

Administration of documentation projects

Organization, administration,
and synchronization of files

Sebastian Drude

DOBES-Workshop June 2008, Nijmegen

Topics

1. The challenge
2. Organization of project files
3. Administration of data files
4. Synchronization of computers and backups

june 2008

DOBES – file management

Topics

1. **The challenge**
2. Organization of project files
3. Administration of data files
4. Synchronization of computers and backups

june 2008

DOBES – file management

1 The challenge

Typical situation:

- Large number of files, growing
- Many have to be checked or processed
- Number of sessions grows, each with different workflow status
- Many different types of related files
- Different versions of the same file
- Different computers – individual laptops, in the field, office, etc.

june 2008

DOBES – file management

1 The challenge

Question:

- How not to get lost in so many files, and different versions of files, on different computers and DVDs, HDs?
- People tend to underestimate the importance of consistent solutions
- These are practical questions, not scientific and even not specifically documentation-methodological

june 2008

DOBES – file management

Topics

1. The challenge
2. Organization of project files
3. Administration of data files
4. Synchronization of computers and backups

june 2008

DOBES – file management

2 Organization of project files

Recommendations:

- Consistent naming of files
- ONE identical directory structure ('tree') for the project on all computers
- Create a virtual drive for this 'tree'
- Use an adequate file manager
- Regular backups and synchronizations

june 2008

DOBES – file management

2 Organization of files: names

- All related files (e.g., same session) should have similar file names, starting identically
- DOBES conventions for names of sessions:
AWSDAM23Jun0201 (-S01)
- Names of DMF and some intermediate files were according to these conventions
- Preliminary annotation files may be named differently:
23Jun02-01-amanajup.sft
- Final names of sessions and associated files may be different, e.g., due to anonymization
153-biogr.sft

june 2008

DOBES – file management

2 Organization of files: one tree

... \ALDP\ can contain subfolders like:

Admin (Reports, Research Permissions,
Lists of files and sessions, Invoices...)

Contains folders such as:

VWS, FUNAI, Internal, Data, Financial

Papers+Results (scientific products)

Contains folders such as:

2008-06-DOBES-filemanagement

MPI-Corpus (the final data tree for uploading)

Media (all audio and video files, images...)

Data (files related to annotation and processing)

june 2008

DOBES -- file management

2 Organization of files: one tree

... \ALDP\Media\2002\ subfolders like:

Audio-Field (audio-files digitized in the field)

Contains files such as:

23Jun02-01-amanajup.wav

Audio-DMF (Digital Master-Files digitalized by the
MPI, each corresponding to one MD / cassette)

Contains files such as:

AWSDAM23Jun0201.wav

Audio-Sessions (ready cut audio files by MPI)

Contains files such as:

amjp-biogr.wav

Photos (Digital Photos or scanned material etc.)

june 2008

DOBES -- file management

2 Organization of files: one tree

... \ALDP\Data\2002\ subfolders like:

Transcriber (Transcriber XML-files such as:)

23Jun02-01-amanajup.trs

Toolbox-texts (Toolbox databases for texts:)

23Jun02-01-amanajup.txt (Econv-generated)

23Jun02-01-amanajup.sft (AG-Syntax format)

Toolbox-lists (Transcribed wordlists)

Print-out-versions (For RTF-files used for
elicitation and translation, such as:)

23Jun02-01-amanajup.doc

june 2008

DOBES -- file management

2 Organization of files: one tree

... \ALDP\Data\2002\ subfolders (cont.):

ELAN (Files converted to or annotated in ELAN)

23Jun02-01-amanajup.eaf

Metadata-work (Metadata files that have still
to be revised or completed, such as:)

AWSDAM23Jun0201-S01.imdi

Metadata-final (IMDI files as uploaded)

153-biogr.imdi

june 2008

DOBES -- file management

2 Organization of files: Drive 'P:'

- It is not only useful to have the same tree on all computers and other devices
- but also to have the tree always on the same (virtual) device, for instance 'P:\'
- In this way, all links and configurations (toolbox...) always remain correct
- It is possible to assign a device letter such as 'P' to a directory (=folder), using the program "SUBST"

june 2008

DOBES – file management

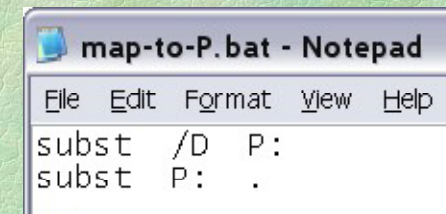
2 Organization of files: Drive 'P:'

For instance, the command

```
SUBST P: D:\ALDP
```

creates the unit/virtual drive **P:**, which in fact is the directory **D:\ALDP**.

