In Case of Failure

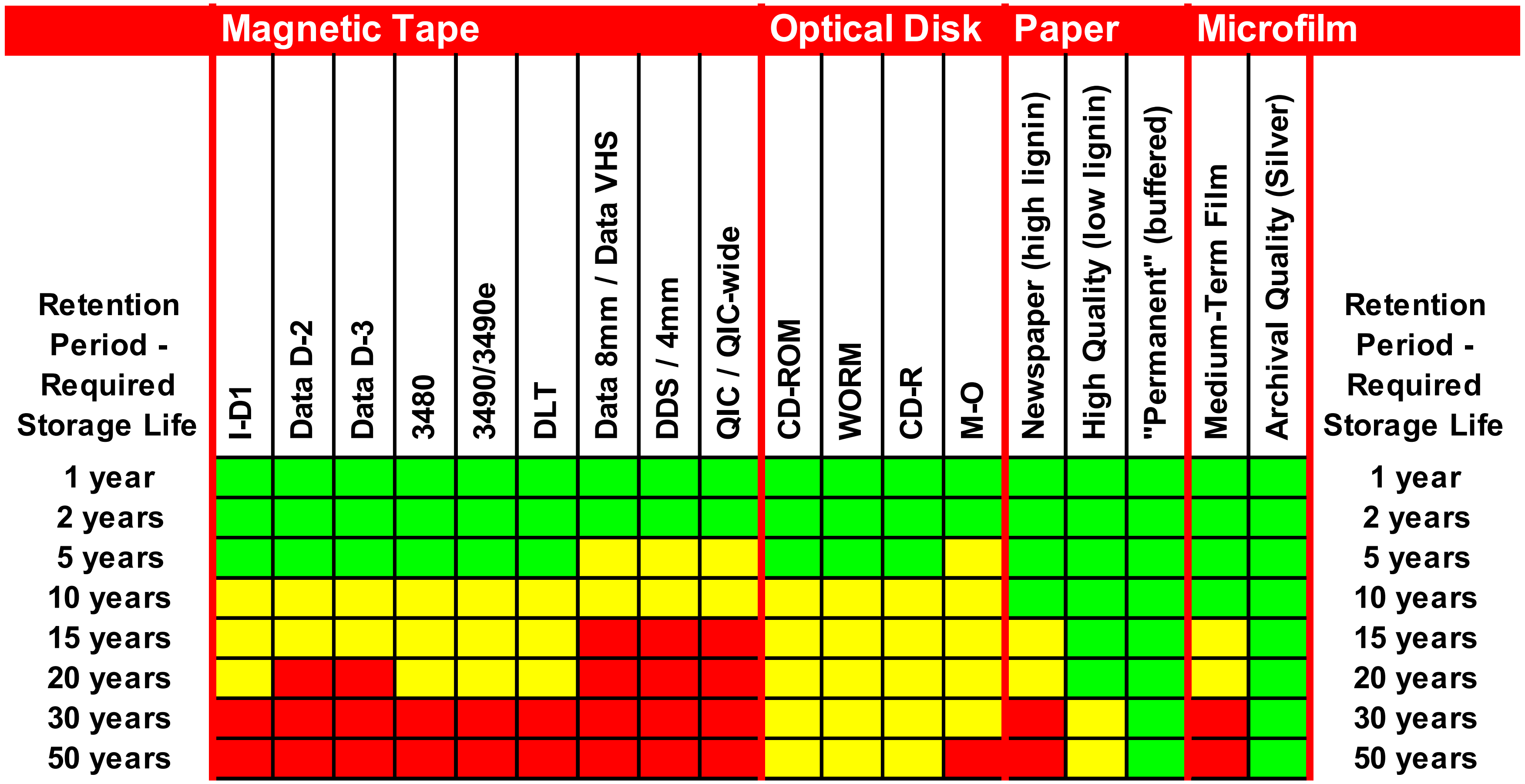
ELAG 2011 Prague

Patrick Hochstenbach * Ghent University
Email: Patrick.Hochstenbach@UGent.be
Twitter: @hochstenbach

BOM-VL/Archipel

http://www.slideshare.net/hochstenbach/20081007-workshop-bomvl-wp3

Life expectancies of media

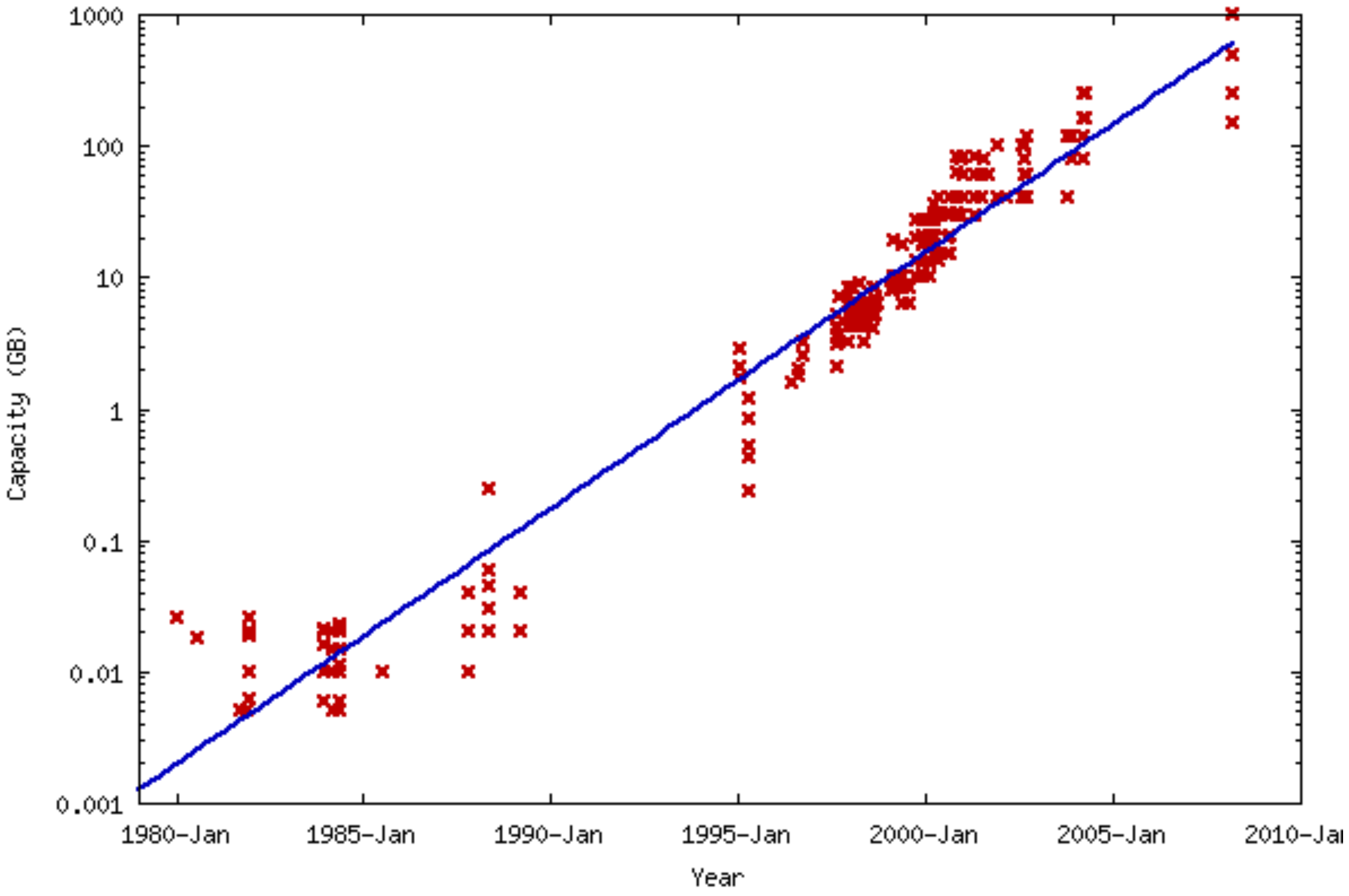


	Magnetic Tape									Optical Disk				Paper			Microfilm		
Retention Period - Required Storage Life	I-D1	Data D-2	Data D-3	3480	3490/3490e	DLT	Data 8mm / Data VHS	DDS / 4mm	QIC / QIC-wide	CD-ROM	WORM	CD-R	M-O	Newspaper (high lignin)	High Quality (low lignin)	"Permanent" (buffered)	Medium-Term Film	Archival Quality (Silver)	Retention Period - Required Storage Life
1 year	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	1 year
2 years	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	2 years
5 years	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟨	🟨	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	5 years
10 years	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟩	🟩	🟩	🟩	🟩	10 years
15 years	🟨	🟨	🟨	🟨	🟨	🟨	🟥	🟥	🟥	🟨	🟨	🟨	🟨	🟨	🟩	🟩	🟨	🟩	15 years
20 years	🟨	🟥	🟥	🟨	🟨	🟨	🟥	🟥	🟥	🟨	🟨	🟨	🟨	🟨	🟩	🟩	🟨	🟩	20 years
30 years	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟨	🟨	🟨	🟨	🟥	🟨	🟩	🟥	🟩	30 years
50 years	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟨	🟨	🟨	🟥	🟥	🟨	🟩	🟥	🟩	50 years

“Storage Media Life Expectancies” - Van Bogart, 1998

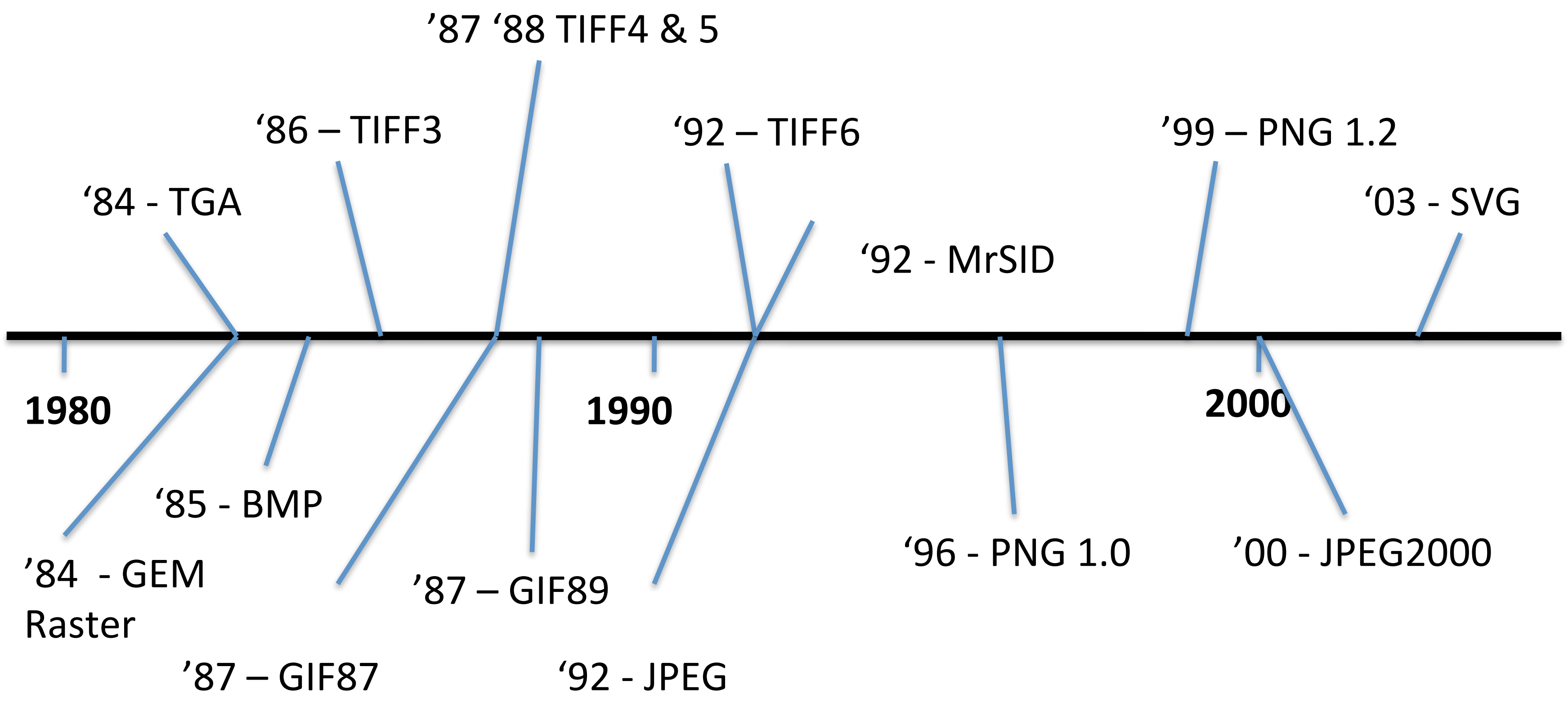
Growth of digital data

Capacity of desktop computers



http://commons.wikimedia.org/wiki/File:Hard_drive_capacity_over_time.png HanKwang (2008)

