

Access IT Training

**Should I know something more
about digital preservation?**

In Star Trek we trust...

- Star Trek : "All Our Yesterdays"
- **Stuart Lee: "Content is king"**
 - <http://digitisation.jiscinvolve.org/2009/07/02/jdcco9-stewart-lee-content-is-king/>

„The problem of digital preservation is one of the most challenging research problems faced by the community of digital libraries today receiving growing interest by researchers and practitioners alike“

„Some Preliminary Ideas Towards a Theory of Digital Presevation“

Giorgos Flouris, Carlo Meghini
CASPAR project

The Past is Prologue

- First digital preservation programs were developed in 1960s
- Each generation of technology brings changes in potential capabilities to
 - creation
 - preservation of digital content
- Each new technology goes through a similar path
 - From **idea** to **implementation** than to **mainstream use** and in most cases to **obsolescence**

The Past is Prologue

- Digital Preservation encompasses a broad range of activities designed to:
 - extend the usable life of computer files
 - protecting files from media failure
 - physical loss, and obsolescence.
- Information must be intact and readable whenever user needs it

The Past is Prologue

- Mentioned accessibility can be divided to:
 - Content renderability
 - Content can be viewed by humans or processed by computers
 - Understandability
 - Content can be interpreted by humans
- This implies main issues:
 - Bitstream preservation
 - Preservation of content, form, style and functionality

The Past is Prologue

- What may happen with the information over time?
- Interpretation requires additional knowledge – a context
- Take a look at this what is going with Flickr:The Commons
 - Can we do something to avoid such a situation in the future?
- There is also an issue of authenticity of information

The Past is Prologue



Source:

“Digital History: A guide to gathering, preserving, and presenting the past on the web.”

by Daniel J. Cohen and Roy Rosenzweig

<http://chnm.gmu.edu/digitalhistory/>

The Past is Prologue



Digital Preservation

- There is no universal solution which could be used for all data types and situations
- There are many different content preservation elements
- The most important includes:
 - Bitstream refreshing
 - Technology preservation
 - Analog backups
 - Migration
 - Reliance on Standards
 - Replication
 - Emulation

Bitstream refreshing/copying

- Bitstream refreshing and copying is more commonly known as “backing up your data”
- Protects data from decay, media failure, malicious destruction etc.
- It should be considered as a minimum maintenance strategy

Technology preservation

- Technology museum
- The idea is to preserve the technical environment that runs the system
 - Including media drives, original applications, OS
- It offers the potential of coping with media obsolescence
- Assuming the media hasn't decayed beyond readability

Technology preservation

- Technology preservation is ultimately a dead end, since no obsolete technology can be kept functional indefinitely
-it is also very expensive

Analog backups (sic!)

- An analog copy of a digital object can, in some respects, preserve its content and protect it from obsolescence
- You are losing all digital qualities, including sharability and lossless transferability
- Text and monochromatic still images are the most amenable to this kind of transfer

Reliance on Standards

- Information about format of a file are crucial for renderability
- What is in a file format specification?
 - e.g. the role of each byte in file header
- Specification is bare minimum, we will also need some software
- Without file format specification the only chance lies in digital archeology

Reliance on Standards

- Why file formats are superseded?
 - Higher user expectations
 - Software upgrades fail to support legacy files
 - The format fails, stagnates, or is no longer compatible with the current environment
- Are Some File Formats Less Vulnerable to Obsolescence than Others?

Reliance on Standards

- Open specification is a minimal requirement
- Wide adoption
- History of backward compatibility
- Good metadata support
- Good range of functionality, but not overly complex
- Available interchange format with usable target
- Built-in error checking
- Reasonable upgrade cycle

Reliance on Standards

- When choosing proprietary format ensure that migration path exist
- PRONOM database
 - <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>
 - British National Archive database of file formats
- “Risk Management of Digital Information: A File Format Investigation”, Lawrence et al.
 - <http://www.clir.org/pubs/abstract/pub93abst.html>

Reliance on Standards

- INFORM Methodology is an approach for measuring the preservation durability of digital formats
 - <http://www.dlib.org/dlib/november04/stanescu/11stanescu.html>
 - Whether or not royalties or license fees are or may be requested
 - Whether the source or specification can be independently inspected
 - Whether revisions have maintained support for backward compatibility
 - Whether it is complex or poorly documented
 - Whether it is widely accepted or simply a niche format

