

Data Migration: Will 1,150 Years of Media Longevity be Enough?

Ph. D. **Hrvoje Stancic**, associate professor
Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences
University of Zagreb, Croatia
hrvoje.stancic@zg.t-com.hr

Contents

1. Approach to the preservation of e-materials
2. Structure of e-information objects
3. Levels of relationship
4. Impact of changes at the preservation of records
5. Data migration
6. Conclusion

1. Approach to the preservation of e-materials

- fast and constant ICT development
 - speed of data processing
 - network throughput
 - ways of connecting
 - services
- impact on MLA institutions, information-documentation centers, firms...
- e-materials/heritage – how to preserve for the long-term

1. Approach to the preservation ...

- fast degradation of:
 - medium
 - format
 - access
 - view
 - search
 - usage
 - trustworthiness / authenticity

This page does not seem to exist...

We apologize for the inconvenience, but the page you were trying to access is not at this address. You can use the links below to help you find what you are looking for.

If you are certain you have the correct web address but are encountering an error, please contact the **Site Administration**.

Thank you.

Question(s)

- How many of you have ever lost valuable data?
and
- Were you able to retrieve lost data?
or
- Did you have a back-up?
- Please share
the experience!



1. Approach to the preservation ...

- Information object
 - any material

- e-Information object
 - any material created by the usage of information technology
 - digitization
 - digitally born materials

1. Approach to the preservation ...

- Every form of information object requires specific approach to the process of preservation!
- Information object
 - physical, logical and conceptual level
 - characteristics and interrelations
 - important for preservations

1. Approach to the preservation ...

- get to know and preserve inter-level relationships
- proactive care for the preserved materials
 - media refreshment
 - migration
 - ...
- preserve authenticity

1. Approach to the preservation ...

- authenticity

- analogue materials (check) – person + material
- e-material (check) – separated from the material ⇒ transmission
- static vs. dynamic materials
- simple vs. complex materials
- content vs. functionality
- context

2. Structure of e-information objects

- e-Information objects
 - creation method unimportant (digitisation / born digitally)
 - each form – specific preservation method
- Levels:
 - physical
 - logical
 - conceptual / intellectual
- Characteristics inter-level relationships

2. Structure – physical level

- Physical level
 - level of recording electronic information object to a medium
 - sign system appropriate for computer recording and processing (0 and 1)
 - record mode dependent on the media
 - magnetic vs. optical media
 - compressed vs. uncompressed
 - device – interface between binary record and media

2. Structure – physical level ...

- no interpretation of meaning
- preservation problems
 - media longevity
 - record longevity
- analogue vs. e-materials – preservation of medium and contents not the same
- separation of the medium longevity problems and record longevity problems

2. Structure – physical level ...

- media longevity
 - depend on the medium type
 - dependant on the physical environmental conditions
 - dependant on the hardware-software system

Question

- What do you think how important it is to have a long-lasting medium?

2. Structure – logical level

- define the way of physical organisation and recording
 - ignores type of medium and format
- must be physically recorded – inter-level relationship
 - simple and complex objects
 - e.g. book
 - chapters (one object – multiple parts)
 - book – logical object of relationships + software information
 - multiple objects – one object (ZIP)

2. Structure – logical level ...

- physical level preservation + logical level preservation
 - physical object loading
 - logical recognition of objects
 - order
 - reading
 - processing
- preservation of inter-level relationship

2. Structure – conceptual level

- object is recognised as a meaningful unit (document, book, melody etc.)
- intellectual level
- content and structure stored at the logical level
- possibility of different organisation at logical level ⇒ several logical records of the same conceptual realisation
- interpretation depends on the physical realisation (.docx vs. .pdf – .mdb vs. .pdf)

2. Structure – conceptual level ...

- permit the possibility of records in the different form – adequate to the conceptual realisation and *intended usage*
- Example!