I usually put little batch-files (....BAT) like the following in the 'root' directory of the tree.



```
map-to-P.bat - Notepad
File Edit Format View Help
subst /D P:
subst P: .
```

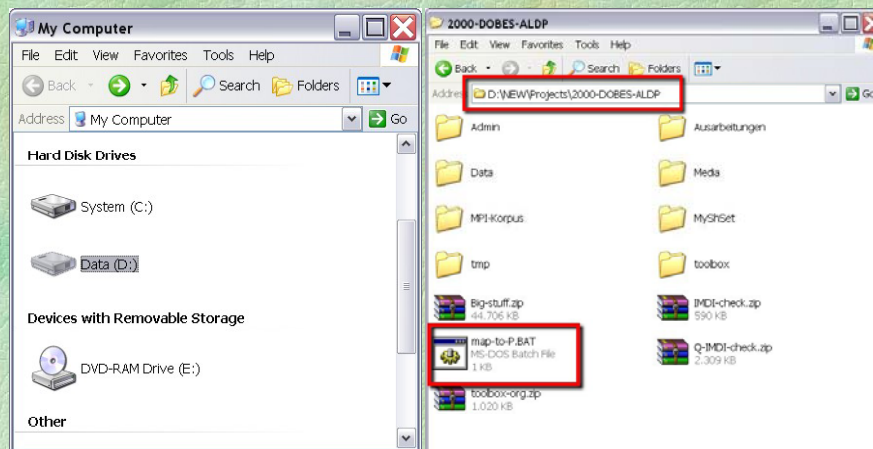
june 2008

DOBES – file management

2 Organization of files: Drive 'P:'

The drives before mapping

Root directory and batch file



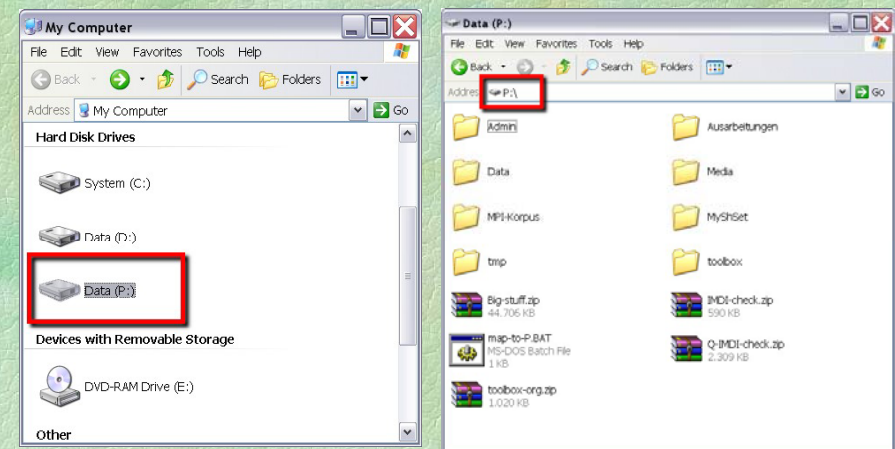
june 2008

DOBES – file management

2 Organization of files: Drive 'P:'

The drives after mapping

Root directory of drive P:\



june 2008

DOBES – file management

2 Organization of files

With such a setting, it is easier to synchronize two versions of the tree, wherever the tree is located ...