Growth in formats



’87 ‘88 TIFF4 & 5

‘86 – TIFF3

‘92 – TIFF6

’99 – PNG 1.2

‘84 - TGA

‘03 - SVG

‘92 - MrSID

1980

1990

2000

‘85 - BMP

‘96 - PNG 1.0

’00 - JPEG2000

’84 - GEM
Raster

’87 – GIF89

’87 – GIF87

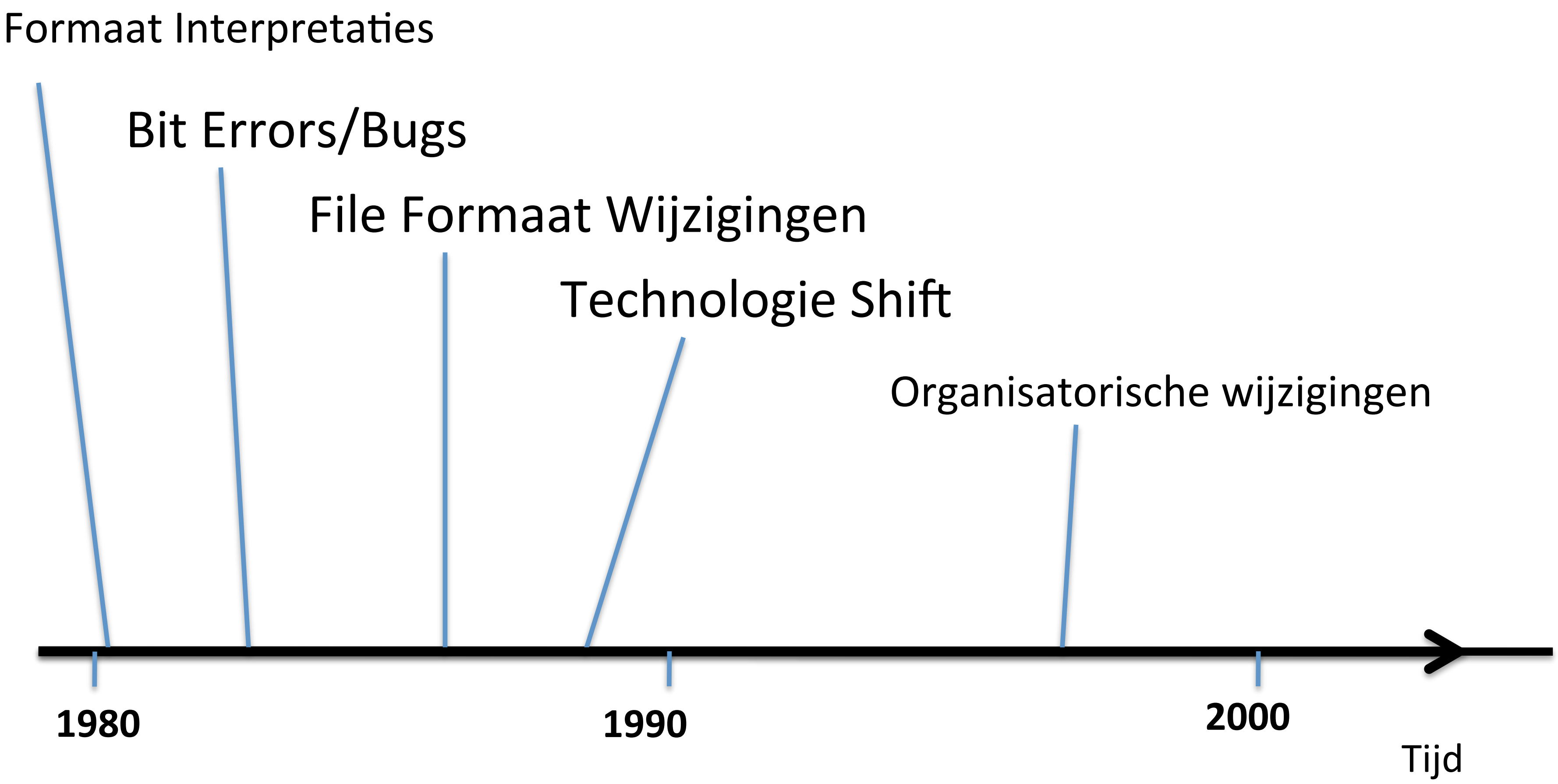
‘92 - JPEG

Formats of formats

MIME type image/tiff:
- TIFF (alle versies)
- TIFF/IT
- TIFF G4/LZW/UNC
- Digital Negative Format (DNG)
- GeoTIFF
- Pyramid TIFF
- …

Bron: PRONOM Technical Registry [http://www.nationalarchives.gov.uk/pronom/]

Short & long term risks



Best practices

Best practices

1. Create a preservation plan

Best practices

1. Create a preservation plan
2. Backup and replicate your data

Best practices

1. Create a preservation plan
2. Backup and replicate your data
3. Store preservation metadata

Best practices

1. Create a preservation plan
2. Backup and replicate your data
3. Store preservation metadata
4. Store technical metadata

Best practices

1. Create a preservation plan
2. Backup and replicate your data
3. Store preservation metadata
4. Store technical metadata
5. Store representation metadata

Best practices

1. Create a preservation plan
2. Backup and replicate your data
3. Store preservation metadata
4. Store technical metadata
5. Store representation metadata
6. Don’t trust software

Best practices

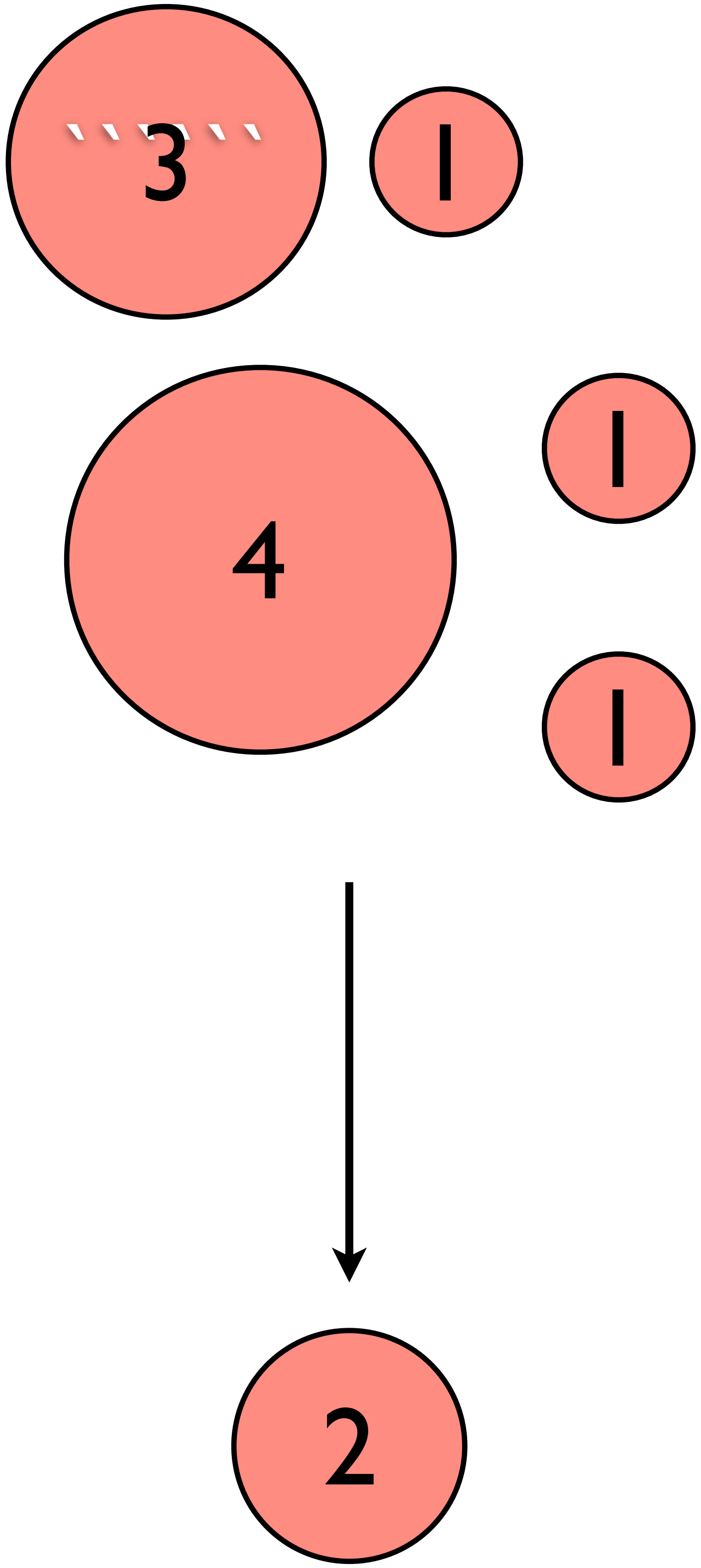
1. Create a preservation plan
2. Backup and replicate your data
3. Store preservation metadata
4. Store technical metadata
5. Store representation metadata
6. Don’t trust software
7. Store descriptive metadata

Preservation Plan

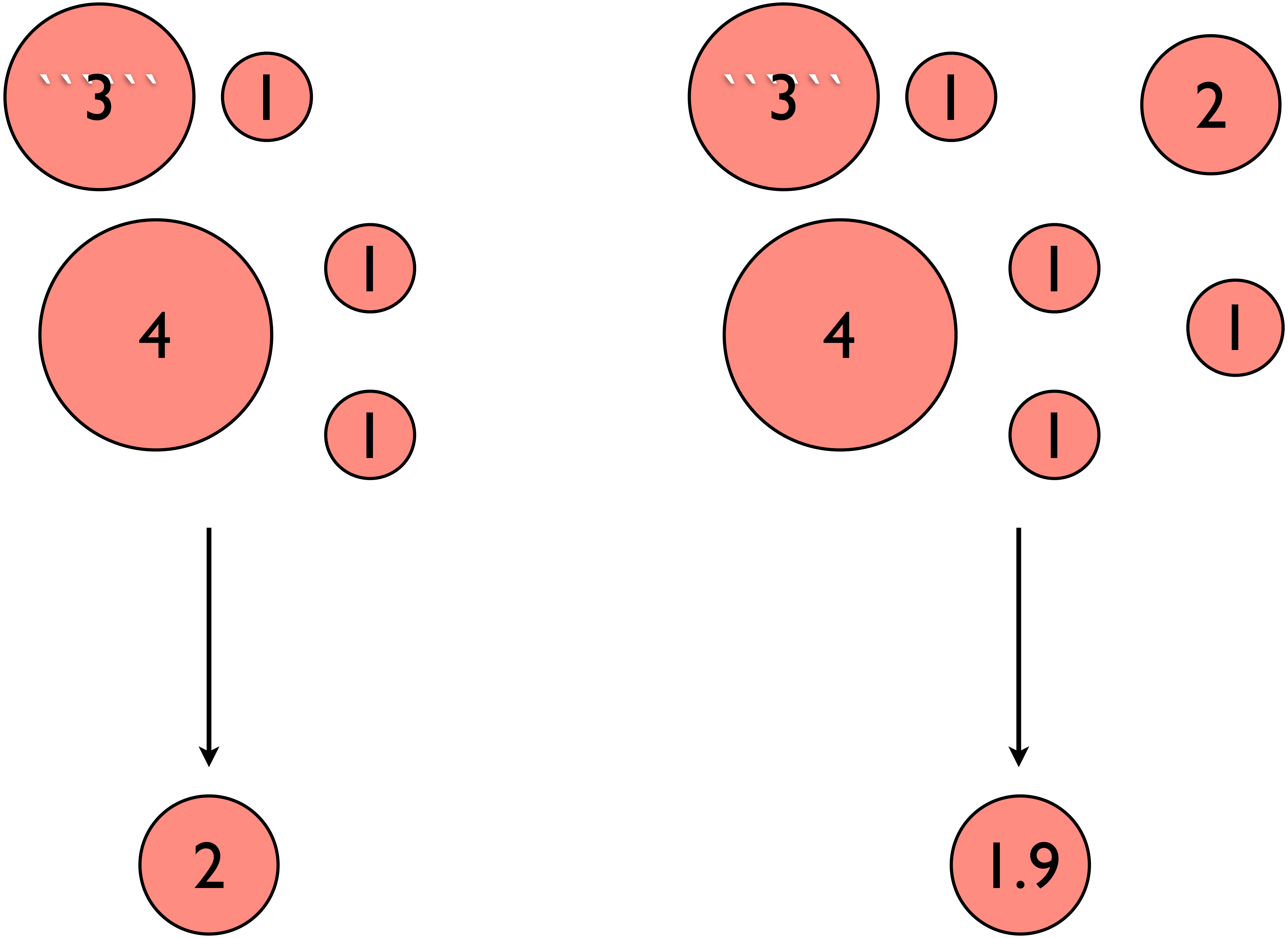
- Preservation policies (what to preserve)
- Legal obligations
- Organizational & Technical constraints
- User requirements
- Context
- http://plato.ifs.tuwien.ac.at:8080/plato