Očuvanje elektroničkih informacijskih objekata: arhivi, knjižnice, muzeji – zajednička koncepcija

Preservation of Electronic Information Objects: Archives, Libraries, Museums – Common Concepts

Sažetak

Uvodno autor određuje kontekst problema očuvanja gradiva u elektroničkom obliku, potom definira koncept elektroničke baštine, te objašnjava najznačajnije oblike u kojima se ona pojavljuje. Nadalje, autor analizira elektroničku baštinu kao elektronički informacijski objekt, pri čemu izlaže o pitanju razlikovanja elektroničkog originala od elektroničke kopije.

Očuvanje elektroničkih informacijskih objekata: arhivi, knjižnice, muzeji – zajednička koncepcija

Preservation of Electronic Information Objects: Archives, Libraries, Museums – Common Concepts

Sažetak

Uvodno autor određuje kontekst problema očuvanja gradiva u elektroničkom obliku, potom definira koncept elektroničke baštine, te objašnjava najznačajnije oblike u kojima se ona pojavljuje. Nadalje, autor analizira elektroničku baštinu kao elektronički informacijski objekt, pri čemu izlaže o pitanju razlikovanja elektroničkog originala od elektroničke kopije.

Očuvanje elektroničkih informacijskih objekata: arhivi, knjižnice, muzeji – zajednička koncepcija

Preservation of Electronic Information Objects: Archives, Libraries, Museums – Common Concepts

Sažetak

Uvodno autor određuje kontekst problema očuvanja gradiva u elektroničkom obliku, potom definira koncept elektroničke baštine, te objašnjava najznačajnije oblike u kojima se ona pojavljuje. Nadalje, autor analizira elektroničku baštinu kao elektronički informacijski objekt, pri čemu izlaže o pitanju razlikovanja elektroničkog originala od elektroničke kopije.

.doc

.pdf

.jpg

Conceptual level

2. Structure – conceptual level ...

- each realisation functions well – depending on the *intended usage*
 - reading (page) (.doc, .pdf, .jpg)
 - printing (.doc, .pdf, .jpg)
 - copying parts of text (.doc, .pdf)
 - text editing (.doc)

.doc

```
D0 CF 11 E0 A1 B1 1A E1 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 3E 00 03 00 FE FF 09 00
06 00 00 00 00 00 00 00 00 00 00 00 02 00 00 00
E3 00 00 00 00 00 00 00 00 10 00 00 E5 00 00 00
01 00 00 00 FE FF FF FF 00 00 00 00 E1 00 00 00
E2 00 00 00 FF FF FF FF FF FF FF FF FF FF FF FF FF
FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
```

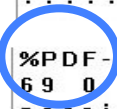
.pdf

```
25 50 44 46 2D 31 2E 33 0D 25 E2 E3 CF D3 0D 0A
36 39 20 30 20 6F 62 6A 0D 3C 3C 20 0D 2F 4C 69
6E 65 61 72 69 7A 65 64 20 31 20 0D 2F 4F 20 37
31 20 0D 2F 48 20 5B 20 31 32 36 31 20 35 37 32
20 5D 20 0D 2F 4C 20 32 30 35 31 34 39 20 0D 2F
45 20 39 33 30 39 38 20 0D 2F 4E 20 31 35 20 0D
2F 54 20 32 30 33 36 35 31 20 0D 3E 3E 20 0D 65
6E 64 6F 62 6A 0D 20 20 20 20 20 20 20 20 20 20
20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20
20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20
20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20
66 0D 36 39 20 34 31 20 0D 30 30 30 30 30 30 30 30
30 31 36 20 30 30 30 30 30 20 6E 0D 0A 30 30 30
30 30 30 31 31 36 38 20 30 30 30 30 30 20 6E 0D
```