- ... for instance on different computers (even remotely, via ftp)
- ... or on a computer and a USB stick or external HD (as backup, or for transfer)
- For synchronization and administration of large quantities of files, I recommend the '**Total Commander**'

june 2008

DOBES -- file management

Topics

1. The challenge
2. Organization of project files
3. **Administration of data files**
4. Synchronization of computers and backups

june 2008

DOBES -- file management

3 Administration of data files

- Frequently, one needs to check a large number of files
- It is useful to produce lists of files, for instance for use with excel
- Using total commander, one can easily:
 - search for files (e.g., of certain types)
 - select the files and put them into a list, even in different directories
 - put their names (with paths) into the clip-board

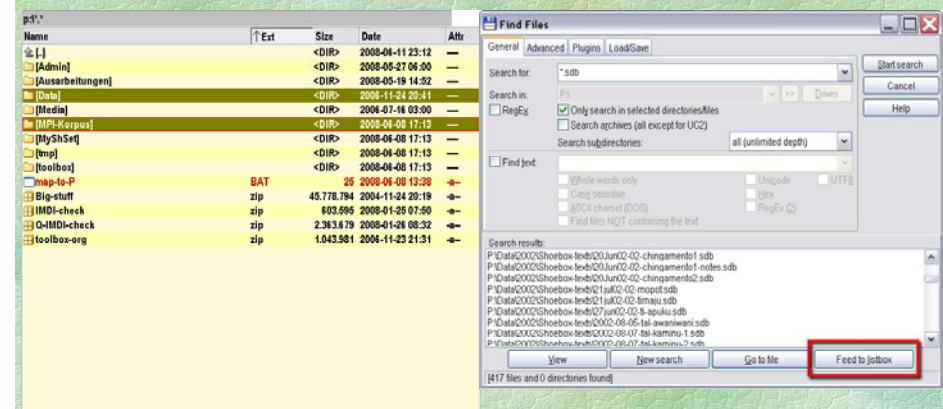
june 2008

DOBES -- file management

3 Administration: listing files

select folders to search

search options and results



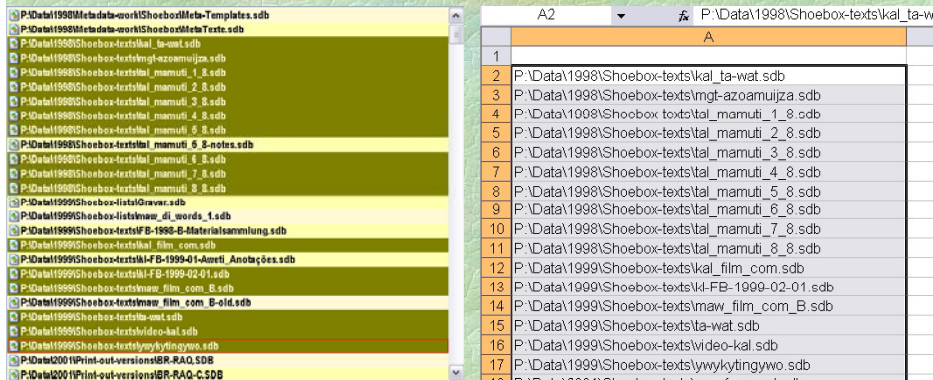
june 2008

DOBES -- file management

3 Administration: listing files

select files among matches
copy list into clipboard

paste list in spreadsheet



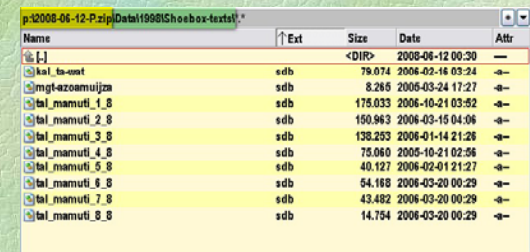
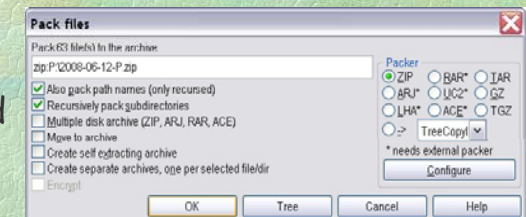
june 2008

DOBES – file management

3 Administration: zipping files

One can do other work with the same selection of matched files

for instance 'zip' the files, preserving the tree-structure
emailing such a 'zip' archive will help to maintain the tree on other computers



june 2008

DOBES – file management

3 Administration: toolbox datab.

Besides spreadsheets, one can use toolbox databases for keeping track of:

