

Jens O. Oberender:



simplify

your backup

IEEE SB Passau

2007-06-27

research **design** experience

<http://jo000.de>

Agenda

- Threats
- Research
- Best-Practice
- Strategies
- Challenges



Problems with your backup?



Threats

- Drive and media failures
- Malicious data modification
- Confidentialty



Threats

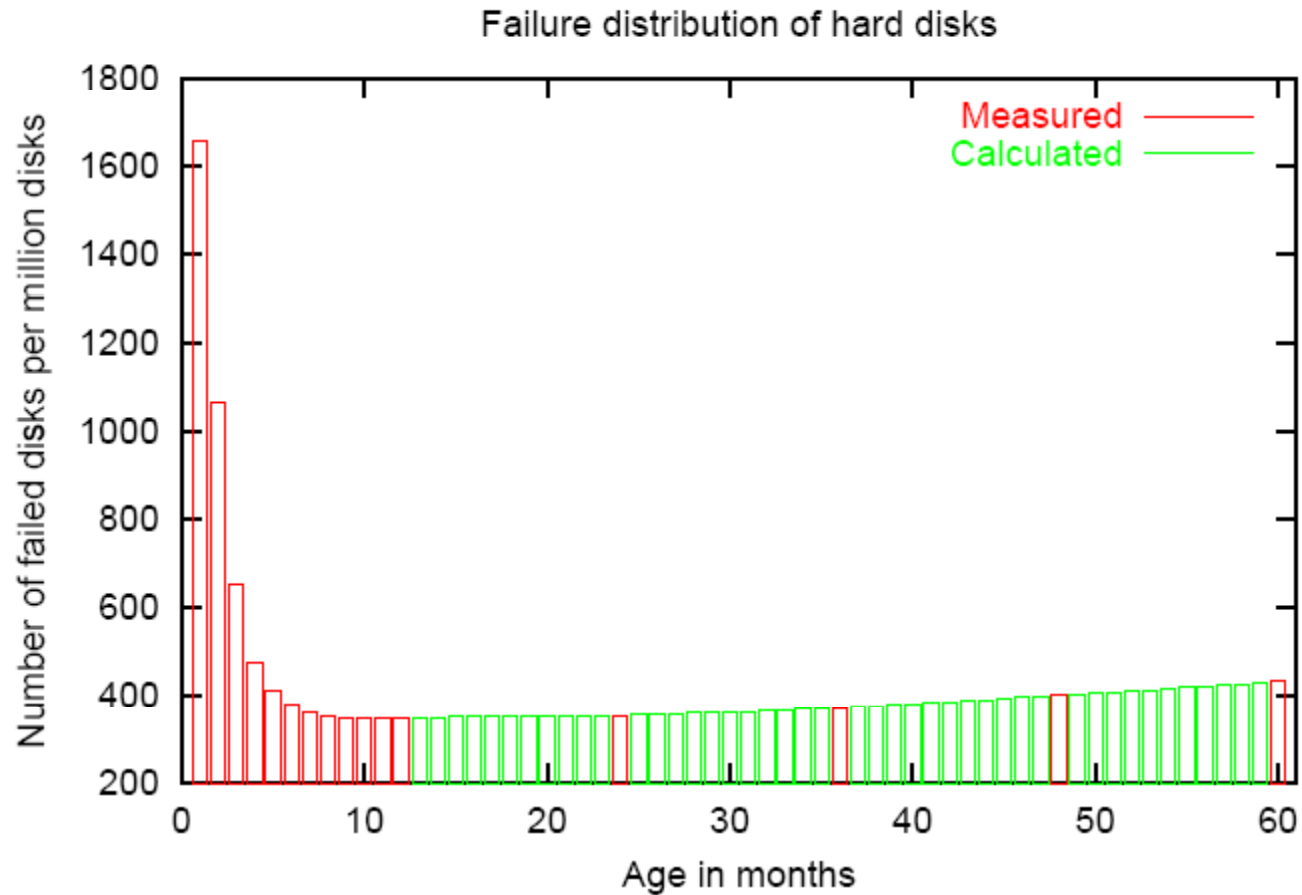


Figure 3: Disk failure distribution

Datenträger	min. Haltbarkeit	Max. Haltbarkeit	Herstellerangabe
XT-Diskette (5 1/4 Zoll, 360 kB)	10 Jahre	??	
AT-Diskette (5 1/4 Zoll, 1,2 MB)	5 Jahre	<10 Jahre	
Moderne Diskette (3 1/2 Zoll, 1,4 MB)	5 Jahre	??	
Magnetband (QIC80, 120 MB/ Travan 1, 400 MB)	1 Jahr	3 Jahre (mit viel Glück!)	
Festplatte	10 Jahre	??	
CD-ROM (gepreßt)	7 Jahre	??	70 Jahre Produkt (ohne Daten!)
CD-ROM (gebrannt)	4 Jahre	??	
DVDs			70 Jahre Produkt (ohne Daten!)
DVD-RAM			Daten 30 Jahre im Cartridge