Risk Analysis

Random error

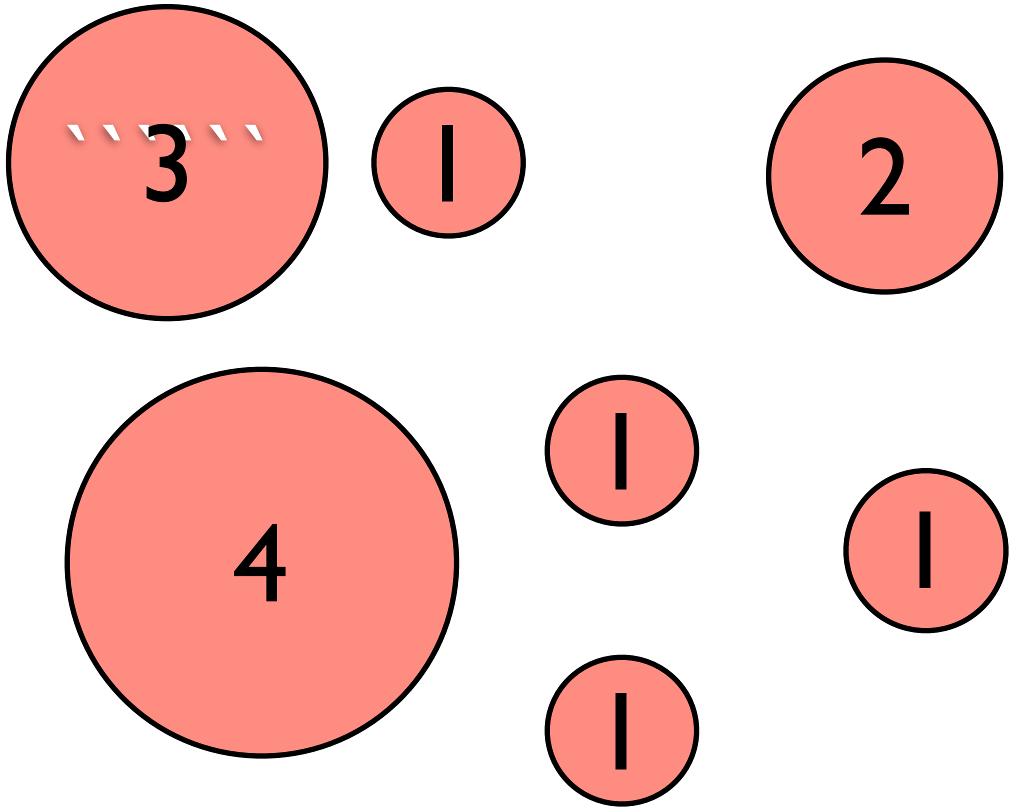


Random error

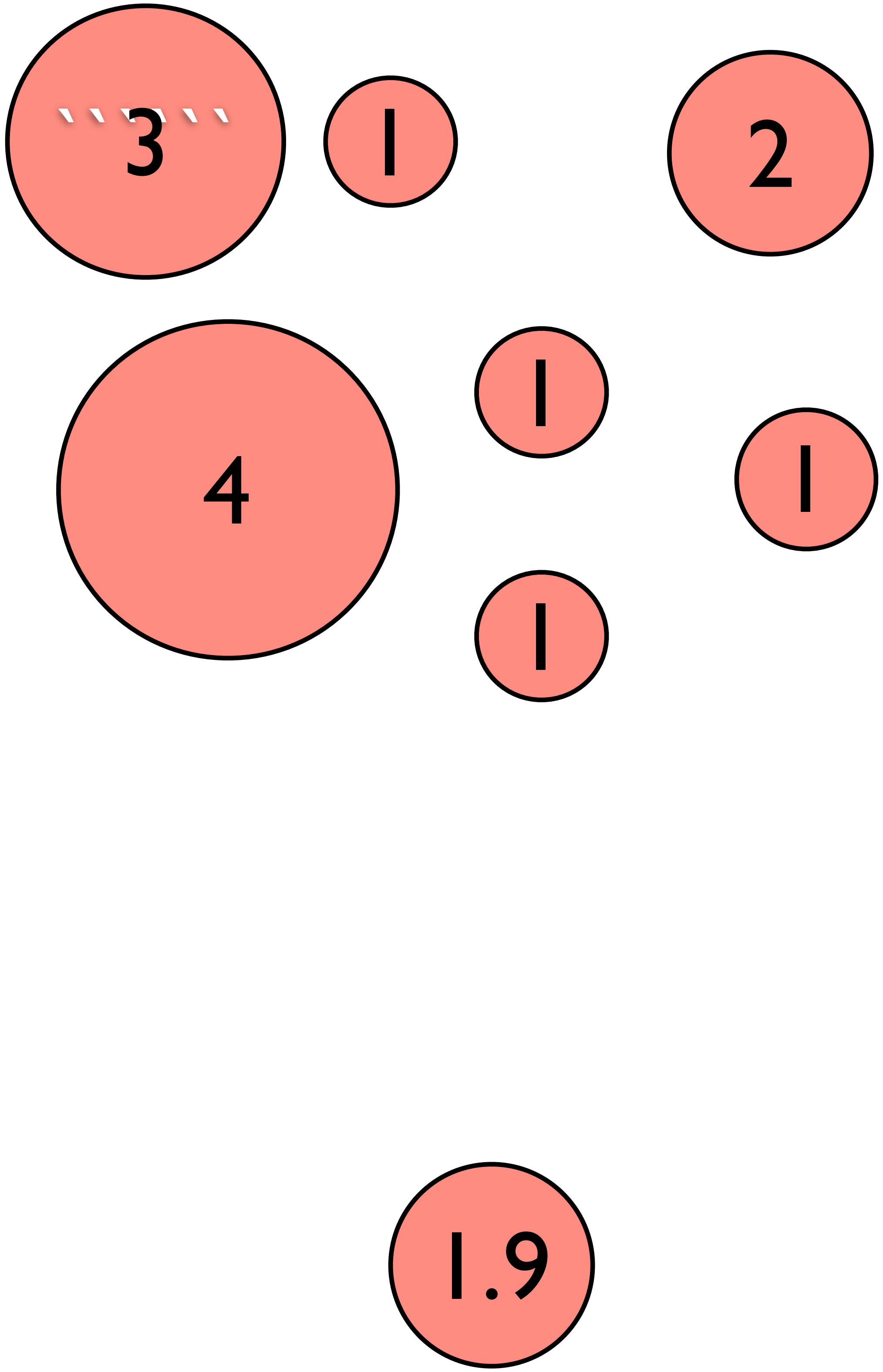


Random error

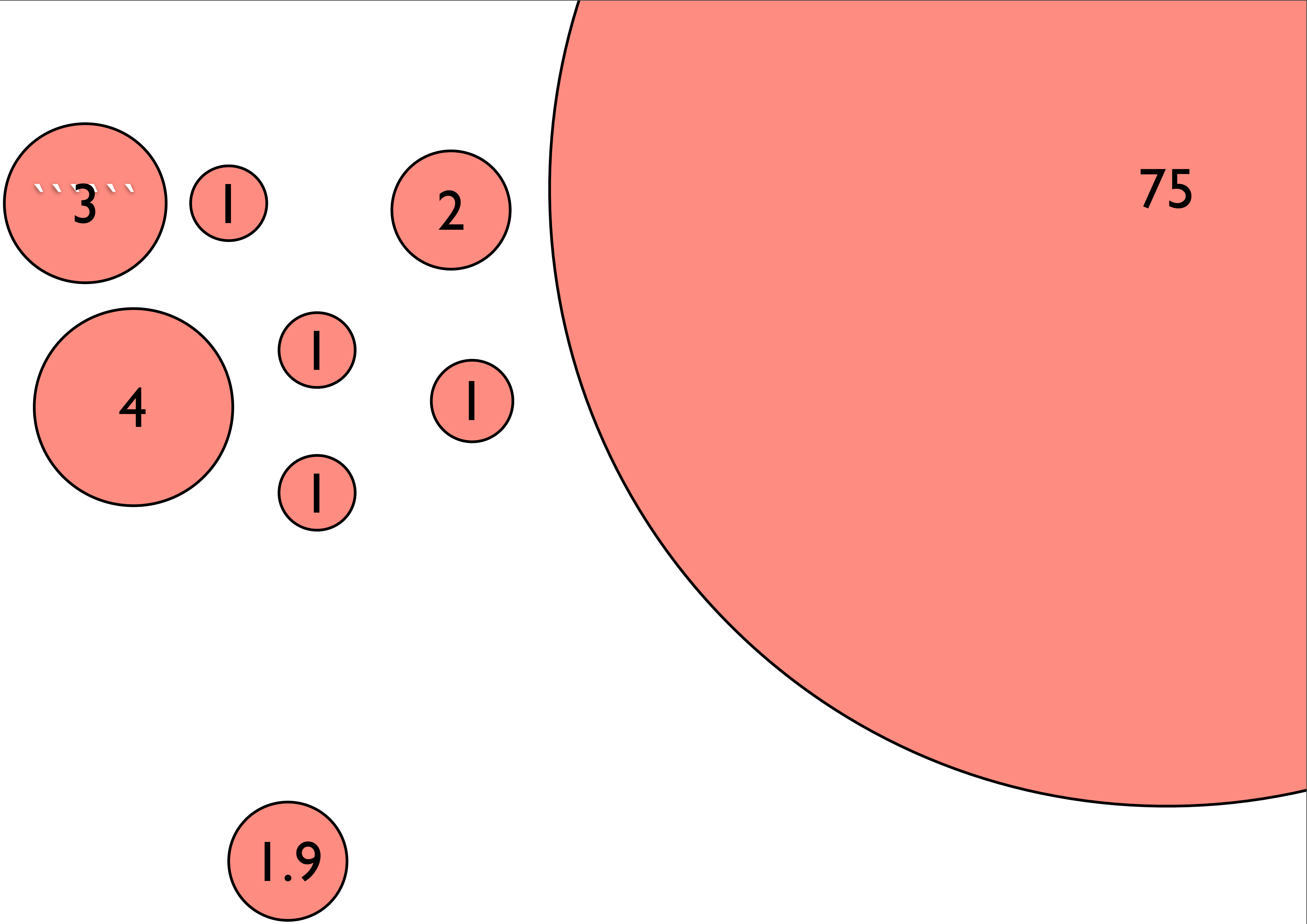
Systematic error



Systematic error



Systematic error



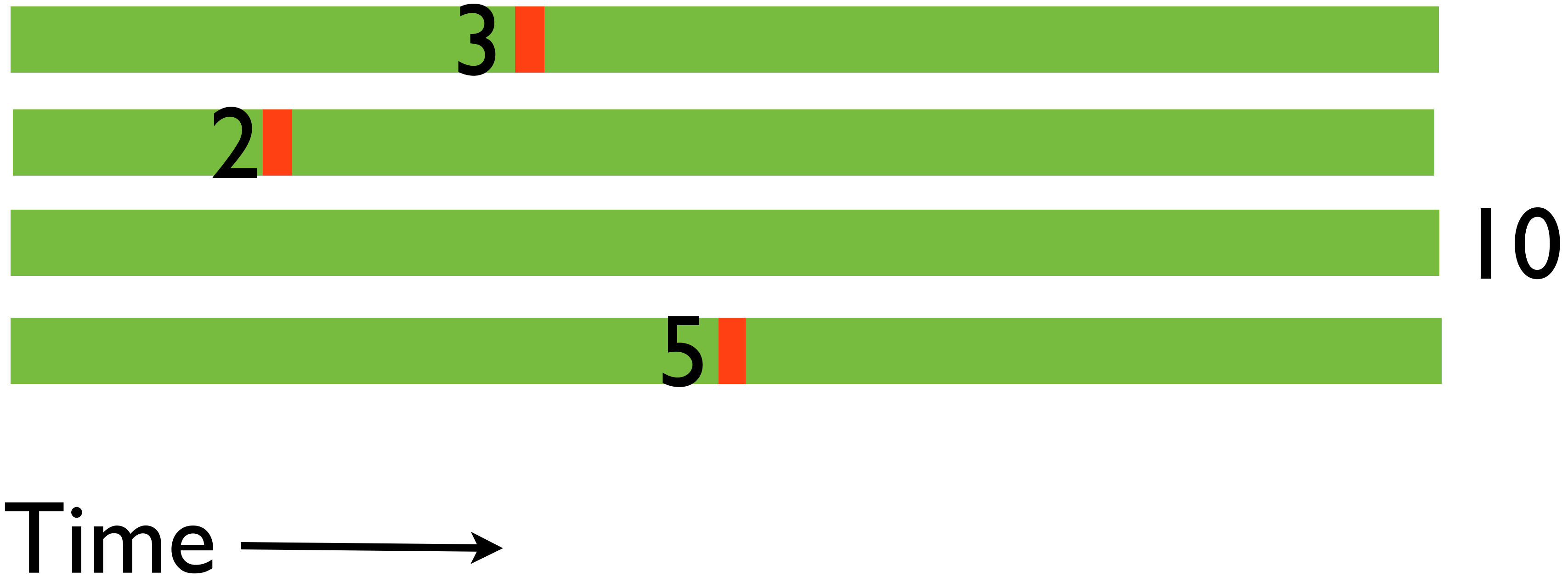
Systematic error

BY DAVID S.H. ROSENTHAL

Keeping Bits Safe: How Hard Can It Be?