Migration

- The goal of migration is to copy data, or convert data, from one technology to another **preserving the essential characteristics of the data**
- It is not always possible to make an exact digital copy or replica of an object
- Migration can deal with obsolescence of the physical storage medium, encodings and formats

Replication

- Intention is to preserve documents through copying and the use of multiple storage locations
- Bitstream copying is a form of replication
- Some consider replication as a form of migration
- LOCKSS (Lots of Copies Keeps Stuff Safe)
 - Peer-to-peer data trading, open, free-market form of replication

Emulation

- Combines software and hardware to reproduce character of another computer, allowing old programs or media to operate in newer environment
- Emulation requires special software (emulators)
 - Virtualization might be also useful here

Emulation

- Win 7 is a good example of how useful emulation can be
- KEEP – Keeping Emulation Environments
 - <http://www.keep-project.eu>
 - Will develop an Emulation Access Platform to enable accurate rendering of both static and dynamic digital objects
 - Including text, sound, image files; multimedia documents, websites, databases, videogames etc.

- Team Digital Preservation and the Aeroplane Disaster
 - <http://www.youtube.com/watch?v=EKnsZZzuUr4>



Sustainability

- Mentioned concepts and actions are only one aspect of the problem
- Another (maybe even more important) issue is sustainability of institution, repository and funding

Sustainability

- There is a number of ongoing projects which deal with long term digital preservation
- There are also some well-established standards which assures that our repository will be useful in the long perspective
 - Open Archival Information System (OAIS)
 - Trusted Digital Repositories : Attributes and Responsibilities (TDR)

OAIS

- Reference Model for an Open Archival Information System
 - <http://public.ccsds.org/publications/archive/650xob1.pdf>
- Prepared by international group working by the NASA Consultative Committee for Space Data Systems
- OAIS is an ISO standard (ISO 14721:2003)

