Personal Archiving

HOPPLA Assumptions and Design Decisions

Andreas Rauber

Department of Software Technology and
Interactive Systems

Vienna University of Technology

http://www.ifs.tuwien.ac.at/~andi

Assumptions / Design Decisions

- Archives of personal data vs. personally archiving
  - persons archiving their own data
  - little to no expertise on preservation issues and IT
  - little cooperation from the user (metadata, …)
  - robustness against interfering user (deleting objects)
  - full automation
  - outsourcing of expertise
  - 80/20 rule, best-practice solutions
  - trade-offs in terms of optimizing tools to specific user needs, focusing on dominant object types

Assumptions / Design Decisions

- System has full responsibility for data preservation
  - bit-stream preservation:
    - managing multiple back-ups on different media
    - manages risks of different storage media
  - logical preservation
    - system manages migrations
    - extracts and stores metadata
    - rules and tools provided by expert-operated web service
  - configuration for levels of risk-mitigation
    - trade-off between storage space and (bit-stream & logical) redundancy
    - quality/cost trade-off for tools, verification,…

Assumptions / Design Decisions

- Stability and system independence
  - plain file system storage & redundant XML metadata
  - original directory structure / e-mail folder structure etc. is re-created on archival media.
  - objects are stored explicitly, essential metadata encoded in filenames
    <name>_<timestamp>[_<tool/rule_applied>_<timestamp>].<ext>
  - detailed metadata stored in XML file per directory tracking all back-up versions, migration versions and paths, potentially even log-file output of migration/validation process
  - can be used by any system without application as long as file system is readable
  - data management can be recreated from XML

Assumptions / Design Decisions

- Trust and accountability, Privacy
  - data stays always with the user
    (may decide to rely on on-line storage)
  - control over information sent to server (file types, sizes, detailed technical characteristics, …)
  - aim to fulfill core requirements of audit and certification initiatives
  - system creates required log-files
  - underlying preservation plans can be provided by server

Assumptions / Design Decisions

- Workflow (rough sketch)
  - User (automatically) performs back-up of registered sources
  - System analyses object formats
  - checks local rule repository to see whether any of them need to be migrated
  - for new object types/characteristics: contact Web-server
  - Server provides best-practice recommendation tools and rules
    - migrate certain type of files up to specific size with some characteristics with tool A, larger files with tool B
  - system regularly checks for updated rules/recommendations
  - system performs migrations of data in archive
  - system informs user which ext. storage media to insert for back-ups

IfS FACULTY OF !NFORMATICS

HOPPLA

- **H**ome **O**ffice **P**ainless **P**ersistent **L**ong-term **A**rchiving

- Based on research initiated in Planets

- Development of prototype within DME

- Further information:

  http://www.ifs.tuwien.ac.at/dp/hoppla