large attendance there have been clear indications since the meeting that some of the issues raised may lead to additional work in particular JTC1 committees
- JTC1 has revised the section on archiving in its directives and should provide clear guidance to its committees on how to meet those requirements and provide long-term storage of relevant content acquired during the development of its standards.
- JTC1 needs to develop a clear retention policy to help convenors and secretariats
- JTC1 needs to consider how older content (eg documents) can be made accessible
- JTC1 needs to ask ISO if its current implementation of Livelink can accommodate the volumes and variety of data accumulated in certain committees eg SC37, SC24 and SC29 where there are considerable numbers of test sequences and evaluation results which need to be retained
- JTC1 NBs need to consider whether they also need to develop a retention policy for material accumulated in their JTC1 mirror committees.