



South-Eastern Finland University of Applied Sciences

Xamk is the 5th largest UAS in Finland

- 9,000 students
- 750 employees
- 1,200 graduates per year
- 60 (approximately) Bachelor's programmes
- 23 Master's programmes
- 4 campuses
- 350 partner institutions around the world



Xamk



Digitalia Research Center

- Digitalia supports and advances the maintenance and preservation of digital information and data.
- Merely obtaining data does not create value. Data must first be transformed into well-defined information, whose origins, features and meaning are known.
- Our mission is to enhance the usability, use and availability of digital data. We also make sure that data is preservable and reliable.

Digitalia

Research Centre for Digital Information



Noora Talsi
Research director
Project manager
Noora.Talsi@xamk.fi



Anssi Jääskeläinen
RDI advisor
Anssi.jaaskelainen@xamk.fi



Miia Kosonen
RDI advisor
Miia.Kosonen@xamk.fi



Liisa Uosukainen
IT specialist
Liisa.Uosukainen@xamk.fi

Legal Aspects

- Finnish Personal Data Act (523/1999)
- EU GDPR (2016/679)
- Obligations for the service provider
- Rights for the user
- Data privacy on a compatible level
- Content supervision

Accessible

- No browser limitation
- No content/file type limitation
- Affordable pricing (forthcoming)

Citizen Archive

My Data

- Citizen in control of his/her own data.

Technical Specs

- OSA - Java application
- Fedora Commons
- Apache Solr
- MariaDB, MongoDB, OpenLDAP

E-Mails In Archive

- National archive solutions are based on preserving the original file as OAIS (Open Archival Information System) AIP (Archival Information Package) packages or XML structure.
- Requires SIP submit

Why E-Mails

Some might be even culturally interesting → heritage

2.56 Million E-Mails Every Second

Online Communication De-Facto

Organizational Power Structures

Knowledge Sharing Patterns

Personal Life Stories

Reveals

- Power of the newspaper head journalist
- Political influence

Starting Points

No decent automated way to do conversion

Format Obsolescence

Proprietary Formats

Conversion Responsibility?

Is .pst file from Outlook 2002 still accessible?

How to open .pst file after Outlook is gone..?

Solution

Automatic conversion from proprietary format to PDF/A-3

- Pffexport
- Convertmetadata.jar
- Ted/Txt2Html
- Imagemagick
- Oowriter
- Wkhtmltopdf
- Ghostscript
- Verapdf

Open Source

Bound Together With Python & Java

Supports any attachment type

PDF/A-3B

Every Original Metadata Included

Metadata

Conversion & Insert To Pdf/A-3B

Attachments

Future Development

.mbox, .msg, .eml

More Supported E-Mail Formats

Independent Test Site

Drop .pst, get pdf files in a downloadable zip

Family Archives

Starting Points

Rely On Usb/Optical

Terms & Conditions Of Clouds

Official Digital Repositories Out Of Reach

Transparency

Difficult to restore data from a broken device

Physical location of the information, inside own country, EU, US?

Precious Information

Pictures

Contracts

Accounting

Mydata

Fitbit

Sports Tracker

Health Measurements

Location Information

New data types and their metadata fields can be determined (data classification)

protect fragile originals and enrich content with metadata

Currently inside manufacturer / vendor clouds. Not accessible via APIs

Increased Interest

Digital Heritage

Digital Footprint

My Life Story

Generation that experienced the WW2 is getting very sparse

"Stories From The War"

Sharing With Distant Relatives

Create user accounts and determine access rights for others to share your data

Future Development

More Facilitation Functionality

Ucd Development According To Feedback

Automate archiving tasks using workflow engine and micro-services

- More customized views.
- Improved search and navigation properties.
- Improved interaction, f.ex space to display and introduce some data sets for others in the archive

Original e-mail (Outlook 2013)

Created PDF/A-3b E-mail (Acrobat DC)

E-mail metadata inside PDF/A-3b file

www.digitalia.fi

Archive front page

Creation of an archive

Advanced search

Search functionality

Leverage from the EU 2014-2020