Threats

<u>Medium</u>	Erwartete Lebensdauer	Speicherdichte (kbit/kg)
Steintafeln und Steinmalereien	mehrere tausend Jahre	0,001...1
Nickelplatte	mehrere tausend Jahre	
<u>Bücher</u> und Handschriften aus säurefreiem <u>Papier</u> und mit säurefreier und nicht eisenhaltiger <u>Tinte</u>	mehrere hundert Jahre	3.000...30.000
Schwarzweißfilme aus <u>Polyethylenterephthalat</u> (PET)	bis zu 1000 Jahre	
<u>Bücher</u> und Handschriften aus säurehaltigem <u>Papier</u> (19. und frühen 20. Jahrhunderts)	70 - 100 Jahre	3.000...30.000
Filme auf <u>Zelluloid</u>	mindestens 50 bis 70 Jahre, > hundert Jahre	
Herkömmliche <u>Bücher</u>	100 - 200 Jahre	
Optische Speichermedien <u>CD-ROM</u> / <u>DVD</u>	<u>CD-R</u> 5 - 10 Jahre <u>DVD-RAM</u> 30 Jahre	CD-ROM: $4 \cdot 10^8$ DVD: $2 \dots 4 \cdot 10^9$
<u>Zeitungspapier</u>	10 - 50 Jahre	
<u>Disketten</u>	5 - 10 Jahre	
<u>Magnetbänder</u>	bis zu 30 Jahre	
<u>REV-Festplatte</u>	bis zu 30 Jahre	
USB-Stick	3 - 10 Jahre	

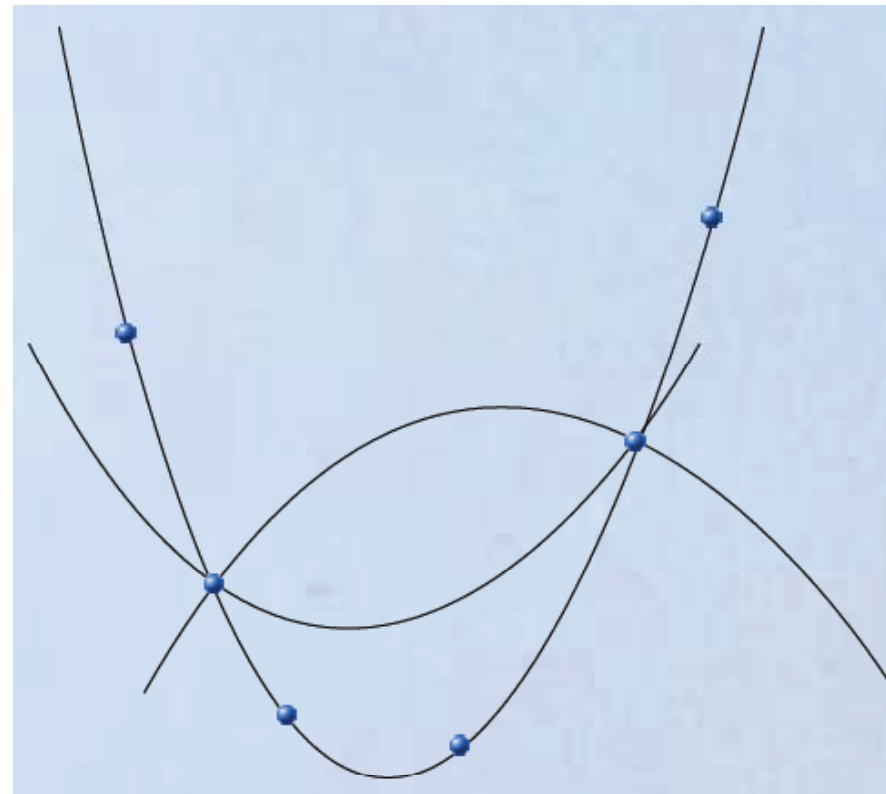
OceanStore

- Replicas
- Distribute data into $k*n$ redundant fragments
 - Random allocation
 - Actively enforce redundancy



Tangler

- Entanglement of data
 - Partition a secret into k keys (Shamir 1979)
 - n recover secret
 - $<n$ dispute access
 - Censorship-resistant