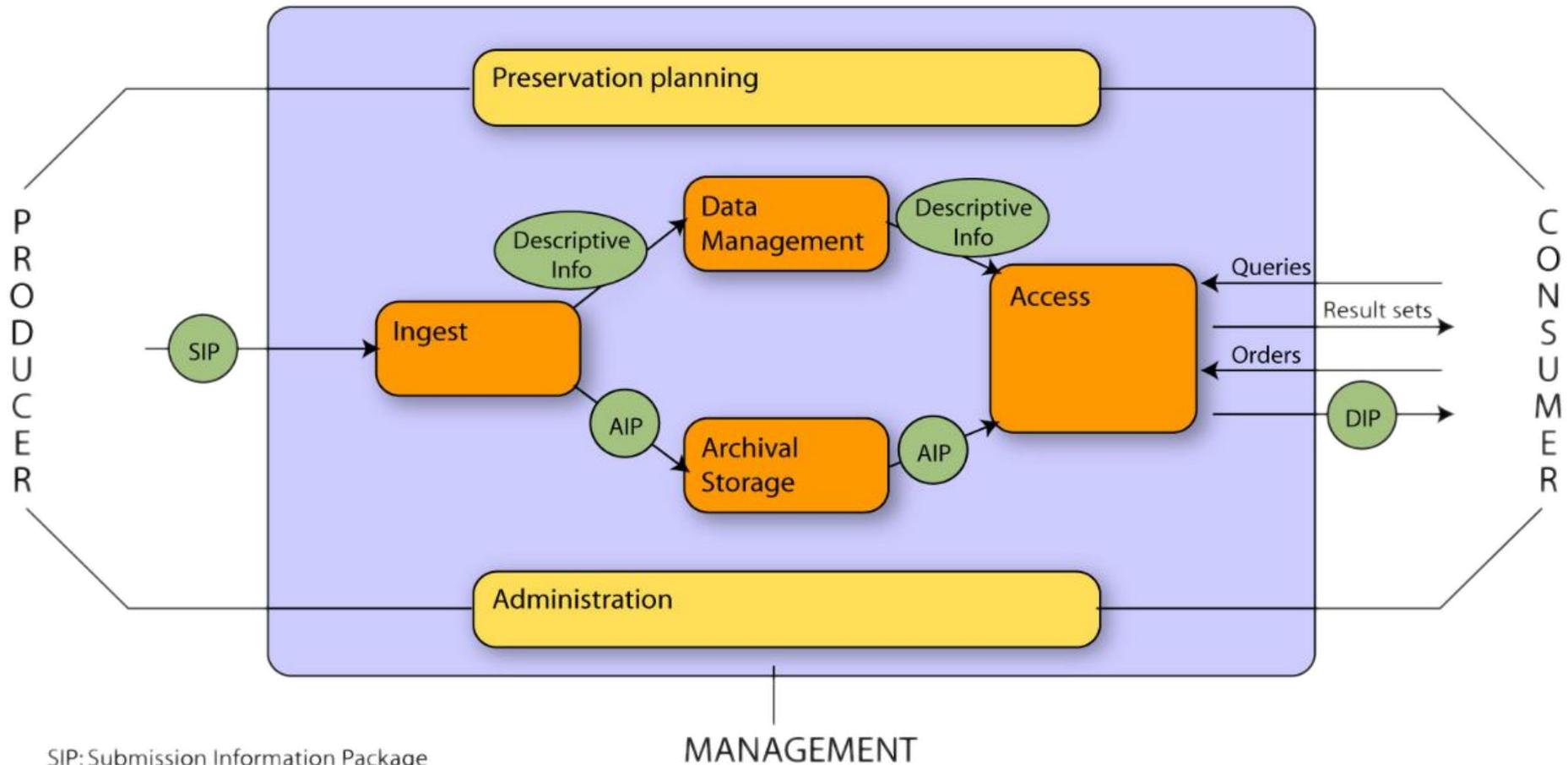
References

- **„Digital Preservation Management Workshops and Tutorial“** - online materials
 - <http://www.icpsr.umich.edu/dpm/>
 - Available also in French and Italian

OAIS

- It provides functional framework for sustaining digital objects in managed repository
- OAIS is a foundation of many digital preservation initiatives
- OAIS defines what is needed but not how to build repository

OAIS



SIP: Submission Information Package
AIP: Archival Information Package
DIP: Dissemination Information Package

OAIS

- National Library of the Netherlands was the first library in the world which has an operational digital deposit system based on the OAIS model
- Some European project are aimed to create tools compliant or extending OAIS model
 - e.g. CASPAR, PLANETS

TDR

- TDR was produced by the Research Libraries Group (RLG) and OCLC
- TDR defines the organizational context for a digital preservation program
- TDR embraces OAIS
- TDR demonstrates what OAIS will mean for an institution
- Enumerates characteristics of a sustainable digital repository

TDR

- TDR addresses the seven attributes such repositories must have:
 - OAIS compliance
 - Administrative responsibility
 - Organizational viability
 - Financial sustainability
 - Technological and procedural suitability
 - System security
 - Procedural accountability

TDR

- Organizational requirements:
 - provide evidence of fundamental commitment to implementing community-agreed standards, best practices
 - commit to understanding and implementing the OAIIS standard
 - meet national/international standards on environment
 - meet or exceed community standards and share measurements with depositors
 - involve external community experts in regularly validating/certifying processes and procedures
 - commit to transparency and accountability in all actions

TDR

- Financial Sustainability:
 - establish and maintain good business practices and an auditable business plan
 - demonstrate financial fitness and ongoing financial commitment
 - balance risk, benefit, investment, expenditure
 - maintain adequate budget and reserves and actively seek potential funding sources

TDR

- Financial Sustainability has until recently been the most overlooked attribute
- Project based funding, especially through grants and gifts, is a common starting point for digital preservation
- This funding approach works for start up costs, but a digital preservation program requires ongoing funding

TDR

- Technological and Procedural Suitability
 - consider and adopt appropriate preservation strategies
 - ensure appropriate infrastructure (hardware, software, facilities) for acquisition, storage, access
 - establish technology management policy for repository (replacement, enhancement, funding)
 - comply with relevant standards and best practices (supported by adequate expertise)
 - undergo regular external audits on system components and performance