MTBF

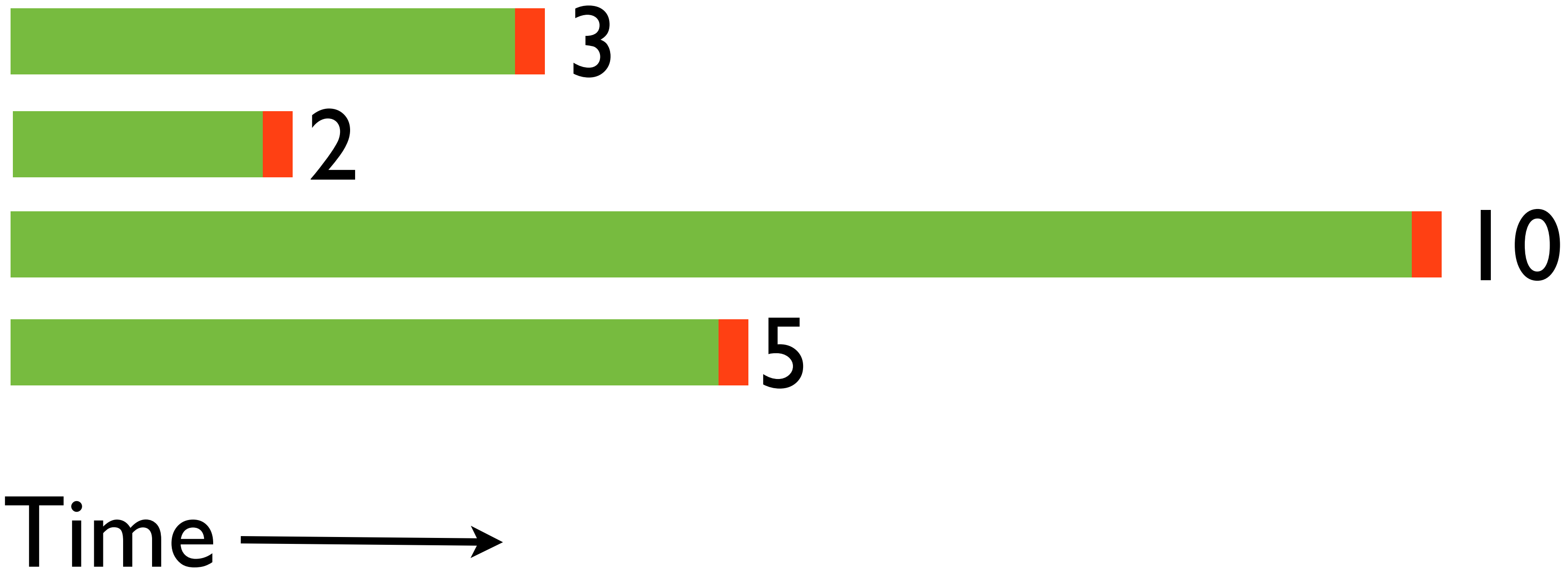
MTBF = Mean Time Between Failure



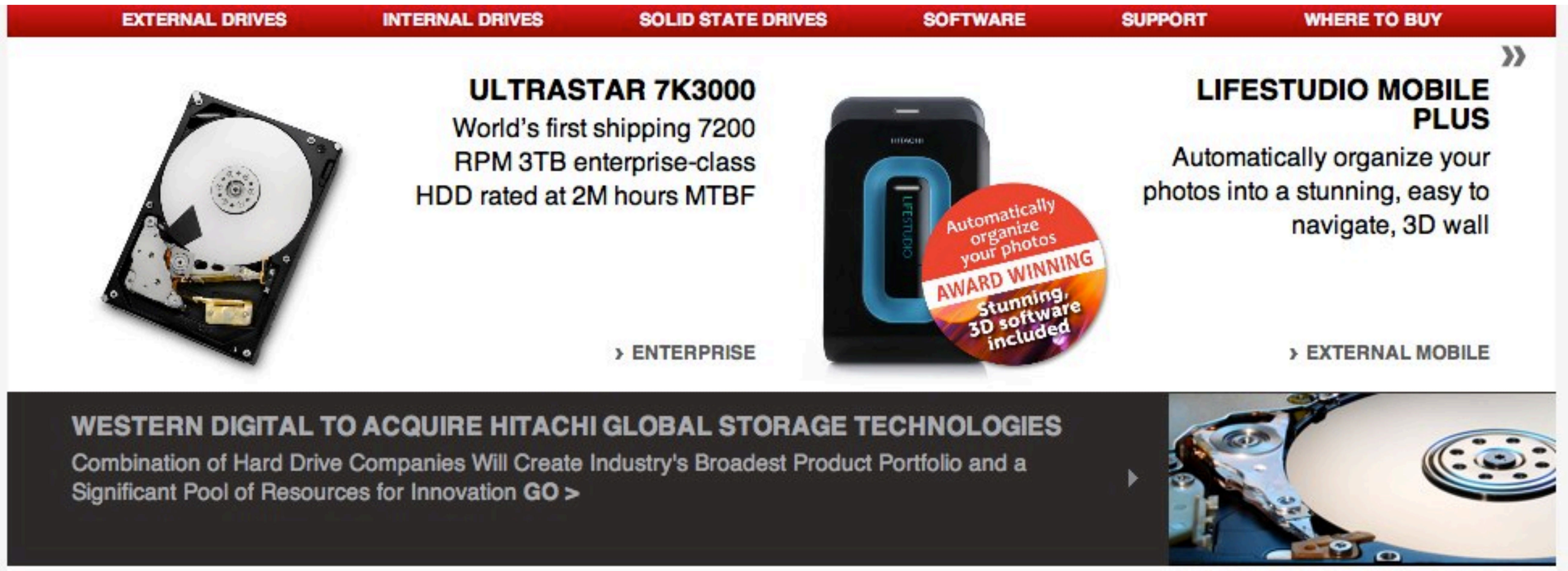
$$\text{MTBF} = \frac{\text{Total Time}}{\text{Number of failures}} = \frac{\text{40 hours}}{\text{3 failures}} = \text{13.3 hrs}$$

MTTF

MTTF = Mean Time To Failure



$$\text{MTTF} = \frac{\text{Total time}}{\text{Number of units}} = \frac{\text{20 hours}}{\text{4 units}} = \text{5 hrs}$$



ULTRASTAR 7K3000

World's first shipping 7200 RPM 3TB enterprise-class HDD rated at 2M hours MTBF

› ENTERPRISE

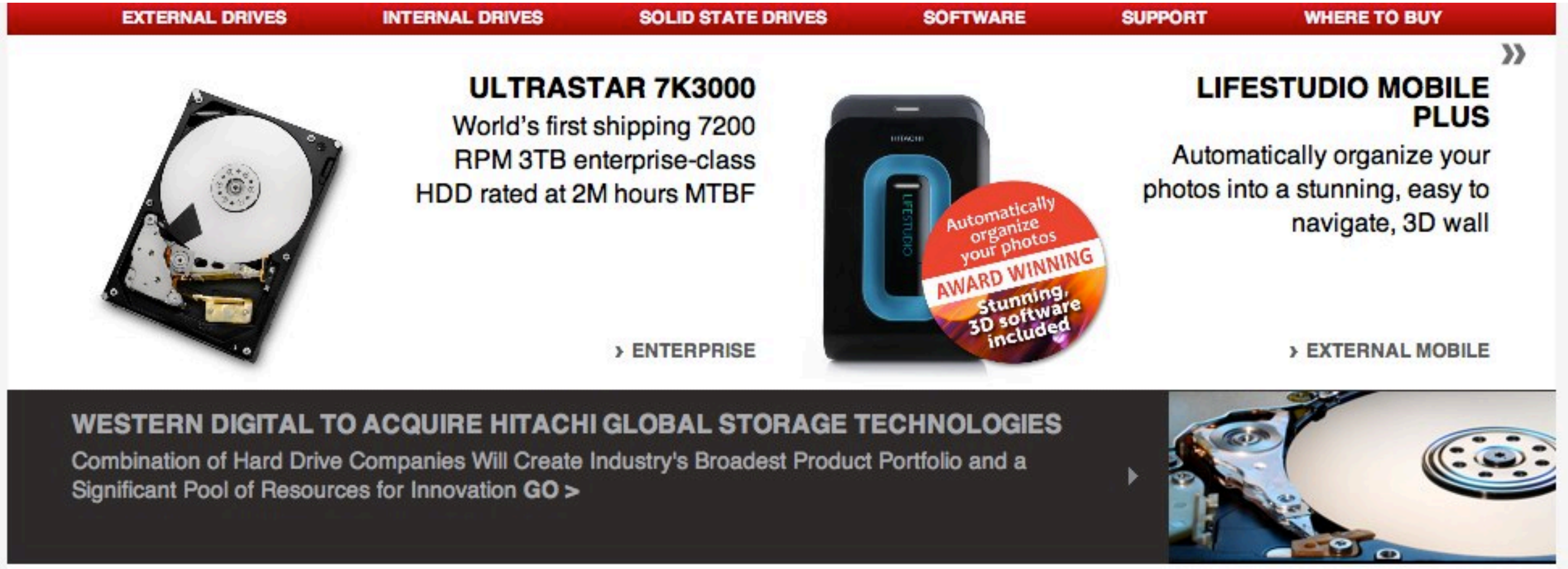
LIFESTUDIO MOBILE PLUS

Automatically organize your photos into a stunning, easy to navigate, 3D wall

› EXTERNAL MOBILE

WESTERN DIGITAL TO ACQUIRE HITACHI GLOBAL STORAGE TECHNOLOGIES

Combination of Hard Drive Companies Will Create Industry's Broadest Product Portfolio and a Significant Pool of Resources for Innovation **GO >**



ULTRASTAR 7K3000

World's first shipping 7200 RPM 3TB enterprise-class HDD rated at 2M hours MTBF

› ENTERPRISE

LIFESTUDIO MOBILE PLUS

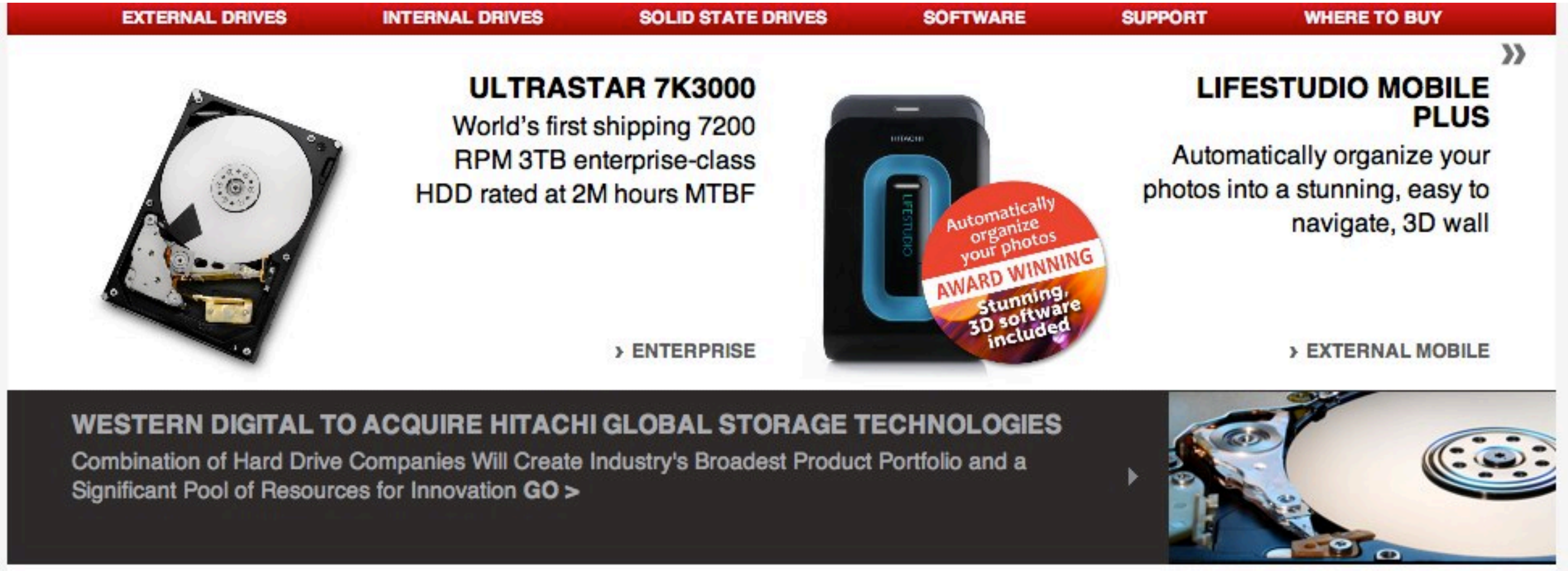
Automatically organize your photos into a stunning, easy to navigate, 3D wall

› EXTERNAL MOBILE

WESTERN DIGITAL TO ACQUIRE HITACHI GLOBAL STORAGE TECHNOLOGIES

Combination of Hard Drive Companies Will Create Industry's Broadest Product Portfolio and a Significant Pool of Resources for Innovation **GO >**

MTTF = 2 M hours = 228 years!



ULTRASTAR 7K3000

World's first shipping 7200 RPM 3TB enterprise-class HDD rated at 2M hours MTBF

› ENTERPRISE

LIFESTUDIO MOBILE PLUS

Automatically organize your photos into a stunning, easy to navigate, 3D wall

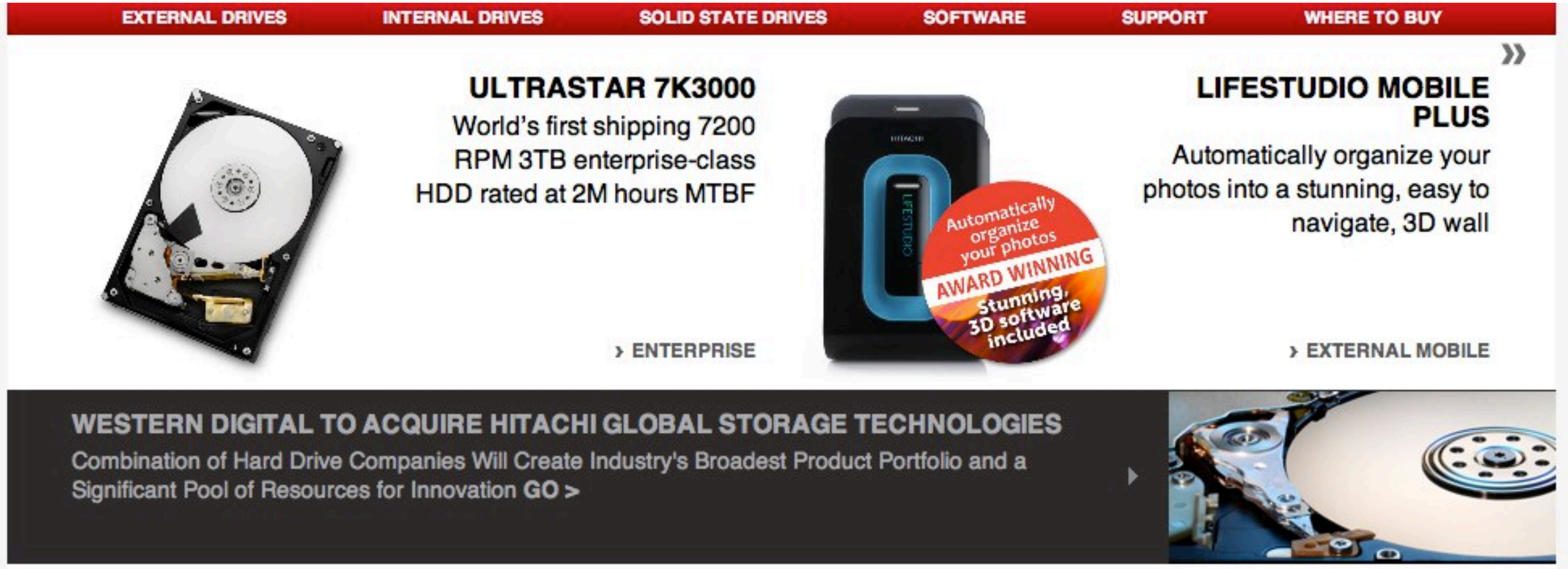
› EXTERNAL MOBILE

WESTERN DIGITAL TO ACQUIRE HITACHI GLOBAL STORAGE TECHNOLOGIES

Combination of Hard Drive Companies Will Create Industry's Broadest Product Portfolio and a Significant Pool of Resources for Innovation **GO >**