.jpg

```
FF D8 FF E0 00 10 4A 46 49 46 00 01 02 01 00 C8
00 C8 00 00 FF E2 04 90 49 43 43 5F 50 52 4F 46
49 4C 45 00 01 01 00 00 04 80 53 4F 4E 4E 02 00
00 00 6D 6E 74 72 52 47 42 20 58 59 5A 20 07 D2
00 0A 00 12 00 10 00 1D 00 00 61 63 73 70 4D 53
46 54 00 00 00 00 50 6E 50 20 49 6D 61 67 00 00
00 00 00 00 00 00 00 00 00 01 00 00 F6 D5 00 01
00 00 00 00 D3 2C 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
01 08 00 00 00 14 63 70 72 74 00 00 01 1C 00 00
00 38 4B 30 30 37 00 00 01 54 00 00 00 E8 64 65
73 63 00 00 02 3C 00 00 01 0E 77 74 70 74 00 00
```

```
Ð Ð í ¸ ± ¡ á
> | } ' |
| |
ã | í
| } ' . . . . . á
â . . . . .
. . . . .
. . . . .
. . . . .
. . . . .
. . . . .
. . . . .
. . . . .
. . . . .
```



```
%PDF-1.3 %ã ã Ð Ó |
69 0 obj << | / Li
nearized 1 | / O 7
1 | / H [ 1261 572
] | / L 205149 | /
E 93098 | / N 15 |
/ T 203651 | >> | e
nd obj |
```

```
x r e
f | 69 41 | 00000000
016 00000 | n | 000
0001168 00000 | n |
```

```
  í | J F I F | | | 
  â | | | | C C _ P R O F
| L E | | | | S O N N I
m n t r R G B X Y Z | Ñ
| | | | a c s p M S
F T | | | | m a g
| | | | ð Ó |
ó ,
| | | | |
| | | | | c a l t
| | | | | c p r t | |
8 K 0 0 7 | | T | |  d e
s c | | < | | w t p t
```

Logical level

2. Structure – conceptual level ...

- conceptual level
 - feeling that all documents are the same
- logical level
 - no feeling that all documents are the same
- Expected! Different file formats!

MS Word 2002

.doc

D0	CF	11	E0	A1	B1	1A	E1	00	00	00	00	00	00	00	00	00	00
00	00	00	00	00	00	00	00	3E	00	03	00	FF	FF	09	00		
06	00	00	00	00	00	00	00	00	00	00	00	02	00	00	00		
E3	00	00	00	00	00	00	00	00	10	00	00	E5	00	00	00		
01	00	00	00	FE	FF	FF	FF	00	00	00	00	E1	00	00	00		
E2	00	00	00	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		

Ð Ñ Ò Ó Ô Õ Ö × Ø Ù Ú Û Ü Ý Þ ß à

> | } ' |

| | | |

ǎ | í | á

| } ' ' ' á

à

.....

.....

.....

.....

.....

.....

.....

.....

.....

.doc

D0	CF	11	E0	A1	B1	1A	E1	00	00	00	00	00	00	00	00	00	00
00	00	00	00	00	00	00	00	3E	00	03	00	FF	FF	09	00		
06	00	00	00	00	00	00	00	00	00	00	00	01	00	00	00		
4A	00	00	00	00	00	00	00	00	10	00	00	4B	00	00	00		
01	00	00	00	FE	FF	FF	FF	00	00	00	00	49	00	00	00		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		
FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF		

Ð Ñ Ò Ó Ô Õ Ö × Ø Ù Ú Û Ü Ý Þ ß à

> | } ' |

| | | |

ǎ | í | á

| } ' ' ' á

à

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Logical level

MS Word 6.0

2. Structure – conceptual level ...

- different / the same file?
- looking at the logical level we do not expect the same document
- loading and saving the file in the new version may impact the authenticity
- loading and saving the file in the same version may impact the authenticity
- new application versions usually correctly read older formats of the same application, but not vice versa

2. Structure – conceptual level ...

- existence of different logical realisation of the same conceptual object
- differences
 - due to different file formats
 - due to differences in versions within the same file format
- important
 - keep correct inter-level relationships