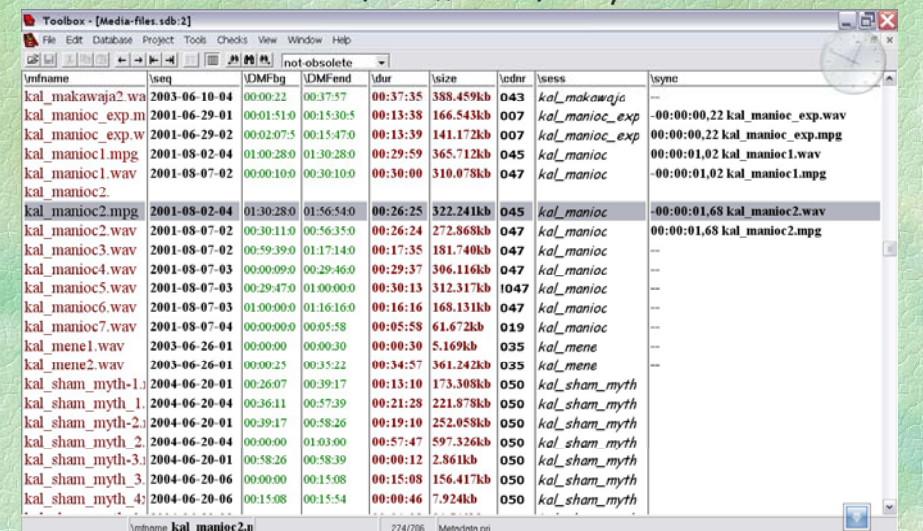
- Sessions (each record contains related files, basic metadata, action points etc.)
- Media files (each record has tape/DMF, start + end times, status etc.)
- Annotation files (each record can keep record of working progress, annotation types, who worked on which file, etc.)

june 2008

DOBES – file management

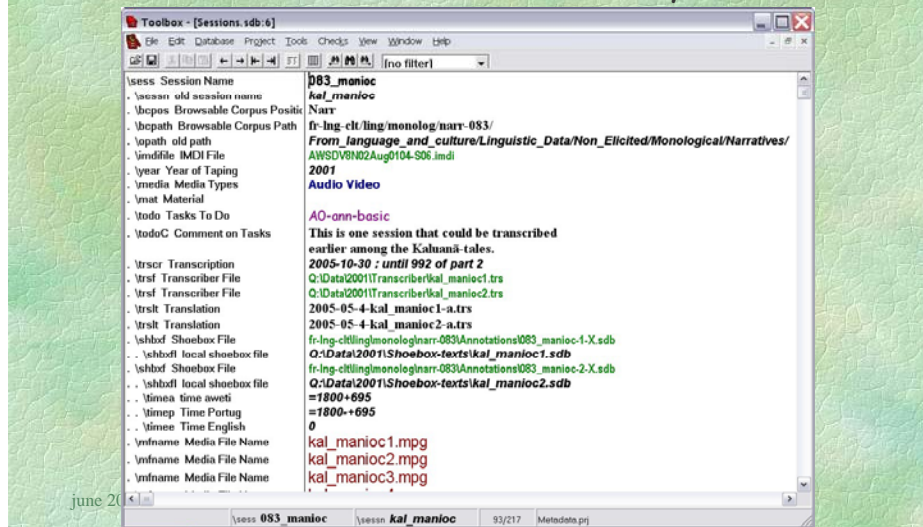
3 Administration: toolbox datab.

Toolbox database for media files, browse view



3 Administration: toolbox datab.

Toolbox database for sessions, entry view



Topics

1. The challenge
2. Organization of project files
3. Administration of data files
4. Synchronization of computers and backups

june 2008

DOBES – file management

4 Synchronization / backup

- It is essential to regularly synchronize the different computers so that all have the same current project tree (in 'P:\')
- There are several tools for synchronizing, I again recommend the 'Total Commander'
- Synchronizing with an external HD (or USB-pen-drive) is an efficient way to backup project data