Best Practice

Methods

- Disk Imaging
- Data De-Duplication

Products

- Tape Technology
- RAID / JBOD

Services

- Maintenance



Bandlaufwerke

Vorteile

- hohe Kapazität
- lange Lebensdauer
- geringer Platzbedarf
- Medienkosten

Nachteile

- Umwelteinflüsse
- hohe Investition
- Serielles Schreiben
- Begrenzte Wiederverwendbarkeit



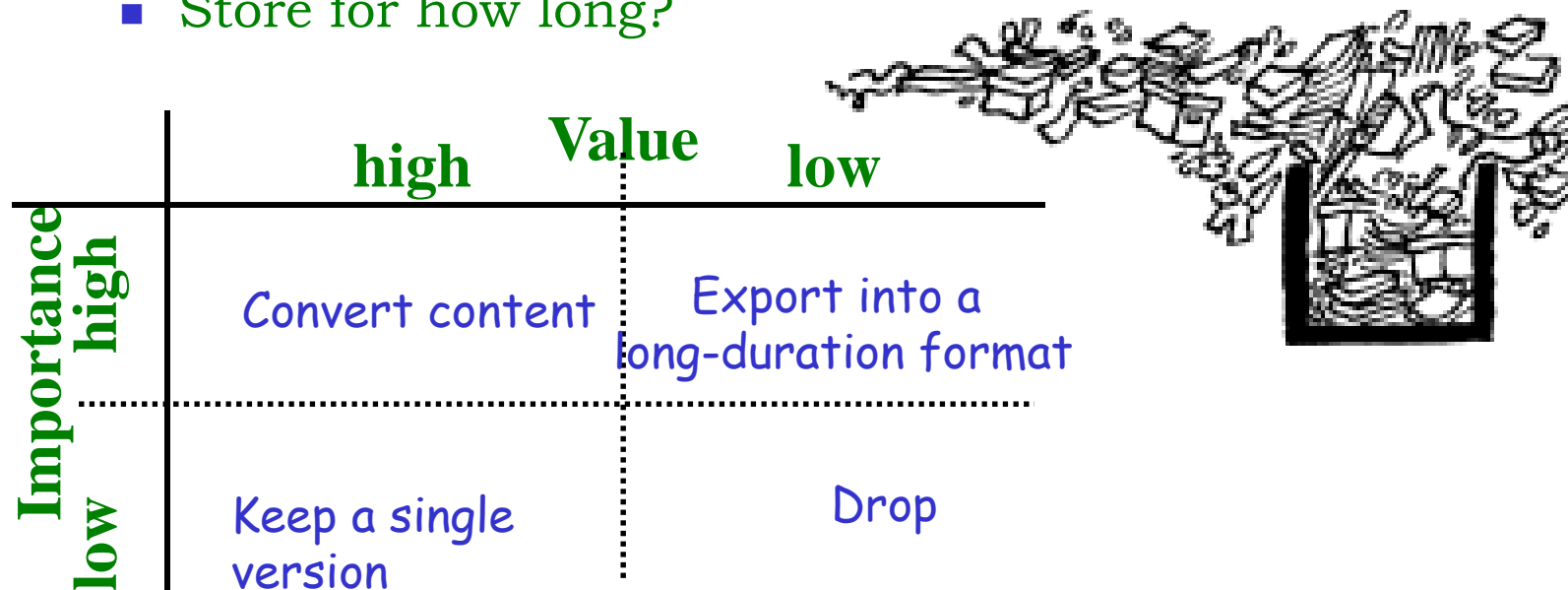
Strategies for Long-time Durability

- Standard File + Metadata Formats
- Migration
- Emulation
- Encapsulation
- Conversion at Runtime



Information Life Cycle Management

- Keep data for the *right* time on the *right* media
- Classify data by value and importance
 - Store for how long?