TDR

- System Security
 - ensure security of systems for digital assets
 - establish policies and procedures to meet requirements
 - copying, authentication, firewalls, backups, disaster preparedness, response, recovery, training
 - stress processes that will detect, avoid, and repair loss
 - It is critical to document changes and resulting actions

TDR

- Procedural Accountability
 - enact all relevant policies and procedures for specified tasks and functions, document all practices
 - establish monitoring mechanisms to ensure continued operation of systems and procedures
 - record and justify preservation strategies
 - set up feedback mechanisms to support problem resolution and negotiate evolving requirements between providers and consumers

Digital Preservation Repository Certification

- In February 2007 Digital Preservation Repository Certification Task Force, published the **Trustworthy Repositories Audit & Certification: Criteria and Checklist**
- Further developments in certification activities include the work of the **Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)**

Digital Preservation Repository Certification

- DRAMBORA provides an online interactive toolkit with an evaluation methodology for repository administrators
 - <http://www.repositoryaudit.eu/>
- Toolkit allows administrators to
 - identify risks at every stage of their activities
 - assess the probability of occurrence
 - determine an action plan should a situation arise

Digital Preservation Repository Certification

- DRAMBORA offers also an live, interactive training
 - Next event will be held in March in Rome

PLANETS

- Planets, Preservation and Long-term Access through Networked Services
 - <http://www.planets-project.eu>
 - Focuses on the needs of libraries and archives
 - Increase Europe's ability to ensure long-term access to its cultural and scientific heritage
 - Improve decision-making about long term preservation
 - Ensure long-term access to valued digital content

PLANETS

- Control the costs of preservation actions through increased automation, scalable infrastructure
- Build practical solutions
- Integrate existing expertise, designs and tools
- Deliver tools and services that can be used in an operational environment

PLANETS

- OAIS breaks preservation planning into 4 functions:
 - Monitor designated community
 - Monitor technology
 - Develop preservation strategy and standards
 - Develop packaging designs and migration plans
- Includes only high-level descriptions; no details for practical implementation
- Important preservation functions seem to be implicit or missing in OAIS

PLANETS

- Broadly map to the OAIS preservation planning functions but also provide added functionality and practical implementation
- Planets experience may lead to refinement / extension of the OAIS reference model

CASPAR

- Cultural, **A**rtistic and **S**cientific knowledge for **P**reservation, **A**ccess and **R**etrieval
 - <http://www.casparpreserves.eu/>
- CASPAR intends to:
 - Implement, extend, and validate the OAIS reference model
 - Enhance the techniques for capturing Representation Information and other preservation related information for content objects

CASPAR

- Design virtualization services supporting long term digital resource preservation
- Integrate digital rights management, authentication, and accreditation as standard features of CASPAR
- Raise awareness about the critical importance of digital preservation among the relevant user-communities

Digital Preservation Europe

- Digital Preservation Europe (DPE)
 - <http://www.digitalpreservationeurope.eu/>
- DPE is an descendant of ERPANET
- Raise the profile of digital preservation
- Promote the ability of Member States acting together to add value to digital preservation activities across Europe
- Use cross-sectoral cooperation to avoid redundancy and duplication of effort

Digital Preservation Europe

- Ensure auditable and certificated standards for digital preservation processes are selected and introduced
- Facilitate skills development through training packages
- Enable relevant research coordination and exchange
- Develop and promote a research agenda roadmap

Digital Preservation Europe

- Help both citizens and specialist professionals recognize the central role that digital preservation plays in their lives and work
- DPE is responsible for development of DRAMBORA

Digital Preservation Coalition

- <http://www.dpconline.org/>
- It's a not-for-profit membership organization established in 2001
- Digital Preservation Coalition is going to support its members to ensure long term access to and management of their digital assets

References

- **„Digital Preservation Management Workshops and Tutorial“** - online materials
 - <http://www.icpsr.umich.edu/dpm/>
 - Available also in French and Italian
- National Library of Australia - **„Policy on preservation copying of collection materials“**
 - <http://www.nla.gov.au/policy/micro.html>

Conclusions

- Open and well-established standards are crucial for LDP
- Long term digital preservation is a really hot topic
 - It can be observed in the number of projects which are dealing with different aspects LDP
- What can I do for my repository?
 - Check out DRAMBORA toolkit!

Q&A