june 2008

DOBES – file management

4 Synchronization / backup

Selection of two analogous directories, see size

Name	Ext	Size	Date	Attr	Name	Ext	Size	Date	Attr
[.]		<DIR>	2008-06-10 15:14	—	[.]		<DIR>	2008-06-10 13:24	—
[Intro+Manuals]		6.001.030	2008-06-10 07:58	—	[Intro+Manuals]		6.001.030	2008-06-10 07:58	—
[demo]		3.347.146	2008-06-10 12:03	—	[demo]		3.347.146	2008-06-10 12:03	—
[Installation]		2.807.527	2008-06-10 01:15	—	[Installation]		2.807.527	2008-06-10 01:15	—
[LIMA]		652.800	2008-06-10 07:42	—	[LIMA]		652.800	2008-06-10 07:42	—
[Utilities]		79.245	2008-06-10 07:42	—	[Utilities]		79.245	2008-06-10 07:42	—
azoamujza wav	wav	2.621.556	2008-06-14 15:27	-a-	azoamujza wav	wav	2.621.556	2008-06-14 15:27	-a-
Toolbox-intr. pps	pps	283.648	2007-07-11 16:30	-a-	Toolbox-intr. pps	pps	283.648	2007-07-11 16:30	-a-
Toolbox-intr. ppt	ppt	282.624	2008-06-10 15:14	-a-	Toolbox-intr. ppt	ppt	282.624	2008-06-10 15:14	-a-
Toolbox-intr. rtf	rtf	27.658	2008-06-10 09:00	-a-	Toolbox-intr. rtf	rtf	27.658	2008-06-10 09:00	-a-
Exercise prj	prj	2.067	2008-06-10 12:22	-a-	Exercise prj	prj	2.067	2008-06-10 12:22	-a-
azoamujza trs	trs	577	2008-06-14 19:50	-a-	azoamujza trs	trs	577	2008-06-14 19:50	-a-
azoamujza sdb	sdb	231	2008-06-14 19:59	-a-	azoamujza sdb	sdb	231	2008-06-14 19:59	-a-
map-to-X bat	bat	23	2008-06-10 10:13	-a-	map-to-X bat	bat	23	2008-06-10 10:13	-a-

0 b 15,6 M in 0/8 files, 0/5 dir(s)

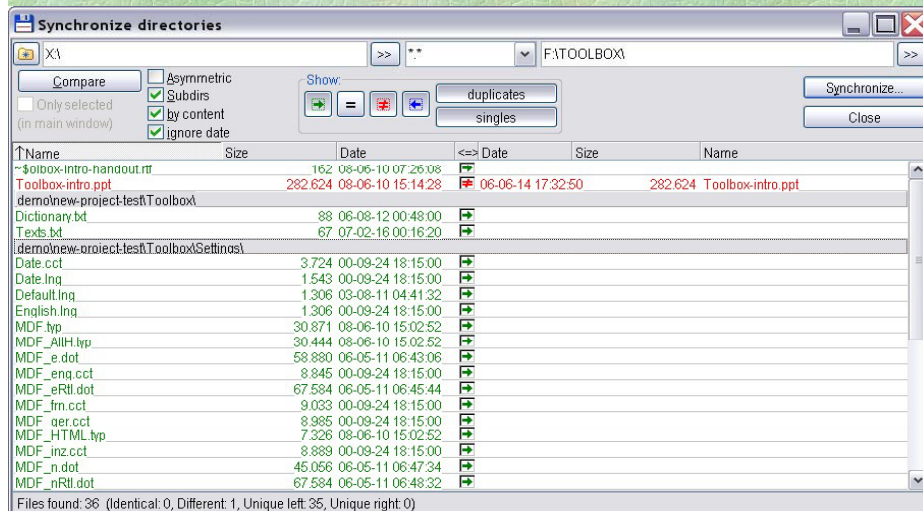
0 b 15,1 M in 0/8 files, 0/5 dir(s)