3. Levels of relationship

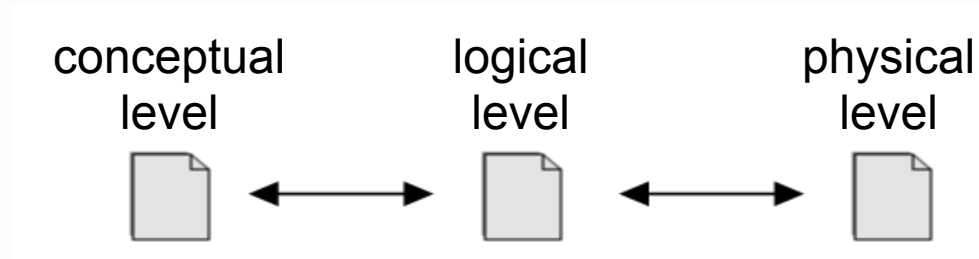
Thibodeau, Kenneth, **Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years**, in: *The State of Digital Preservation: An International Perspective*, Council on Library and Information Resources (CLIR), Washington, D.C., SAD, July 2002, pp. 4-31, <http://www.clir.org/pubs/reports/pub107/pub107.pdf>, 9 February 2014

✂ multiple inter-related relationships of physical, logical and conceptual levels

3. Levels of relationship ...

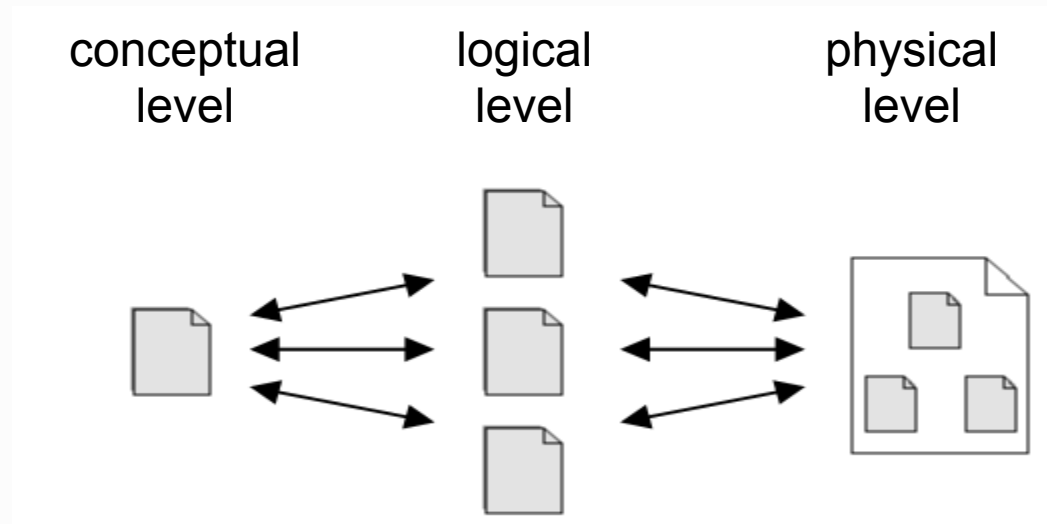
- one-to-one

- one document saved in one file is a single object at all three levels



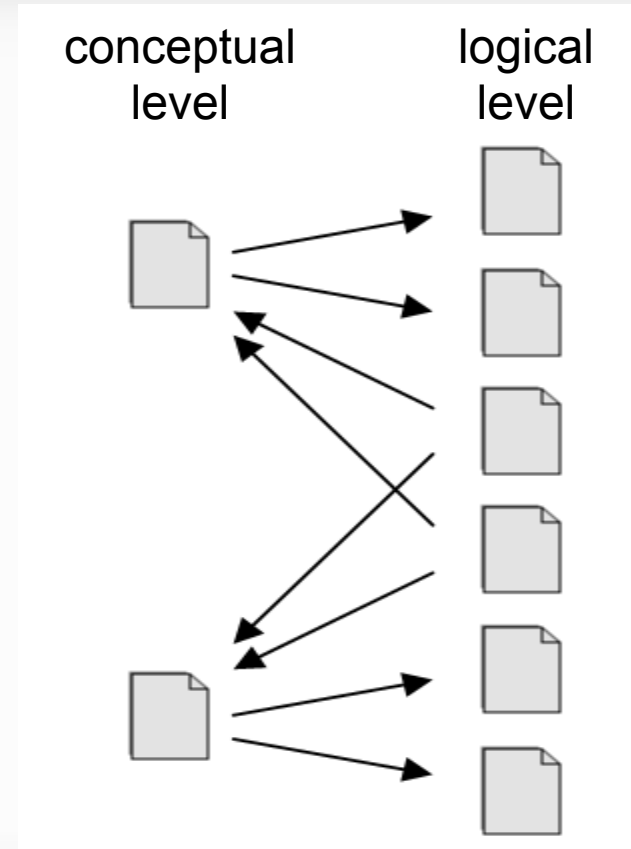
3. Levels of relationship ...

