



Long Term Data Storage: Are We Getting Closer to a Solution

A.Stander, S. Rossouw &
N. van der Merwe



Introduction

- Problems
 - Storage media unstable
 - Equipment rapidly outdated
 - Software versions incompatible



Archival Format

- Possible solution is to strip away machine and software dependence and to recopy it regularly to newer media
- Convert to common character format such as ASCII or Unicode
- Unicode supports all characters across all platforms, languages and programs
- Standards needed for data that is only available online



Problems

- Copyright when recopying
- Software needed to make digital documents accessible often outdated or cannot run on later versions of the operating system
- Later hardware often incompatible



SGML & XML

- Standard Generalised Markup Language (SGML) – standard for representing texts in electronic format
- Extensible Markup Language (XML) – cut down version for ease of use
- Can be used to store and transfer any kind of structured data between different computer systems



SGML (cont..)

- Descriptive markup system that allows different processing instructions to be associated with the same part of the file eg. To extract data or to format it
- A special file, the Document Type Definition (DTD) defines the contents of the file and allows for it to be checked by a special program (parser)



SGML & XML (cont...)

- Data Independence is facilitated by SGML & XML by the provision of a mechanism that allows the replacement of a particular string by another so that different computer systems can understand each other's character sets



Magnetic Media

- Tends to lose magnetism and must be regularly rewritten
- Substrate also deteriorates
- Needs proper storage and operating environments



Optical Media

- CD-ROM (Compact Disc - Read Only Memory)
- CD-R (Compact Disc -Recordable)
- DVD-ROM (Digital Versatile Disc - Read Only Memory)



Optical Media

- Use laser to read data stored as series of pits in metallic layer
- CD-R uses dye layer that is changed by laser light (can be affected by strong light)
- Long lifetime attractive



Media Life

- D3 Magnetic tape: 1 – 50 years
- DLT magnetic tape cartridge: 1 – 75 years
- CD/DVD: 2 – 75 years
- CD-ROM: 3 months – 30 years



Smaller Amounts of data

- Store as XML on CD-ROM or CD-R in ISO 9660 format
- Regularly recopy
- Use suitable storage facilities
- Create policies to make data available to other users



Management

- Standards needed for data that depends on software eg. To take snapshots for archival
- Standards needed for online data
- Proper migration policies needed
- Legislative issues must be addressed



Conclusion

- Software & Hardware independent data and optical media can solve many of the current problems
- Standards are needed for transient data and proper migration policies must be developed