$$\text{MTTF} = 2\ \text{M hours} = 228\ \text{years!}$$

$$\text{AFR} = 1/\text{MTTF} = 0.004 = 0.4\ \%$$



ULTRASTAR 7K3000

World's first shipping 7200 RPM 3TB enterprise-class HDD rated at 2M hours MTBF

› ENTERPRISE

LIFESTUDIO MOBILE PLUS

Automatically organize your photos into a stunning, easy to navigate, 3D wall

› EXTERNAL MOBILE

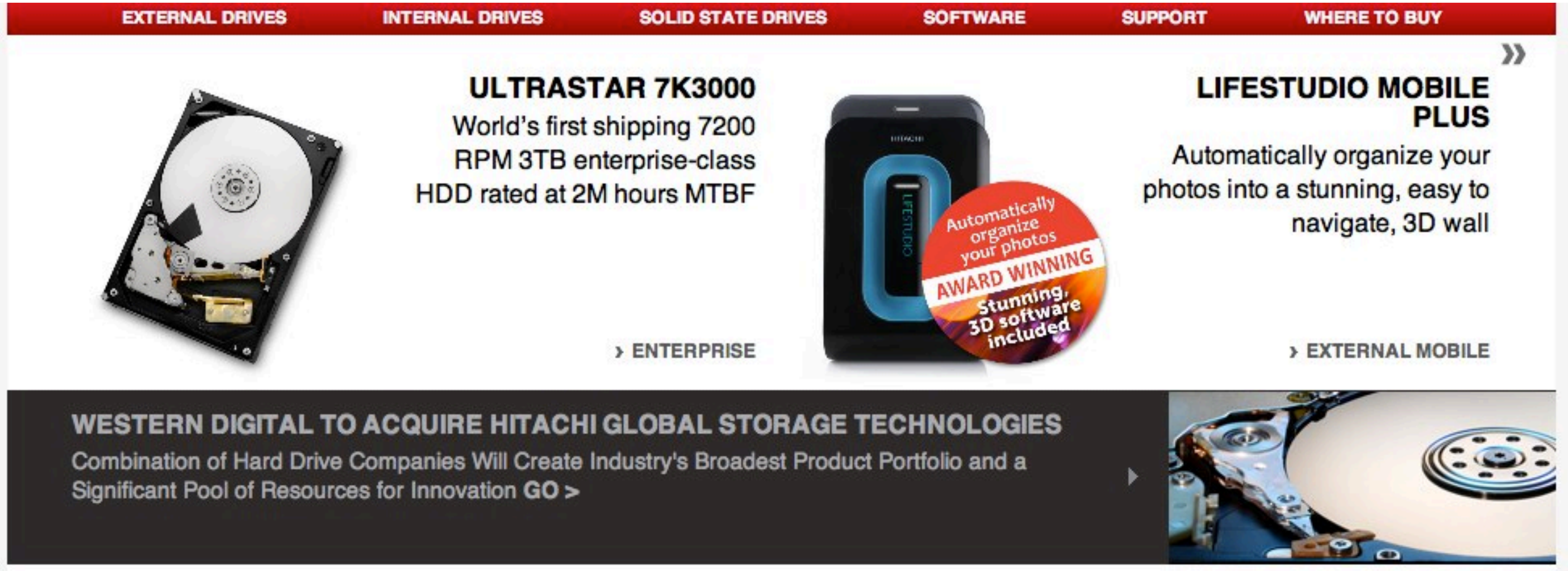
WESTERN DIGITAL TO ACQUIRE HITACHI GLOBAL STORAGE TECHNOLOGIES

Combination of Hard Drive Companies Will Create Industry's Broadest Product Portfolio and a Significant Pool of Resources for Innovation **GO >**

MTTF = 2 M hours = 228 years!

AFR = 1/MTTF = 0.004 = 0.4 %

$$R(t) = \exp(-t/\theta)$$



ULTRASTAR 7K3000

World's first shipping 7200 RPM 3TB enterprise-class HDD rated at 2M hours MTBF

LIFESTUDIO MOBILE PLUS

Automatically organize your photos into a stunning, easy to navigate, 3D wall

WESTERN DIGITAL TO ACQUIRE HITACHI GLOBAL STORAGE TECHNOLOGIES

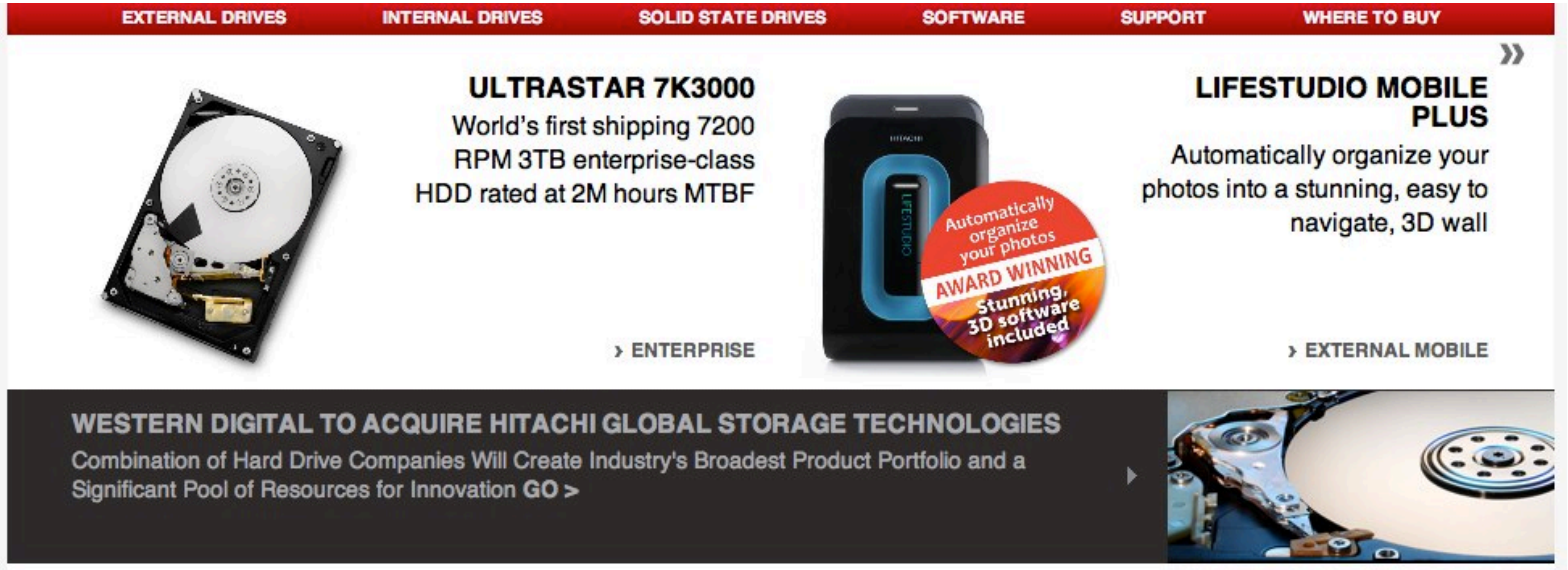
Combination of Hard Drive Companies Will Create Industry's Broadest Product Portfolio and a Significant Pool of Resources for Innovation **GO >**

$$\text{MTTF} = 2\ \text{M hours} = 228\ \text{years!}$$

$$\text{AFR} = 1/\text{MTTF} = 0.004 = 0.4\ \%$$

$$R(t) = \exp(-t/\Theta)$$

$$R(5) = \exp(-5/228) = 0.98 = 98\%$$



ULTRASTAR 7K3000

World's first shipping 7200 RPM 3TB enterprise-class HDD rated at 2M hours MTBF

LIFESTUDIO MOBILE PLUS

Automatically organize your photos into a stunning, easy to navigate, 3D wall

WESTERN DIGITAL TO ACQUIRE HITACHI GLOBAL STORAGE TECHNOLOGIES

Combination of Hard Drive Companies Will Create Industry's Broadest Product Portfolio and a Significant Pool of Resources for Innovation **GO >**

MTTF = 2 M hours = 228 years!

AFR = 1/MTTF = 0.004 = 0.4 %

$$R(t) = \exp(-t/\theta)$$

$$R(5) = \exp(-5/228) = 0.98 = 98\%$$

50 disks = 0.98^50 = 0.36 = 36%

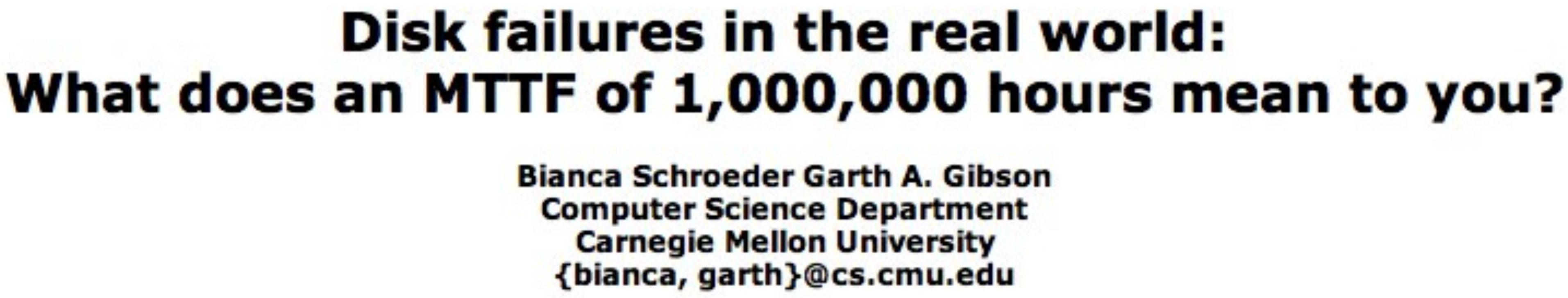
Experiments

- Simulate 100 disks with a 200 MTTF using Processing. What happens if the AFR is not 0.4% but 4% (hint: what is MTTF in that case)?
- Given a MTTF of 200 years and 50 disks what is the reliability in 1,2 and 5 years?

Experiments

- Amazon S3 claims an AFR per object of 0.000000001% [1]. What is the MTTF?
- There are 100 billion objects in S3. Given an estimated average size of 1 MB how big is S3?
- What is the chance (reliability) none of these 100 billion objects are lost in 1 year?

[1] http://aws.amazon.com/s3/faqs/#How_reliable_is_Amazon_S3#How_durable_is_Amazon_S3

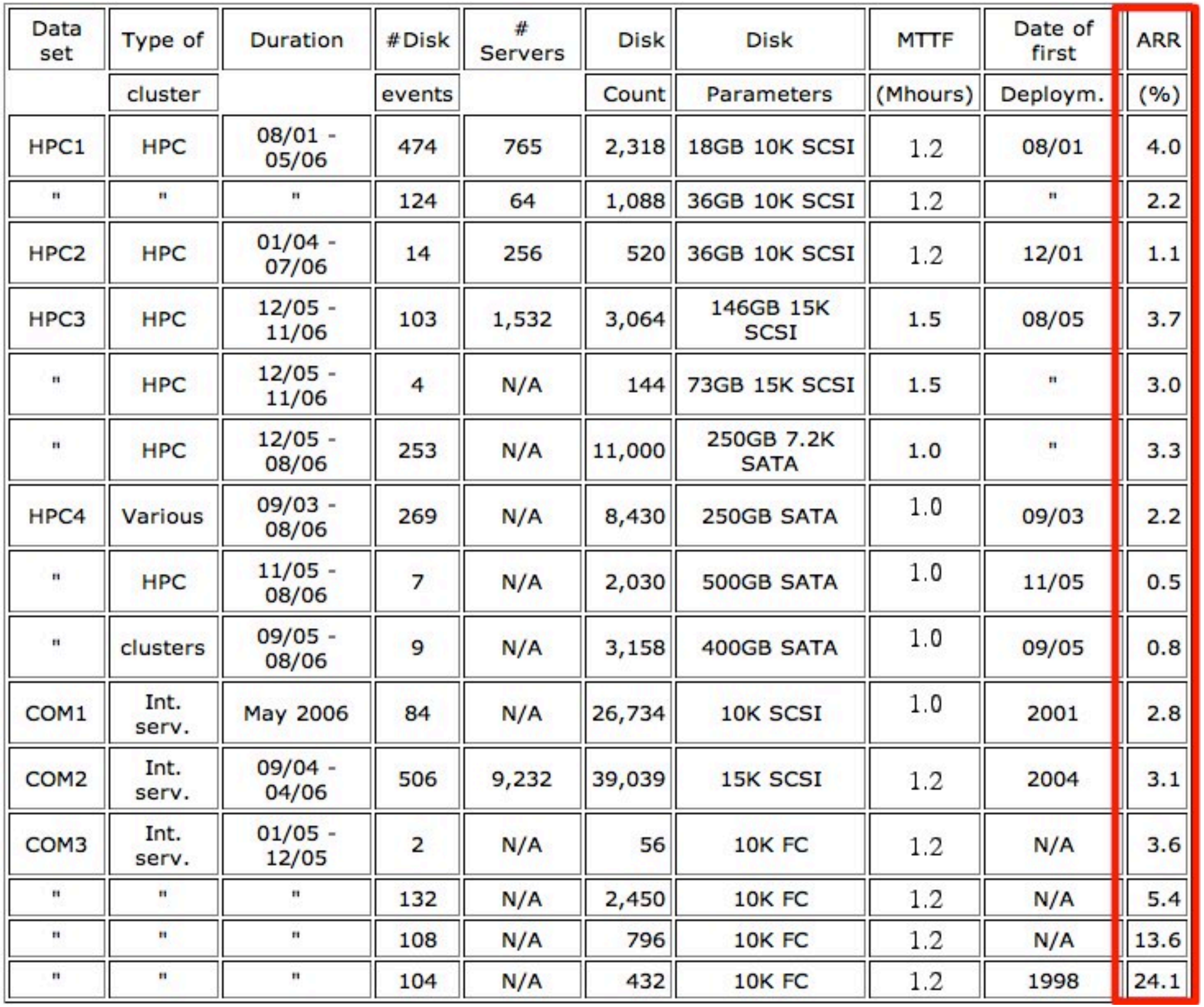


Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you?

