

ODF Workshop

29-30 October, Berlin



Federal Foreign Office

OPEN DOCUMENT FORMAT

ODF ALLIANCE

HWD:ODFTrans 2.0

a meta-converter approach

Hans W. Dahm

HWD IT-Consulting

data migration/data conversion

Examples:
change of

operating
system

messaging
system

office suite

document conversion

- not a „one time move“
- ongoing process
- develop a migration and coexistence strategy
- exchange data with people or systems
 - within the organization
 - outside the organization
- need for appropriate tools
 - intuitively to use
 - users should know about the quality

From idea to product

- developing a migration and coexistence strategy
- what about interoperability?
- available converters
 - OpenXML/ODF Translator Addin for Office (MS/Clever Age Addin)
 - Sun Plugin for Microsoft Office
 - OpenOffice.org
- MS/CleverAge Addin failed to install on MS-OfficeXP
- central conversion possible?

From idea to product

- starting point was the existing tool email2PDF
- email2PDF
 - converts documents (and other formats) to pdf
 - email-based
 - central service
 - no additional software at the client-side
 - no need for training or special instructions
 - no license fees 😊

From idea to product

- HWD:ODFTrans 1.0 with MS/Clever Age Addin only
use of commandline translator
results were not satisfying
sometimes no output

HWD:ODFTrans 2.0

ODFTrans is a metaconverter using

MS/Clever Age Addin

Sun ODF-Plugin

OpenOffice.org API

- central service for >10.000 user in the Federal Foreign Office
- other German authorities are evaluating

How to use?

- compose a new mail
- attach documents to be converted
- send a simple mail to the converter mailbox
- optional: specify a special output format or engine

DOC_ODT_SUN

DOC_ODT_OO

DOC_ODT_MS

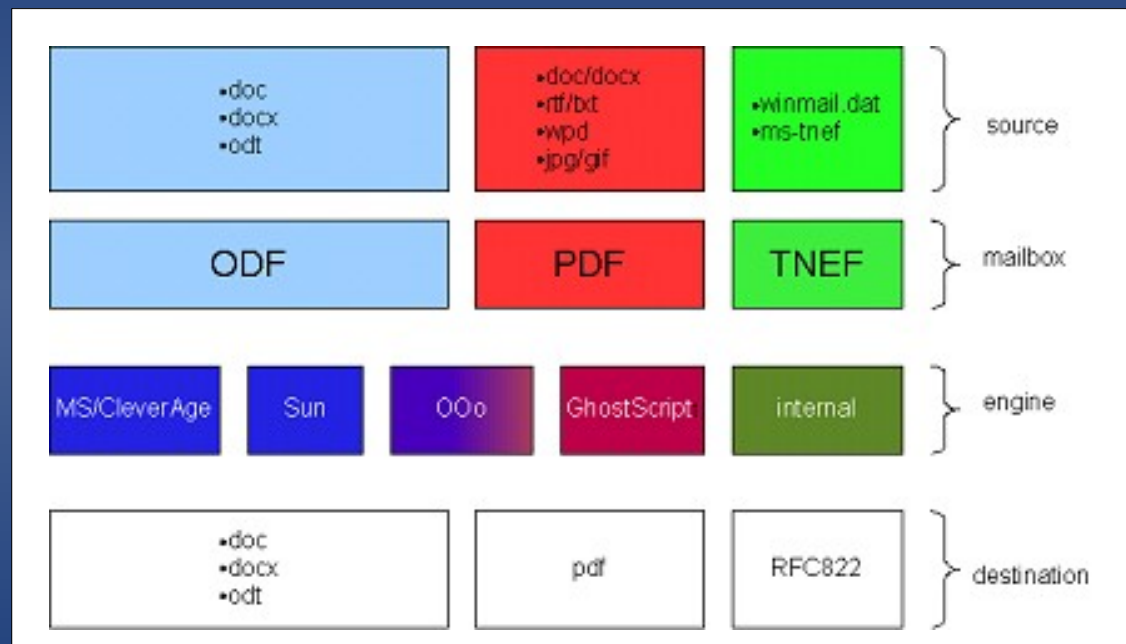
ODT_DOC_SUN

...

Central conversion

- ODFTrans polls the mailbox at regular intervals
- If a job is found, the requested conversion is done
- output (or an error report) is sent back to the user
- if the requested (or default) engine fails, another engine will be used instead

Features



Software requirements

- MS Windows (XP SP2, 2003 Server)
- MS Office 2003 SP1
- .Net 2.0 framework
- compatibility pack
- MS/CleverAge Addin
- Sun Plugin
- OpenOffice
- optional: ghostscript (for pdf)

Advantages

- easy to use
- no client-side installation (e.g .NET framework)
- no need for upgrading MS Office
- MS-Office based plugins and addins can be used by Linux clients

Disadvantages

- access restrictions
- local resources (e.g. fonts)
- password protection
- waiting time (mailqueues, heavy load)

- transport of confidential documents
(but: central conversion does not mean public conversion)

experience report on using different converter engines

- fidelity
- performance
- size difference

Fidelity

Fidelity is a measure of similarity between the source document and the output. Full or perfect fidelity is the goal.