Home Use

Services

- RAID 1
- Amazon S3
- Public Data
 - Sourceforge
 - Wikipedia

Software

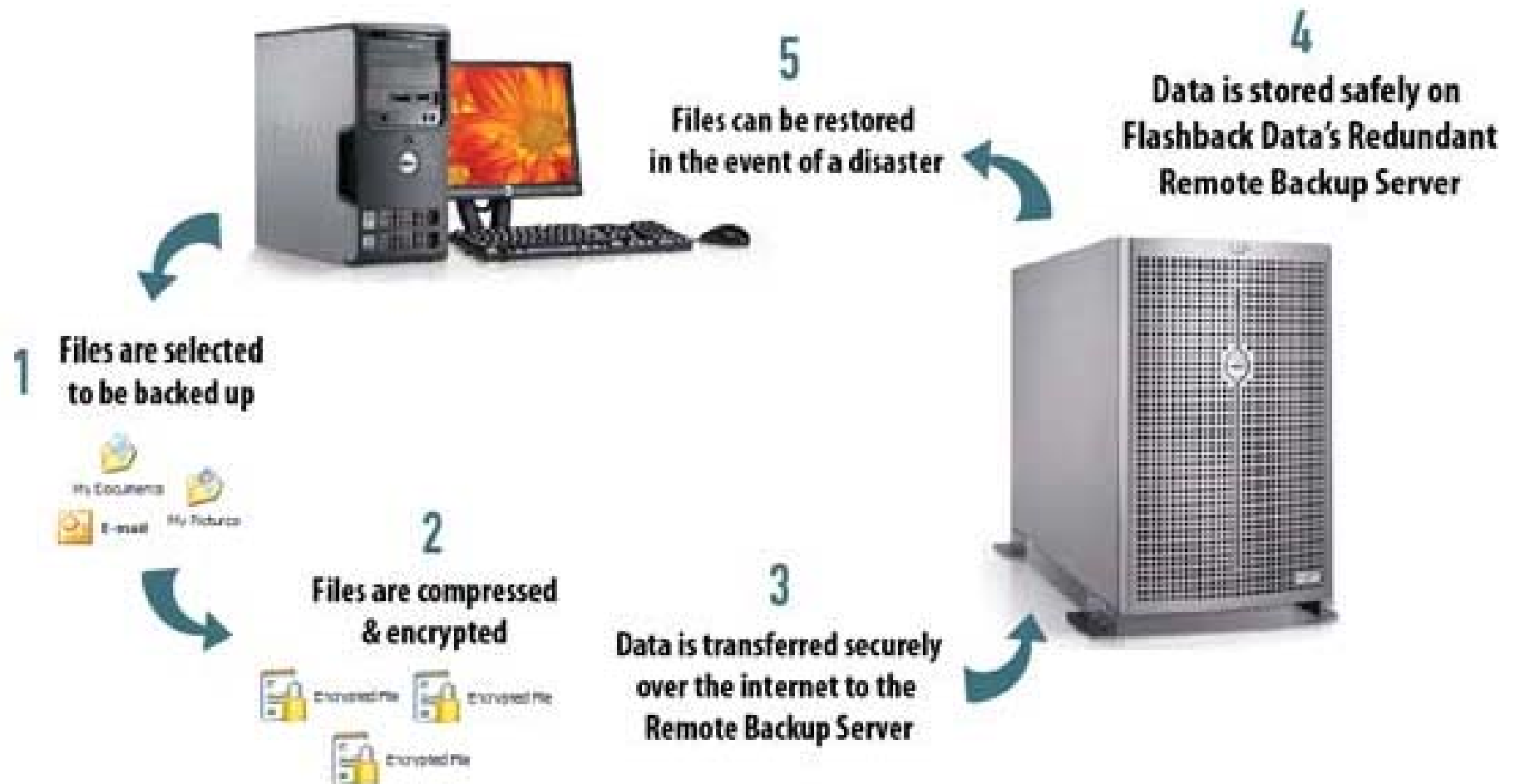
- Duplicity



WIKIPEDIA
Die freie Enzyklopädie



Online Backup



When things go wrong...

1. Have a plan
2. Stick to the plan

**Our Disaster Recovery Plan
Goes Something Like This...**



Challenges

- Data Lookup
- Consistency
- Test restore frequently



Conclusions



- „*Electronic Archives are the Memory of the Information Society.*“
Erkki Liikanen, EU-Kommissar für die Informationsgesellschaft, 2003
- 1. Setup online backups for networked sites
- 2. Test frequently



Backup!
Day!

Email: jens.oberender@joooo.de

References

- Articles
 - Datendeduplikation
http://www.computerzeitung.de/article.html?thes=&art=/articles/2007002/30920017_ha_CZ.html
 - Verfallsdatum unklar (Deutschlandfunk)
<http://www.dradio.de/dlf/sendungen/forschak/453943/>
 - **Best Practices for Digital Archiving**
<http://www.dlib.org/dlib/january00/01hodge.html>
 - **Marc Farley: Does a "best practice" exist for backup strategies?**
http://searchstorage.techtarget.com/ateQuestionNResponse/0,289625,sid5_cid423460_tax286191,00.html
 - **Co-operative Development of a Long-Term Digital Information Archive**
<http://www.kopal.langzeitarchivierung.de/>
- Products
 - Rsnapshot
http://www.mikerubel.org/computers/rsync_snapshots/
 - Backup Over Network To Multiple Incremental Archives
<http://folk.uio.no/johnen/bontmia/>
 - Amazon S3
<http://aws.amazon.com/s3>
 - Duplicity
<http://www.nongnu.org/duplicity/>
 - Google Desktop
<http://desktop.google.de>
 - Duplicate Finder
<http://www.easyduplicatefinder.com/>