june 2008

DOBES – file management

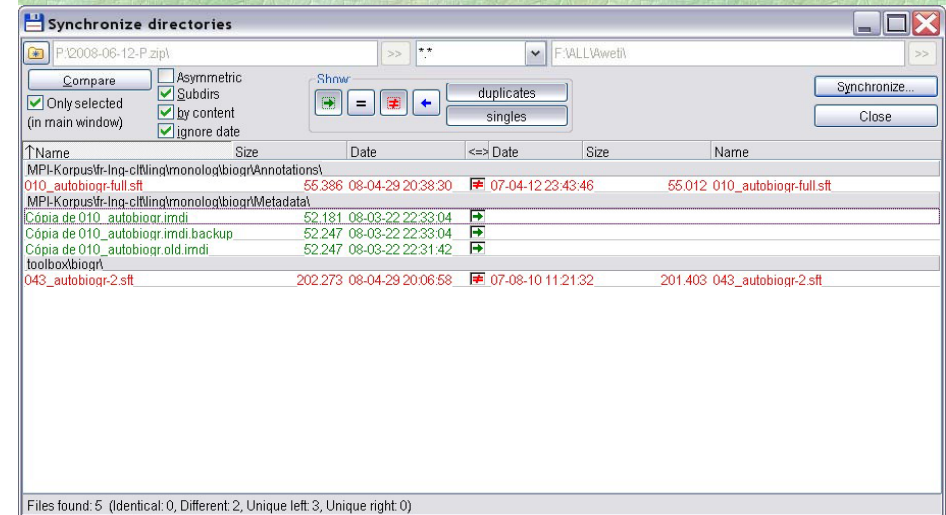
4 Synchronization / backup

Synchronize directories in TC



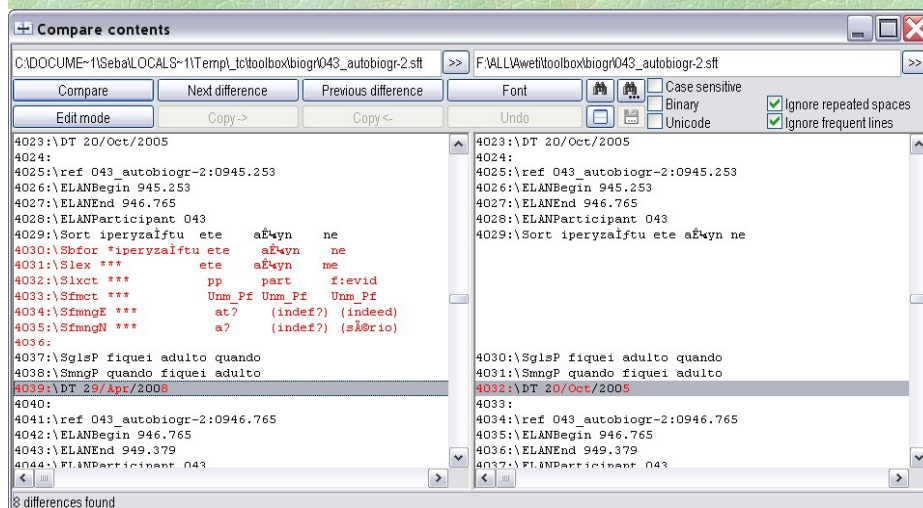
4 Synchronization / backup

Synchronize zip with directory in TC



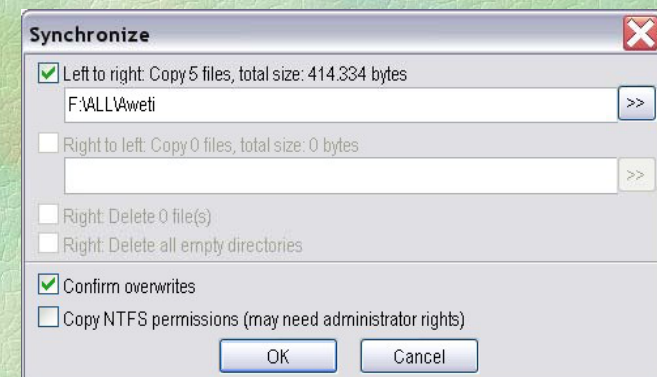
4 Synchronization / backup

Compare content of two text files in TC



4 Synchronization / backup

Synchronizing in TC: actions and targets



4 Synchronization / backup

- Using the possibility to change the target directory, one can easily keep older versions of deleted or changed files as a backup
- The synchronization can be done with a remote server (for instance, via FTP)

Get organized! Backup your data!
The sooner, the better...

Administration of documentation projects

Organization, administration,
and synchronization of files

Sebastian Drude

DOBES-Workshop June 2008, Nijmegen