- No converter achieves full fidelity
- The majority of documents is converted correctly related to the content
- differences in layout (often marginal)
- Problems arise, if the content has been modified

Example 1: content changed

original

- 1. Part 1
 - a) Part a:
 - b) Part b:
 - aa) Part aa
 - bb) Part bb

- 1. Part 1:
 - a) Part a:
 - b) Part b:
 - b.a) Part aa
 - b.b) Part bb

MS Clever Age

- 1. Part 1:
 - a) Part a:
 - b) Part b:
 - b.a) Part aa
 - b.b) Part bb

Sun

- 1. Part 1:
 - a) Part a:
 - b) Part b:
 - b.a) Part aa
 - b.b) Part bb

OpenOffice.org

Example 2: layout problems



original



MS Clever Age



Sun



OpenOffice.org

Example 3: revealing hidden information



original



MS Clever Age

Performance

Text conversion

Source MS Word Document	CleverAge	Sun	OpenOffice.org
1 page– text only	11 sec	8 sec	1 sec
2 pages– text only	11 sec	8 sec	1 sec
3 pages– text only	11 sec	8 sec	1 sec
10 pages– text only	11 sec	8 sec	1 sec
100 pages– text only	12 sec	10 sec	2 sec

Source ODT Document	CleverAge/ MS	Sun	OpenOffice.org
1 page– text only	13 sec	2 sec	1 sec
2 pages– text only	13 sec	2 sec	1 sec
3 pages– text only	13 sec	2 sec	1 sec
10 pages– text only	13 sec	2 sec	1 sec
100 pages– text only	14 sec	3 sec	1 sec

Performance

Graphics

Source MS Word Document	CleverAge/ MS	Sun	OpenOffice.org
1 page– 1 graphic	13 sec	2 sec	1 sec
2 p. – 1 graphic each	13 sec	2 sec	1 sec
3 p. – 1 graphic each	13 sec	2 sec	1 sec
10 p. – 1 graphic each	14 sec	3 sec	1 sec

Source ODT Document	CleverAge/ MS	Sun	OpenOffice.org
1 page– 1 graphic	12 sec	3 sec	1 sec
2 p. – 1 graphic each	12 sec	4 sec	1 sec
3 p. – 1 graphic each	12 sec	5 sec	1 sec
10 p. – 1 graphic each	12 sec	12 sec	1 sec

Performance

Text conversion

source: doc „text only“ target: odt	CleverAge	Sun	OpenOffice.org
1 page– text only	6 sec	2 sec	1 sec
2 pages– text only	6 sec	2 sec	1 sec
3 pages– text only	6 sec	2 sec	1 sec
10 pages– text only	6 sec	2 sec	1 sec
100 pages– text only	7 sec	5 sec	2 sec

source: odt „text only“ target: doc	CleverAge/ MS	Sun	OpenOffice.org
1 page– text only	5 sec	1 sec	1 sec
2 pages– text only	5 sec	1 sec	1 sec
3 pages– text only	5 sec	1 sec	1 sec
10 pages– text only	5 sec	1 sec	1 sec
100 pages– text only	6 sec	1 sec	1 sec

Performance

Graphics

source: doc target: odt	CleverAge/ MS	Sun	OpenOffice.org
1 page– 1 graphic	7 sec	3 sec	1 sec
2 p. – 1 graphic each	8 sec	3 sec	1 sec
3 p. – 1 graphic each	8 sec	3 sec	1 sec
10 p. – 1 graphic each	11 sec	4 sec	1 sec

source: odt target: doc	CleverAge/ MS	Sun	OpenOffice.org
1 page– 1 graphic	7 sec	2 sec	1 sec
2 p. – 1 graphic each	8 sec	2 sec	1 sec
3 p. – 1 graphic each	8 sec	2 sec	1 sec
10 p. – 1 graphic each	8 sec	5 sec	1 sec

Size

Text

source: doc „text only“ target: odt	Size	CleverAge/ MS	Sun	OOo
1 page	31 kB	4 kB	8 kB	8 kB
2 pages	34 kB	4 kB	9 kB	9 kB
3 pages	37 kB	4 kB	10 kB	10 kB
10 pages	70 kB	4 kB	12 kB	12 kB
100 pages	466 kB	6 kB	13 kB	13 kB

source: odt „text only“ target: doc	Size	CleverAge/ MS	Sun	OOo
1 page	8 kB	27 kB	23 kB	70 kB
2 pages	9 kB	30 kB	26 kB	76 kB
3 pages	10 kB	32 kB	29 kB	82 kB
10 pages	12 kB	63 kB	61 kB	147 kB
100 pages	13 kB	459 kB	457 kB	938 kB !!