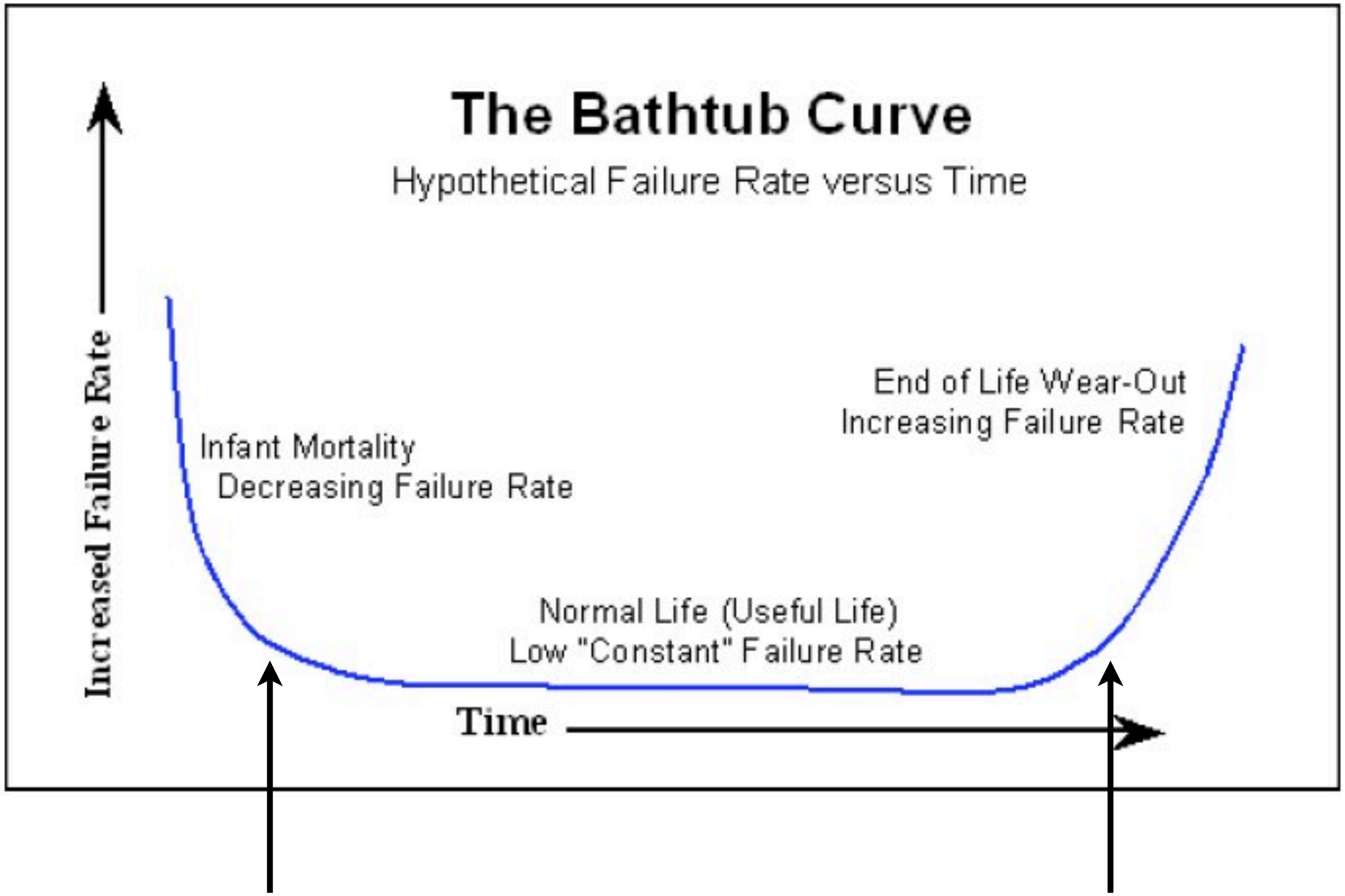
Bianca Schroeder Garth A. Gibson
Computer Science Department
Carnegie Mellon University
{bianca, garth}@cs.cmu.edu

http://db.usenix.org/events/fast07/tech/schroeder/schroeder_html/index.html

Shroeder & Gibson

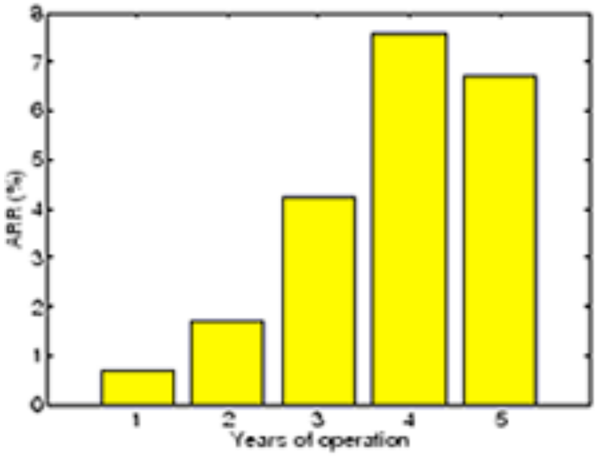


Data set	Type of	Duration	#Disk	# Servers	Disk	Disk	MTTF	Date of first	ARR
	cluster		events		Count	Parameters	(Mhours)	Deploym.	(%)
HPC1	HPC	08/01 - 05/06	474	765	2,318	18GB 10K SCSI	1.2	08/01	4.0
"	"	"	124	64	1,088	36GB 10K SCSI	1.2	"	2.2
HPC2	HPC	01/04 - 07/06	14	256	520	36GB 10K SCSI	1.2	12/01	1.1
HPC3	HPC	12/05 - 11/06	103	1,532	3,064	146GB 15K SCSI	1.5	08/05	3.7
"	HPC	12/05 - 11/06	4	N/A	144	73GB 15K SCSI	1.5	"	3.0
"	HPC	12/05 - 08/06	253	N/A	11,000	250GB 7.2K SATA	1.0	"	3.3
HPC4	Various	09/03 - 08/06	269	N/A	8,430	250GB SATA	1.0	09/03	2.2
"	HPC	11/05 - 08/06	7	N/A	2,030	500GB SATA	1.0	11/05	0.5
"	clusters	09/05 - 08/06	9	N/A	3,158	400GB SATA	1.0	09/05	0.8
COM1	Int. serv.	May 2006	84	N/A	26,734	10K SCSI	1.0	2001	2.8
COM2	Int. serv.	09/04 - 04/06	506	9,232	39,039	15K SCSI	1.2	2004	3.1
COM3	Int. serv.	01/05 - 12/05	2	N/A	56	10K FC	1.2	N/A	3.6
"	"	"	132	N/A	2,450	10K FC	1.2	N/A	5.4
"	"	"	108	N/A	796	10K FC	1.2	N/A	13.6
"	"	"	104	N/A	432	10K FC	1.2	1998	24.1

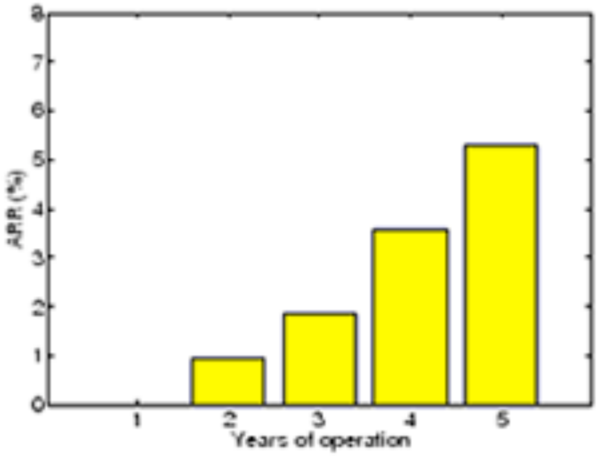

The Bathtub Curve
Hypothetical Failure Rate versus Time
Increased Failure Rate
Infant Mortality
Decreasing Failure Rate
Normal Life (Useful Life)
Low "Constant" Failure Rate
End of Life Wear-Out
Increasing Failure Rate
Time

1 yr

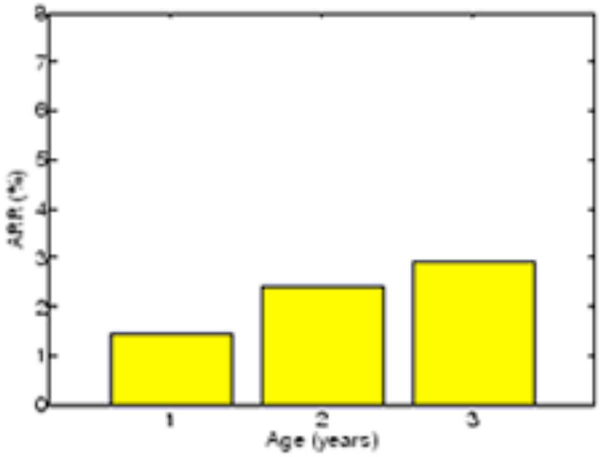
3-5yr



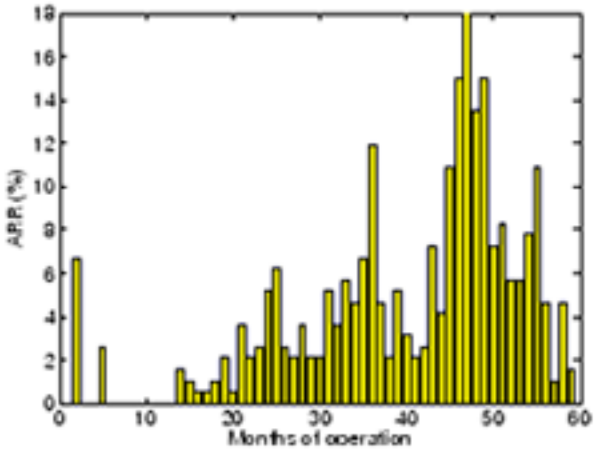
HPC1 (compute nodes)



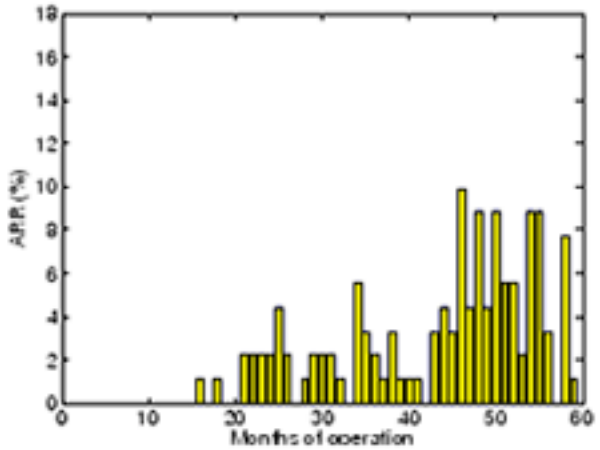
HPC1 (filesystem nodes)



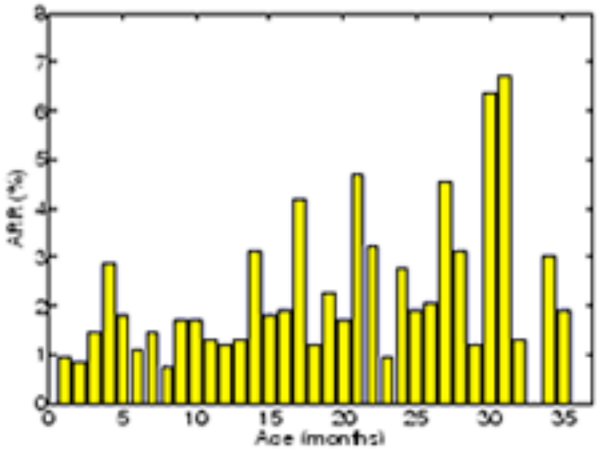
HPC4



HPC1 (compute nodes)



HPC1 (filesystem nodes)



HPC4

Experiments

- Given the lifetime of the universe (13 billion years) as the lifetime of one storage byte. What is the probability one Tera byte (1 billion bytes) will survive 100 years?
- Discuss

Why Do Computers Stop and What Can Be Done About It?

http://www.hpl.hp.com/techreports/tandem/TR-85.7.html

Jim Gray

Serial Failures

Serial Failures

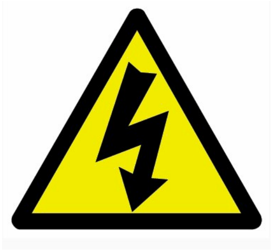
⚡ 87 years

Serial Failures

⚠ 87 years

75 years

Serial Failures

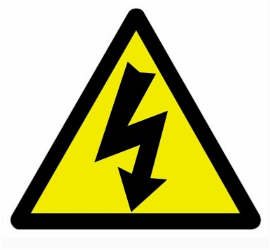


87 years

75 years

50 years

Serial Failures



87 years

75 years

50 years

[redacted]

31 years

Serial Failures

- A→B→C→D→.... →SYSTEM

$$\frac{1}{\text{SYSTEM}} = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D} +$$

E.g. : components : 1 , 100 , 1000, 10000
System: 0.989 years

Parallel Failures

= 200 years

= ?? years

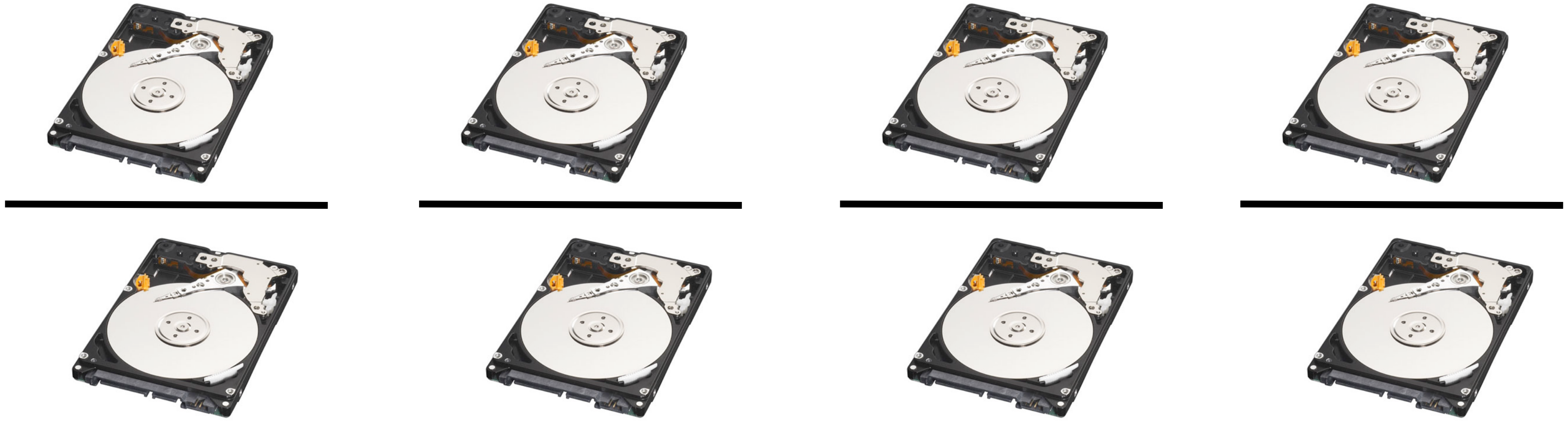
Parallel Failures

$$\text{SYSTEM} = \left\{ \begin{array}{l} A \\ B \\ C \\ D \end{array} \right.$$

