planets

Planning the Future with Planets
Preservation Planning
Gathering requirements
Discussion session

Presented by **Hans Hofman**
Nationaal Archief Netherlands
13 April 2008

Goal of this session

❑ To discuss what (types of) requirements will influence the preservation planning process?

❑ Requirements on various levels:

   ▪ high-level requirements that are identified in policies (cf. examples)

   ▪ requirements from user communities

   ▪ requirements on a more technical level (cf. tomorrow's programme)

   ▪ requirements for (types of) objects

Preservation policy: situation now

- Trying to understand what organisations do in this area:
  - Large institutions are accumulating expertise and are building trusted digital repositories
  - Small institutions generally lack expertise and funding to build a digital repository
  - Large institutions have formulated various *requirements* – as can be discovered in different types of documents
- No coherent picture (yet)
- Very high level and abstract

Preservation policy: examples

❑ Example policy statements of institutions with a digital preservation programme

- UK Data Archive
- State Library of Victoria
- National Archives of Australia
- ISO/TR 18492:2005

UK Data Archive

- ❑ < UK Data Archive Preservation Policy
- ❑ http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0308.pdf
- ❑ p. 11: "The UKDA has chosen to implement a preservation strategy based upon open and available file formats, data migration and media refreshment."

- ❑ What does this choice mean in practice? Two examples:
  - ▪ Emulation is –apparently– not a preservation strategy that will be chosen; all obsolete files will be migrated.
  - ▪ Migration to open file formats will be preferred.

State Library of Victoria (Australia)

❑ < Digital Preservation Policy, State Library of Victoria
❑ http://www.slv.vic.gov.au/about/information/policies/digitalpreservation.html
❑ **"Storage.** Born-digital objects published on disk (CD-R or DVD) are considered the archival copy and will be stored appropriately. When needed and authority granted, the physical format data may be copied to another storage carrier in order to preserve its contents. The master TIFF files shall be stored appropriately in a secure location on the Library's LAN, and back-ups made in accordance with TSD policy."
❑ What does this choice mean in practice? Three examples:
  ▪ CD-R and DVD should be stored in appropriate places (with appropriate temperatures and relative humidity).
  ▪ It is allowed –if needed– to make copies to ensure long-term preservation.
  ▪ For files stored on the Local Network, back-ups are made.

National Archives of Australia

- ❑ < An Approach to the Preservation of Digital Records
- ❑ http://www.naa.gov.au/images/an-approach-green-paper_tcm2-888.pdf
- ❑ p. 14: "The digital preservation program must be able to preserve any digital record that is brought into National Archives' custody regardless of the application or system it is from or data format it is stored in."
- ❑ What does this choice mean in practice? One example:
  - ▪ all records that are accepted, should be preserved, regardless file format, medium, application, etc.
  - ▪ transform to open standard + keep 'original' format

ISO/TR 18492:2005

❑ International standard: Long-term preservation of electronic document-based information

❑ p. 12: Migration to standard formats

Storage repositories should consider migrating electronic document-based information from the wide variety of formats used by creators or recipients to a smaller number of "standardized" formats upon their transfer to the custody of the repository.

"Standardized" formats could be a consensus on formats that are widely used and are likely to cover a majority of a particular class of electronic document-based information. Proprietary file formats should be avoided. Among the technology neutral formats that merit consideration are PDF/A-1, XML, TIFF and JPEG.

Preliminary conclusion

❑ The variety of requirements in institutional policy documents is very large

❑ Some of these requirements are very general, others are a bit more specific.

❑ Questions

▪ How does one make these requirements operational for preservation planning purposes?

▪ What are 'valuable' requirements in a process of preservation planning?

▪ …

Policies

❑ Benefits of digital preservation policy

< ERPANET research, 2003

❑ Examples of benefits

- Planning coherent digital preservation programs
- Ensuring *accountability*
- Allocation of funds
- Ensuring digital materials are available for current and future use
- Defining significant properties that need to be preserved for particular types of resources
- Providing a comprehensive statement on digital preservation
- Providing security measures

Policies (2)

❑ The framework allows that various requirements are identified

❑ However, on this high level, it is possible and feasible that some requirements are already made explicit.

❑ Examples:

- choice for one strategy (e.g. migration to open document format)
- choice that some types of records/documents can be denied because e.g. an *exotic* file format is used
- …

Policies (3)

❑ Positive aspects related to preservation planning
  - Some strategies are excluded as viable solutions: no investments have to be made
  - Some requirements are identified at a high level
  - …

❑ Possibly negative aspects
  - Some decisions/choices may prove to be difficult to implement
  - At this moment, lack of information about some strategies
    -> possibility to miss interesting new solutions
  - …

User perspective

❑ Goal of digital preservation is to serve (future) users in providing usable and authentic information

❑ What are needs/requirements of users?
  ▪ easy access
  ▪ knowledge about origin of documents/ to be able to interpret them
  ▪ to use them for their own convenience

❑ Example requirements
  ▪ some users prefer that all information is presented in a uniform way
  ▪ some users prefer that they can search full-text in documents
    • consequence: don't migrate texts to image files
  ▪ …

User perspective (2)

- ❏ Some user requirements will affect decisions for preservation actions
  - Different manifestations of 'deliverable units'?
- ❏ However, difference between preservation and presentation copy
  - Not necessarily the same 'object'/ format...
  - Should it be the same, in order to reduce costs for preservation strategies/ actions?
  - Providing presentation copy 'on the fly'/ on demand?
- ❏ Not all users necessarily have the same requirements
  - requirements based on user segmentation

User perspective (3)

❑ Difficult to assess what users want in the digital era
  ▪ because they are (often) not used to work with digital information/documents
  ▪ because they are possibly not aware of the possibilities of different ways of presentation

❑ Compilation of requirements is based on user studies in which the participants combine the use of paper and digital documents

❑ Predict what users in the future want is impossible….!

'Technical' requirements

❑ These could be based on requirements as identified by users or the institution

❑ Is it –from a user's point of view- necessary that the page layout is as close as possible to the 'original'?
  ▪ Yes -> page layout, fonts, links, headers & footers, titles & subtitles should be preserved
    • migrate
    • emulate
  ▪ No -> only the textual content is important
    • don't bother about emulation
    • migrate to an 'easy' format

'Technical' requirements

❑ Is it –from the institutional perspective– important as part of authenticity requirements that specific lay-out characteristics are preserved?

▪ Yes -> then choose the migration or emulation strategy that best corresponds to this requirement

▪ No -> then choose whatever strategy that seems the best

❑ Similarly: behavioural or dynamic characteristics?

Essential characteristics of 'digital objects'

❑ What needs to be preserved?
- content
- context
- structure
- form / appearance
- (sometimes) behaviour

❑ What criteria for determining these essential characteristics?