Size

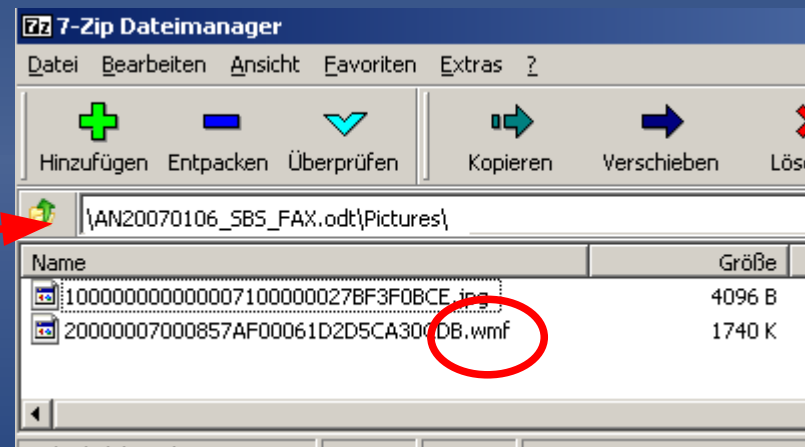
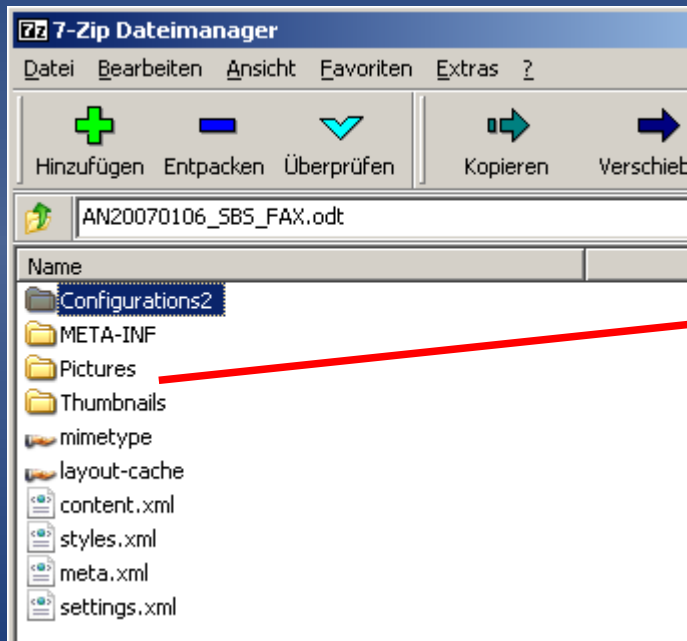
Graphics

source: doc target: odt	Size	CleverAge/ MS	Sun	OOo
1 page– 1 graphic	48 kB	27 kB	37 kB	37 kB
2 p. – 1 graphic each	74 kB	27 kB	37 kB	37 kB
3 p. – 1 graphic each	100 kB	27 kB	37 kB	37 kB
10 p. – 1 graphic each	281 kB	27 kB	37 kB	37 kB

source: odt target: doc	Größe	CleverAge/ MS	Sun	OOo
1 page– 1 graphic	75 kB	84 kB	84 kB	126 kB
2 p. – 1 graphic each	76 kB	84 kB	84 kB	127 kB
3 p. – 1 graphic each	76 kB	85 kB	84 kB	127 kB
10 p. – 1 graphic each	76 kB	86 kB	84 kB	128 kB

Problems

Firewall-Restrictions



Conclusion

- no full fidelity
- need to check the output
- roundtripping or workflow need agreements on what formats and objects are allowed
- improving fidelity will help odf adoption

Summary

- People often have documents in the wrong“ format (mail archives, zip files etc.)
- HWD:ODFTrans is a tool to give people a quick and comfortable access to the content of office documents.
- setting up special workstations is expensive and inconvenient
- HWD:ODFTrans can help organizations on their way to ODF

Outlook

- HWD:ODFTrans is free.
- We are working on the next release
 - with support for spreadsheets and presentations
 - with a web interface to easily upload documents for conversion
 - as FOSS under the terms of the GPL

Thank you for your attention

for further information: hwd@hwd-online.de