SYSTEM = A * B * C * D

E.g. : components : 200,200
System: 40000 years

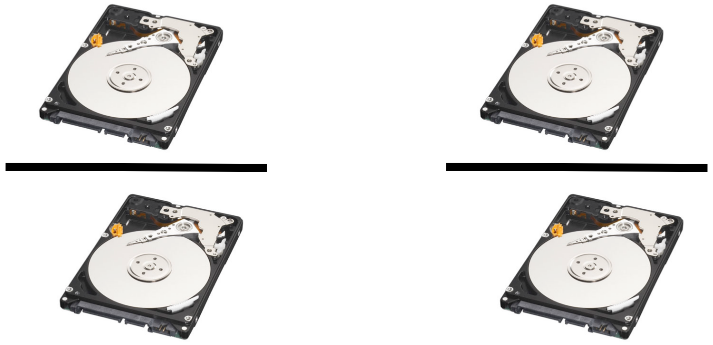
Composite Failures



= ?? years

Composite Failures

= 40.000 years



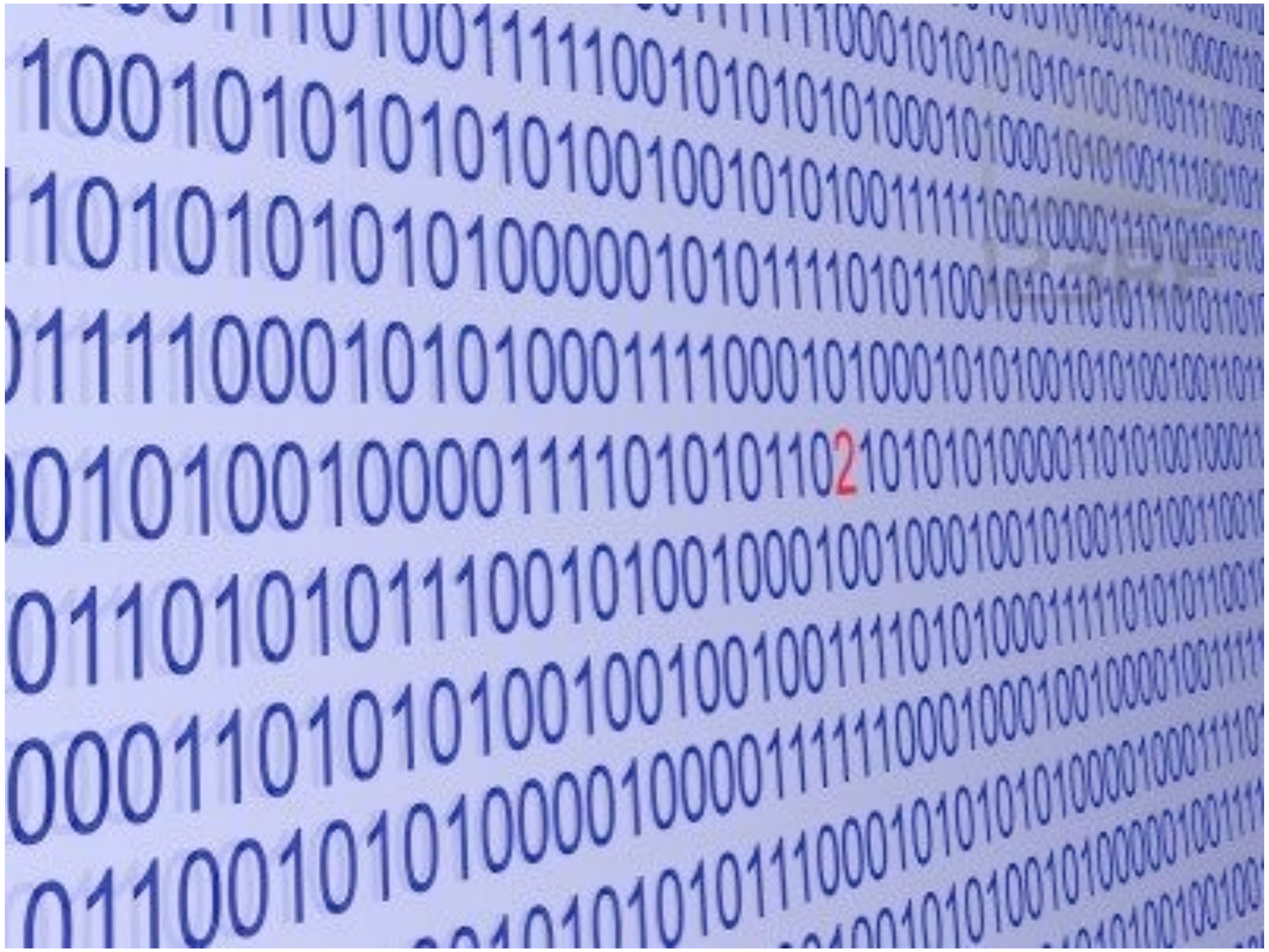
= SYSTEM

$$\frac{1}{\text{SYSTEM}} = \frac{1}{40.000} + \frac{1}{40.000}$$

SYSTEM = 20.000

Experiments

- Calculate the composite failure of the Tandem example (administration, software, hardware, environment)
- How would you make this setup more reliable? Calculate the effect
- What is the MTTF of a 5-way mirror of 7K3000 disks?

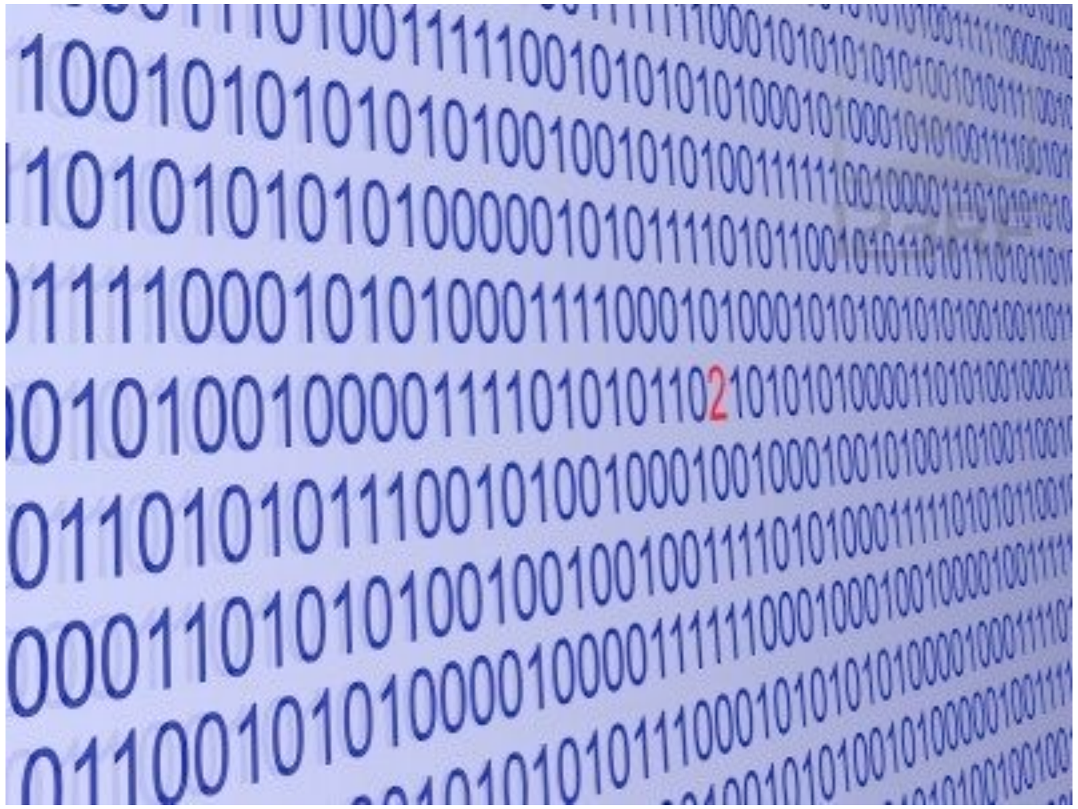


Analysing the Impact of File Formats on Data Integrity

Volker Heydegger; Universität zu Köln; Cologne, Germany

http://old.hki.uni-koeln.de/people/herrmann/forschung/heydegger_archiving2008_40.pdf

Bit Errors



0 1 1 0 0 0 1 0 1 1

↓

0 0 1 0 1 0 1 0 0 1

BER = Bit Error Rate = 3/10 = 0.3 = 30 %

Bit Errors

- Soft error - repeat the operation
- Hard error - after some repeats data is lost
- Typical disk BER = 10^{-5} to 10^{-6} (every 10KB to 100 KB read)

Bit Errors

Drive Type	Hard Error
Consumer SATA	10^{-14}
Enterprise SATA	10^{-15}
Enterprise SAS	10^{-16}

*) BER-s are in bit = 1/8 byte

10^{14} =~ 10 TB

10^{15} =~ 100TB

10^{16} =~ 1 PB

1 sector error for every 10 TB -> 1 PB read

Experiments

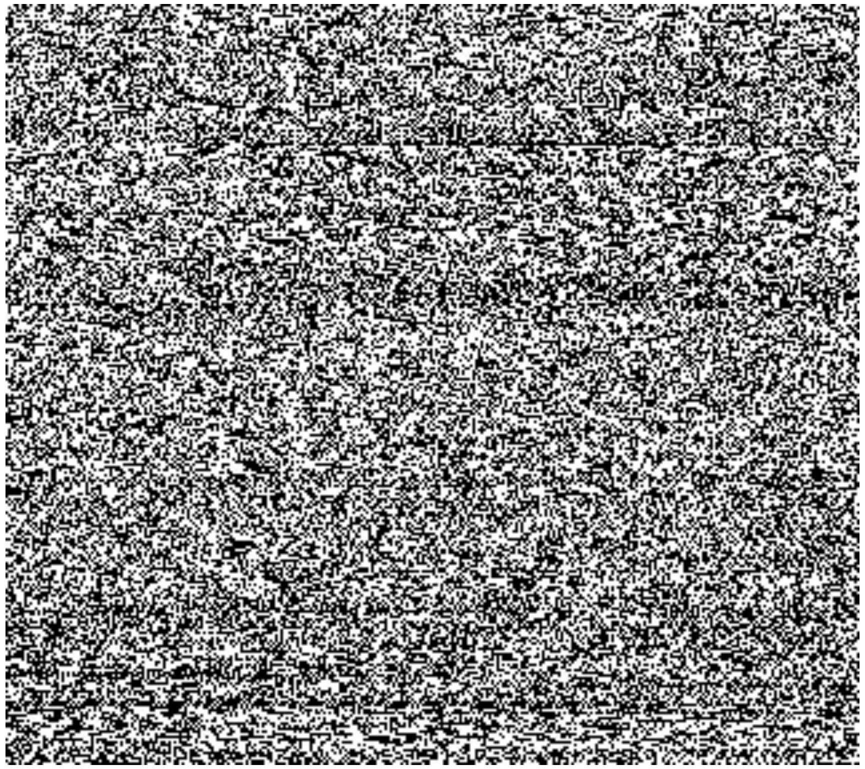
- Collect a few sample document from the web (images, documents, executables, etc); flip one or more random bits; explain the resulting effect
- Use the visual defects experiment to measure the effect of flipping bits on images files with various compressions
- Open and save an image file. Measure the visual effects.
- Calculate the checksum of the files and repeat the experiments. Check results.

PDF
@
HTTP
PNG
JPEG
GIF
MOVIE
ZIP
.html
.css
.js
XML

File Formats

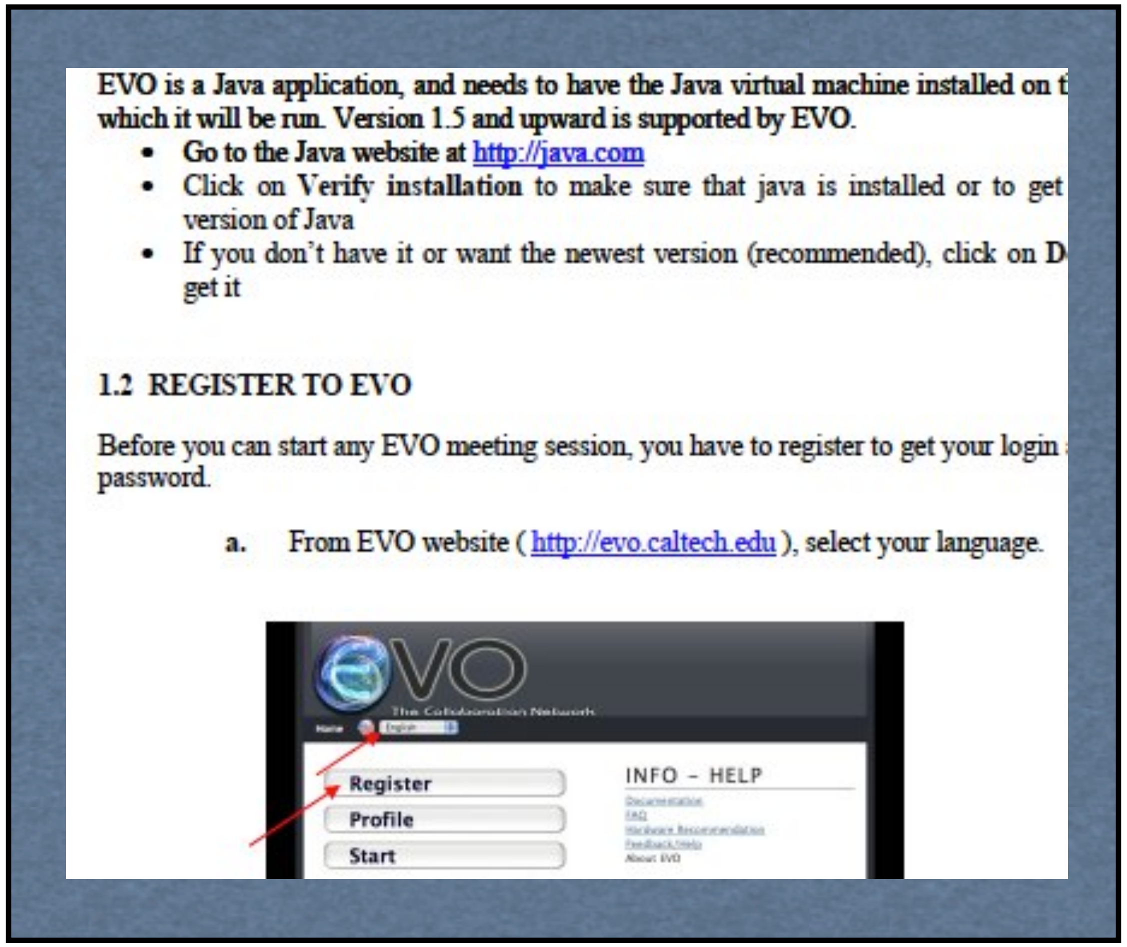
- The goal of digital preservation is not preserving the bits and bytes but the means to access and use the information represented by them.