- one-to-many or many-to one
 - one document separated in three parts compressed in one .zip file



3. Levels of relationship ...

- many-to-many
 - repeated usage of a report template for appearance, content and structure while the content is filled-in from the different tables of the database

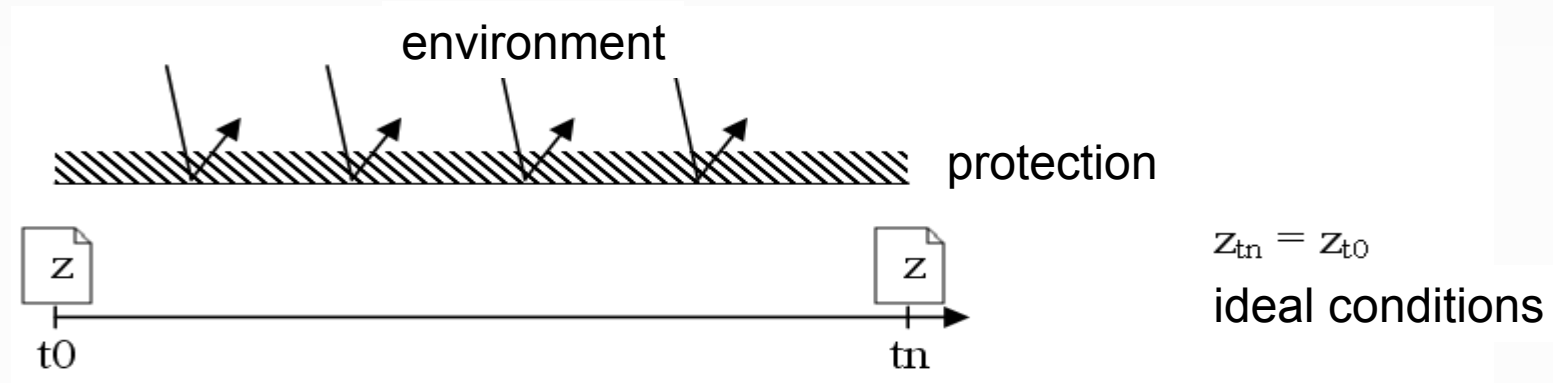


3. Levels of relationship ...

- long-term preservation
 - know and understand inter-level relationships
 - know and understand inter-object relationships
- the impact of time
 - physical changes on media
 - media generations
 - logical level – file versions
 - intellectual – understanding a document from the context – importance of authenticity

