



# TSM

## part of our cultural heritage?

Gerhard Schneider

University of Freiburg, Germany



# Off topic: an experiment

- Outsourcing: get rid of your current problems, pay money, lose knowledge
- Insourcing: grow, become more important and listen to your customers
  - Who have just lost their knowledge
- Universities compete
  - Giving resources to other universities may be a strategic disadvantage
    - Out of sight (money...)
    - Needs of your own researchers (and research)



# Off topic: an experiment

- Heidelberg and Freiburg try a different way:
  - The servers remain at both ends
    - Although the network is powerful enough
  - HD operates both servers and FR helps
    - And therefore maintains some competence
  - Both sides reinvest
    - Not much saving if you invest on one side only
    - And have something to show to visitors
    - 2 sites (disaster recovery) for free
    - Both parties continue to keep their own data



– *joint-sourcing?*

# The past 15 years

- In 1995
  - a database size of 2GB was considered to be excessively large
  - Restore of 10 GB of user data was a major task
  - An IBM cartridge tape stored about 10 GB
- Today
  - Petabytes are on everybody's mind
  - User data consists of more than just a few lines of code
    - Photos, scans,....
  - User data is living much longer
    - In my younger days I needed my data for a few years only



# What is stored?

- Backup and archival of PCs, mainframes, servers
  - Originally to repair a server crash
- Today we still restore data that was archived 15 years ago
- More institutions are generating data
  - Libraries: digital scans of (valuable) books
    - You don't want to scan again after 15 years
  - Student admission office
    - Keep the data for decades
  - Scientific data from expeditions
    - Spacecraft, expensive experiments,
- and want to retrieve their data in n years (n=??)



# Ongoing EU projects

- Planets
- Caspar
- Nestor
- are addressing issues of long term archival of digital data
- Technical solution sound easy – at first
  - Copy from old media to new media
    - How long does it take to migrate petabytes?
  - Who owns the data?
    - Owner may be gone – does the new owner (TSM definition) have the right to own the data?
  - What does the data mean?



# Nestor: 10 principles for trusted digital repositories

- The repository commits to continuing maintenance of digital objects for identified community/communities.
- Demonstrates organizational fitness (including financial, staffing structure, and processes) to fulfill its commitment.
- Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.
- Has an effective and efficient policy framework
- Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
- Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.
- Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.
- Fulfills requisite dissemination requirements.
- Has a strategic program for preservation planning and action.
- Has technical infrastructure adequate to continuing maintenance and security of its digital objects.



# Questions from my talk 4 years ago

- Will we be able to
  - load data onto a platform in 50 years?
    - Try to load unix data onto windows platforms today ☹
  - understand what to do with the data?
- What are the data formats that have a chance to survive the many fashion trends in computer science?
  - Tif, openxml, .....
  - Royalty free formats
- Notion of living data
  - Requires an environment to execute
    - Like games

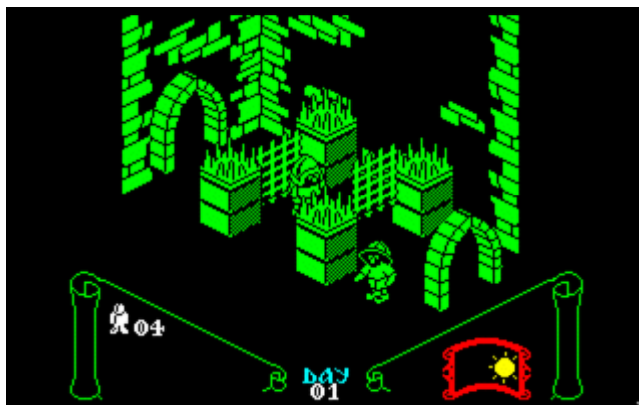




# Living Data ☺



Commodore C64



Sinclair ZX



Atari



# Emulation

- Emulation (Vmware, XEN) are possible approaches towards keeping an environment alive
  - Several Ph.D. Theses on the way
  - Store complete images of the opsys environment and the data environment together with the information to rebuild it on a new system
  - Run directly from the opac system of your library.



# Ask again: what is stored?

- Today we see more and more data
  - that is vital to (parts of) society
  - Impossible to regenerate
    - Lack of resources (e.g. time)
  - With a life expectancy of 50 years or more
    - Digital photos should last at least as long as their b&w predecessors.
    - Why not 500 years? Like the Bodleian....
- Can we trust our data to TSM?
  - i.e. will we be able to get it out again?
    - In 50 years??



# Prepare the path

- How to worry about the future?
  - Like NASA sending a plate on a spacecraft to outer space: will they be able to read
  - Like Robinson Crusoe: sit at the beach with your tape and try to understand what is on it.
  - Most of the sophisticated talks of today will be forgotten in 50 years (old knowledge)
    - How do we really operate a steam engine?
    - At least two generations of TSM admins



# The easy step

- Fight the loss of media with migration
  - Who can still read 8“ floppies?
    - Who knows that they once existed?
  - Use TSM to copy all the data onto the most modern media
    - And throw away the old media (10GB tapes...)
- How long does it take to migrate 1 PByte?
  - When LRZ moved to its new premises, tapes were moved by truck – so much for networks
  - How much of your investment is necessary just for migration?



# TSM – a choice to store our cultural heritage?

- Institutions produce data without knowledge of the background mechanisms
  - Servers, backup, archive
- Are the internal storage formats of TSM known?
  - For digital archeology in 2100
  - Suppose you find a TSM tape in 2100 and there is no computer to run TSM
- Is there a way to export data from TSM without any loss of information to some future (obscure) system?
  - Data files
  - Metadata, like file system structure, dependencies
  - data base contents



# TSM in 2057

- Many institutions trust their data to TSM:
  - Oxford University Library, Deutsche Nationalbibliothek, SUB Göttingen
- What about file formats?
  - Cross platform today (Unix, Windows) and tomorrow
- What about user interface
  - The Archive Interface is not quite what you would expect in 50 years
  - Home grown solutions/frontends will not survive



What about encryption ????

# TSM – a choice to store our cultural heritage?

- TSM must offer a conversion or export to open standards
  - This version does not have to be efficient
  - Survival of data is what counts
- What are the open standards that we need?
- Preserving the digital cultural heritage is becoming an issue
- TSM could be a candidate to help solve this issue
- But it has to evolve
  - Or deteriorate into a system for today's use only





# The future

In other words:

- There is a huge market for the TSM developers
  - Job security better than civil service
  - *ditto* for those who run the day-to-day business
  - Librarians will love you
- Would you contribute to the heritage issue?
  - In terms of discussion, solution, open formats, export facilities, file formats, simple archive, etc.