File Formats



Bits

Software
+
Environment



EVO is a Java application, and needs to have the Java virtual machine installed on t
which it will be run. Version 1.5 and upward is supported by EVO.

- Go to the Java website at http://java.com
- Click on Verify installation to make sure that java is installed or to get
  version of Java
- If you don't have it or want the newest version (recommended), click on D
  get it

1.2 REGISTER TO EVO

Before you can start any EVO meeting session, you have to register to get your login
password.

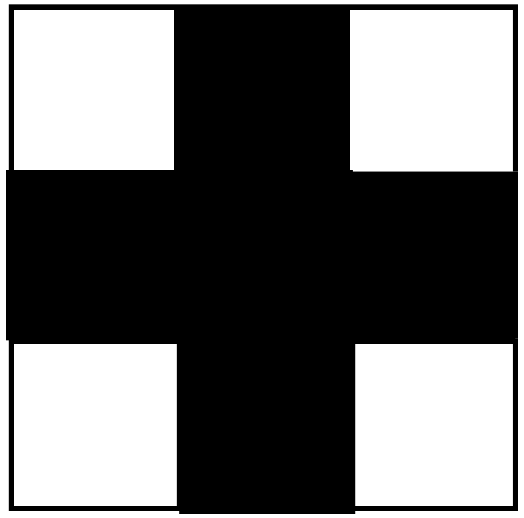
a. From EVO website (http://evo.caltech.edu), select your language.

Information

File Formats

hypothetical 3-bit format

1 1 0 1 1 0 0 1 0 1 1 1 0 1 0



Width = bit [1 .. 3]
Height = bit [4 .. 6]
Data = bit [7 .. 15]

File Formats

With software you have only two options:

1. The software works and is maintained
2. The software doesn’t work and is not maintained

File Formats

1. The software works and is maintained

- Your designated community has the software tools
- Your archive has the software tools
- In both cases you need to provide information which software you need and the steps required to get access to the data

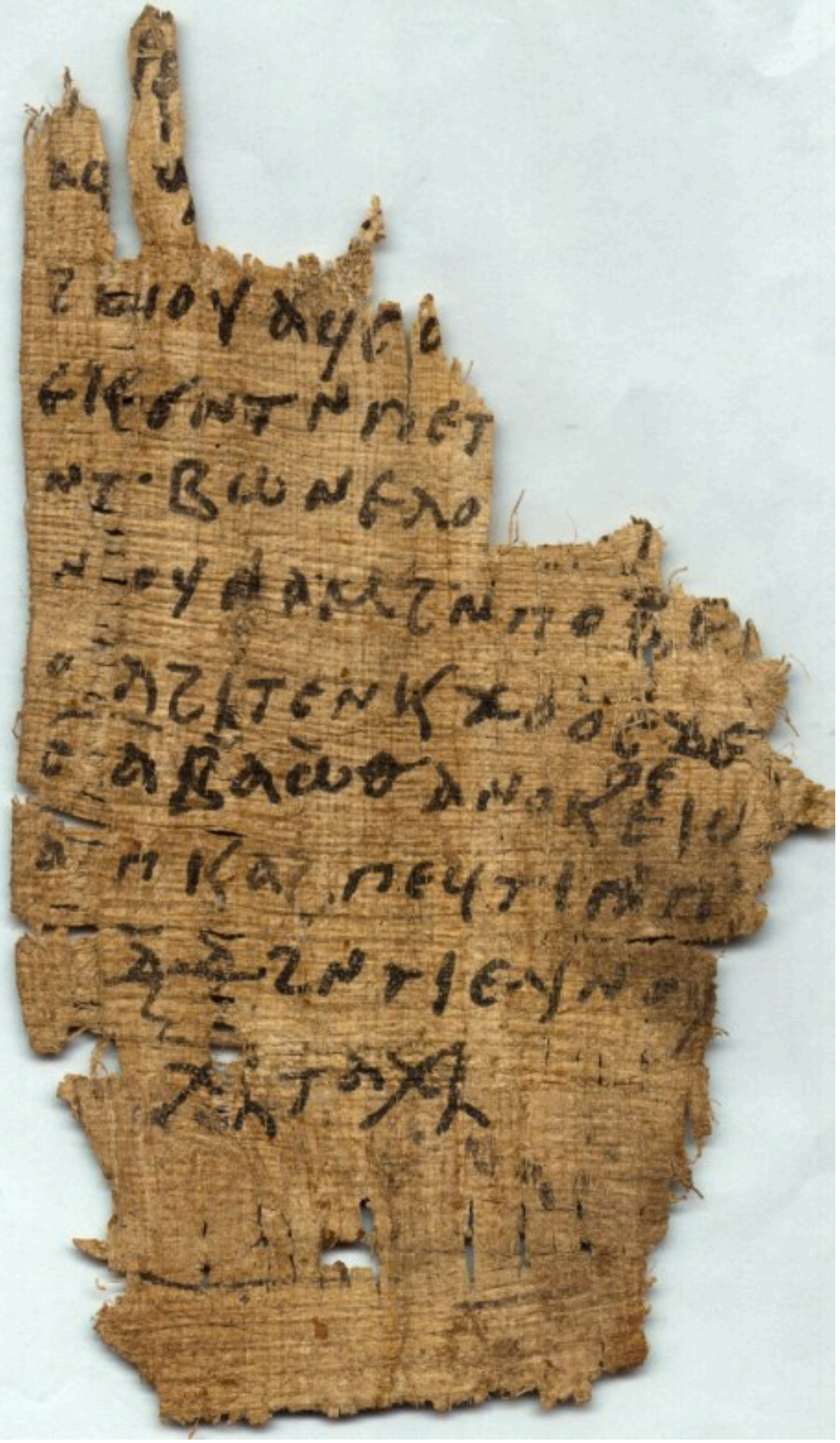
File Formats

2. The software doesn’t work and is not maintained

- Archive the source code of the orginal software
- Emulate the original software

Experiments

- Experiment with different textencoding demo files to discover the bit content of these files.
- Use droid and jhove to characterize and validate the demo files.
- Invalidate the files using truncation, bit errors. Check the results.
- Use migration and emulation to get access to the demo.wp file.



ⲁϥ ⲙ
ϩⲙⲟⲟⲩ ⲁⲩⲱ
ⲉⲓⲉ ϭⲛⲧ ⲛ̄ ⲡⲉⲧ
ⲛ̄ⲧ · ⲃⲱⲱⲛ ⲉⲃⲟⲗ
ⲛ̄ⲟⲩϣⲁⲁⲕⲉ ϩⲛ̄ ⲡⲟⲃϣ
ⲉⲇⲓⲧⲉⲛ ⲕ ϫⲟⲟⲥ
ⲁⲃⲃⲁ ⲁⲗⲱⲟ ⲁⲛⲟⲕ ⲉⲓⲱ
ⲁ ⲙ ⲕⲁⲓ ⲡⲉϥⲧⲓⲙⲉ
ⲇⲉ ϩⲛ̄ ⲧⲓⲉⲩⲛⲟ
ⲡⲗⲧⲟⲭ

Metadata

- Descriptive Metadata
- Administrative Metadata
- Structural Metadata
- Rights Metadata
- Representation Metadata

Packaging

- Digital objects are composite structures
- Need to be described, validated and accessed as a whole
- Complex Objects

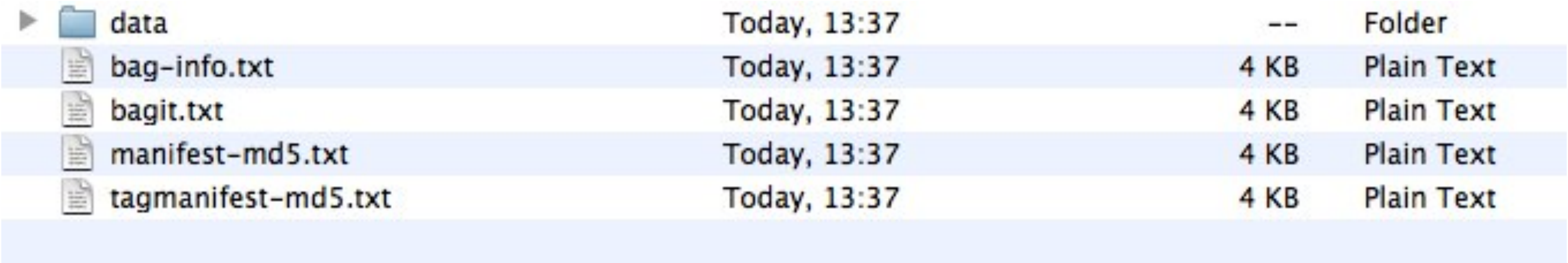
Package Formats

- METS
- MPEG-21/DIDL
- LOM/IMS
- BagIt
- TIPR RXP

BagIt

- Library of Congress & California Digital Library
- NDIIP
- Generic Format

BagIt

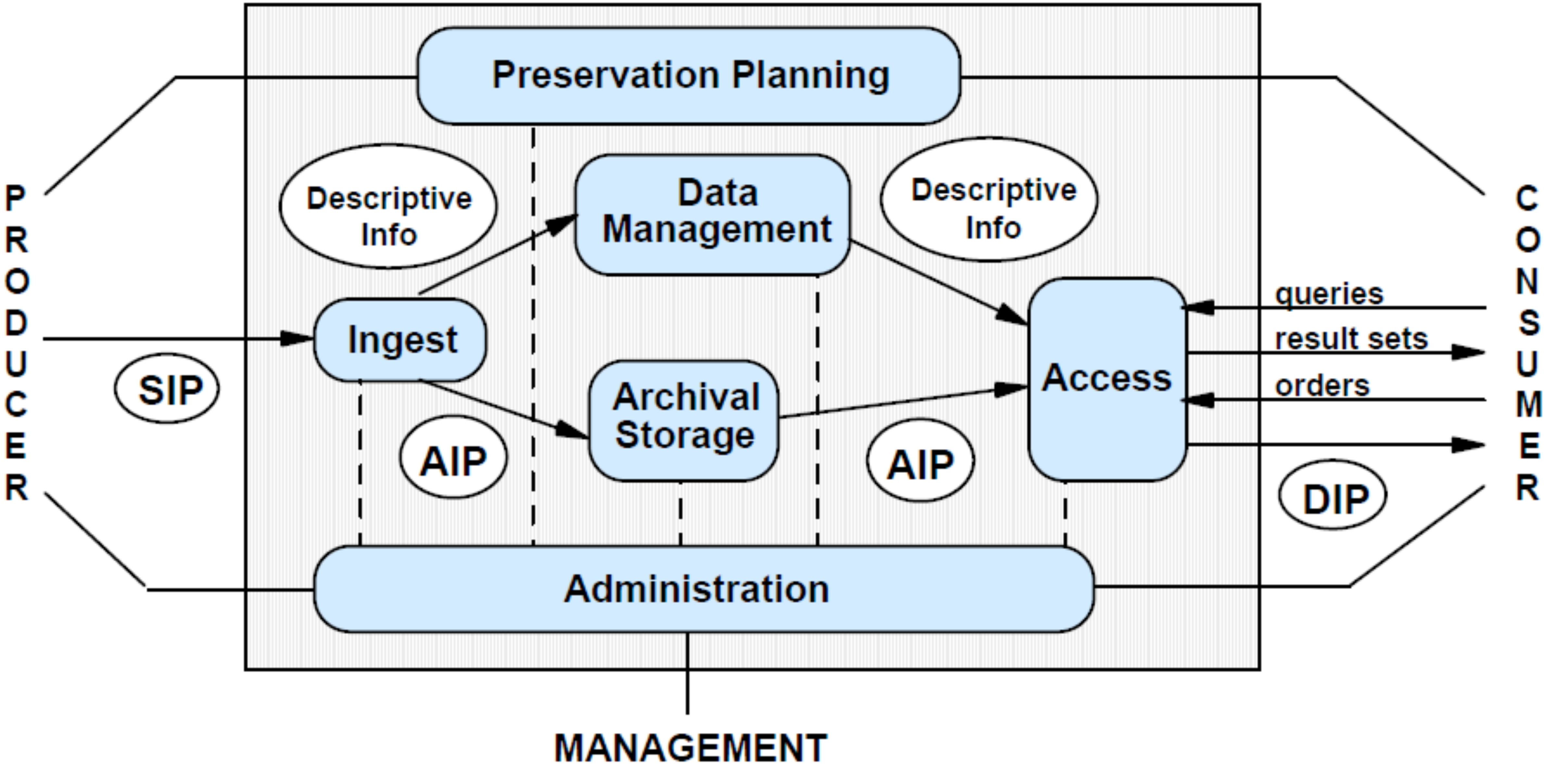


data	Today, 13:37	--	Folder
bag-info.txt	Today, 13:37	4 KB	Plain Text
bagit.txt	Today, 13:37	4 KB	Plain Text
manifest-md5.txt	Today, 13:37	4 KB	Plain Text
tagmanifest-md5.txt	Today, 13:37	4 KB	Plain Text

Experiments

- Create using the Bagger toolkit a bag. Add Dublin Core descriptive metadata.
- Save the bag as ZIP-file and deposit it do the demo archive.
- As archivist access the deposit and validate its contents.

Conclusions


Preservation Planning
PRODUCER
Descriptive Info
Data Management
Descriptive Info
CONSUMER
queries
result sets
orders
Ingest
SIP
Access
Archival Storage
AIP
AIP
DIP
Administration
MANAGEMENT