4. Impact of changes at the preservation of records

- ideal conditions – no impact on objects



z – record

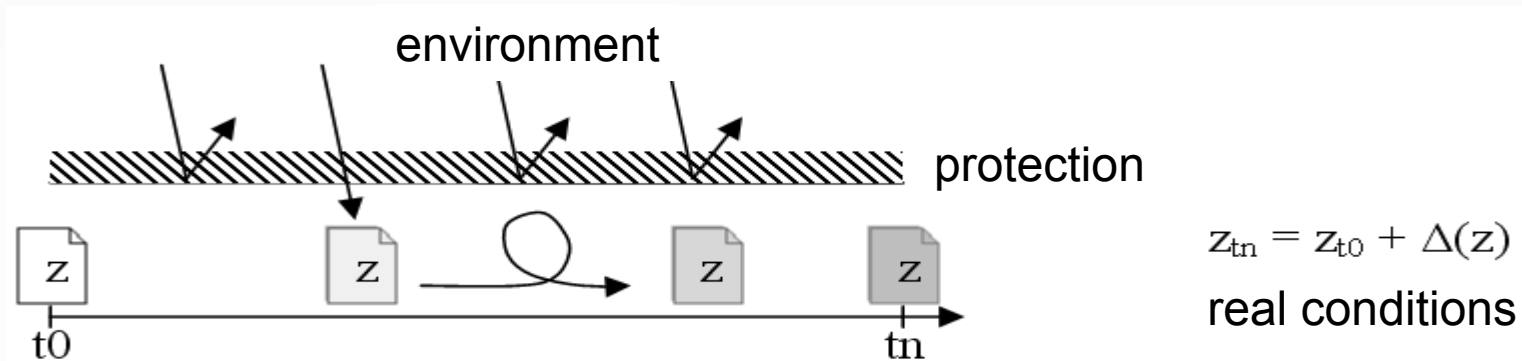
t0 – time of ingesting an object into the preservation process

tn – later time of access to the preserved record

- ideal conditions do not exist!

4. Impact of changes at the preservation of records

. real conditions



z – record

t_0 – time of ingesting an object into the preservation process

t_n – later time of access to the preserved record

$\Delta(z)$ – expected or unexpected changes of the record during the preservation time (impacts from the environment, media refreshment, migrations

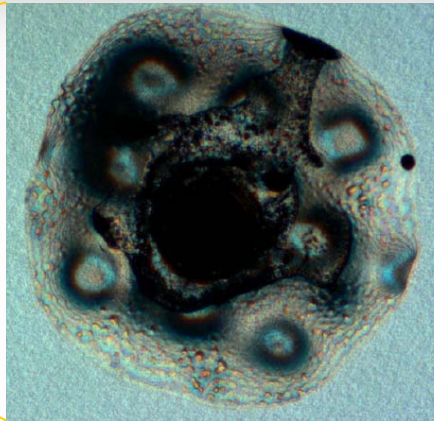
etc.)

5. Migration

- Migration is needed!
- E.g. THE CURRENT PROBLEM OF DATA LOSS WITH OPTICAL MEDIA
 - "Commonly used recording layer materials for CD-R/ DVD±R media are organic substances from the group 'cyanine' and 'phtalocyanine'. The disadvantages of the organic materials are the crystalline structure that absorbs humidity and the material properties change with time due to the impact of light.
Due to humidity, after writing data onto the organic recording layer, a change in crystals' size occurs. Consequently, it results in change of pit size to such extent that the reading drives lose their ability to recognize the size of each pit and so the disc becomes unreadable."

Data Tresor Disk, <http://www.datatresordisc.com/english/by-product-dtd.html>, 9 February 2014

Optical media degradation



Degenerative spot as seen through microscope

Chemical degradation

Optical media degradation – research

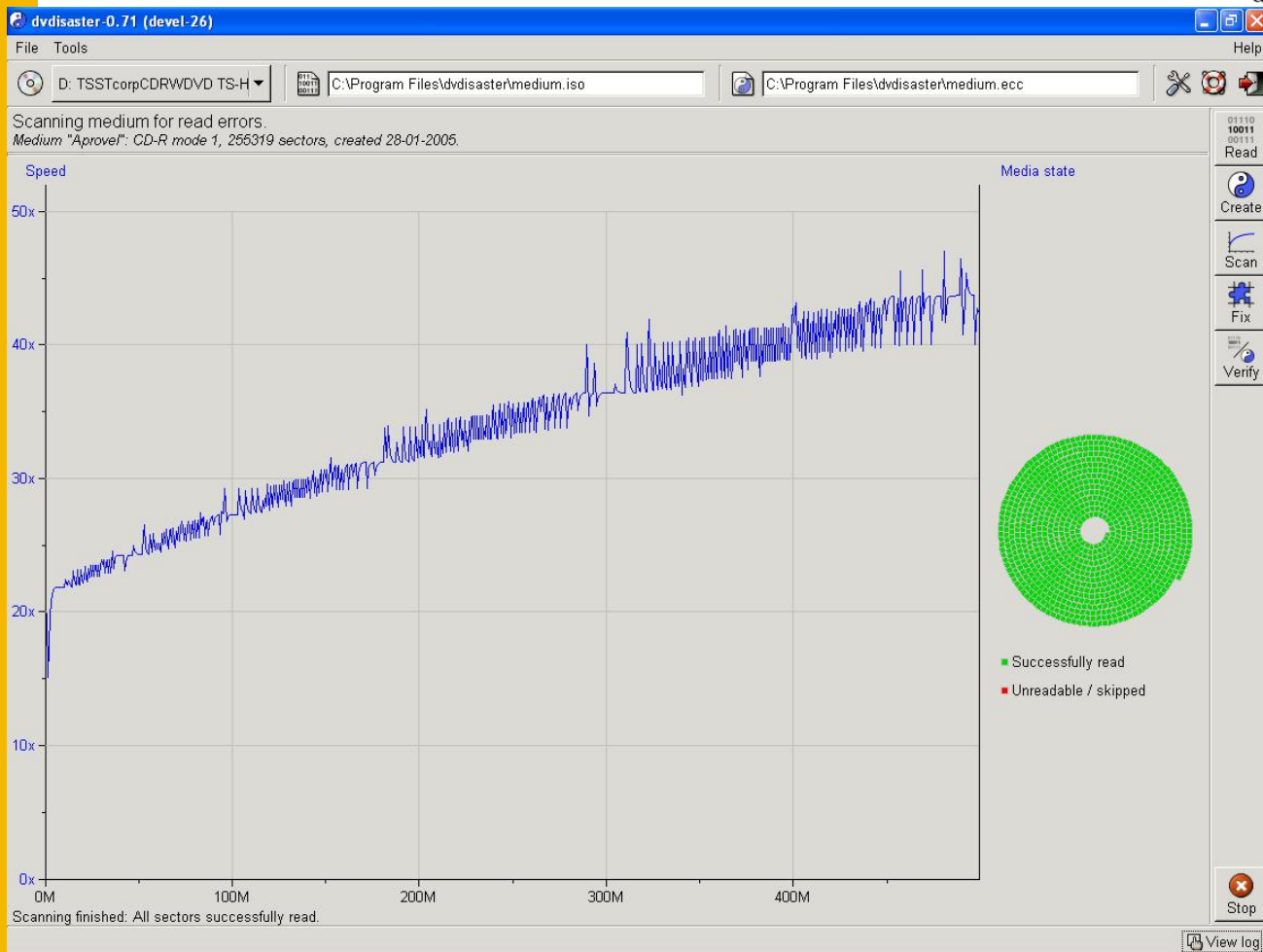
- Pilot study done at the Croatian Agency for Medicinal Products and Medical Devices – in 2009
- Research was one of the activities of DPE-Croatia
- Checking of the readability and sustainability of optical media (CDs)
- Aim: determine possible correlations of different factors and their impact on the long-term preservation (LTP) of the content on optical media
- 279 tested CDs

Optical media degradation – research ...

- We have tested and investigated CDs:
 - visually (scratches on surface, media brand)
 - using programs (Windows Explorer, CDR Media Code Identifier, CD-R Diagnostic, DVDDisaster)
 - size of the contents, number of files and folders
 - disc manufacturer, nominal capacity
 - recording application, recording date, chemical type of the layer
 - number of recording sessions, media type, total number of recorded sectors, number of unreadable sectors



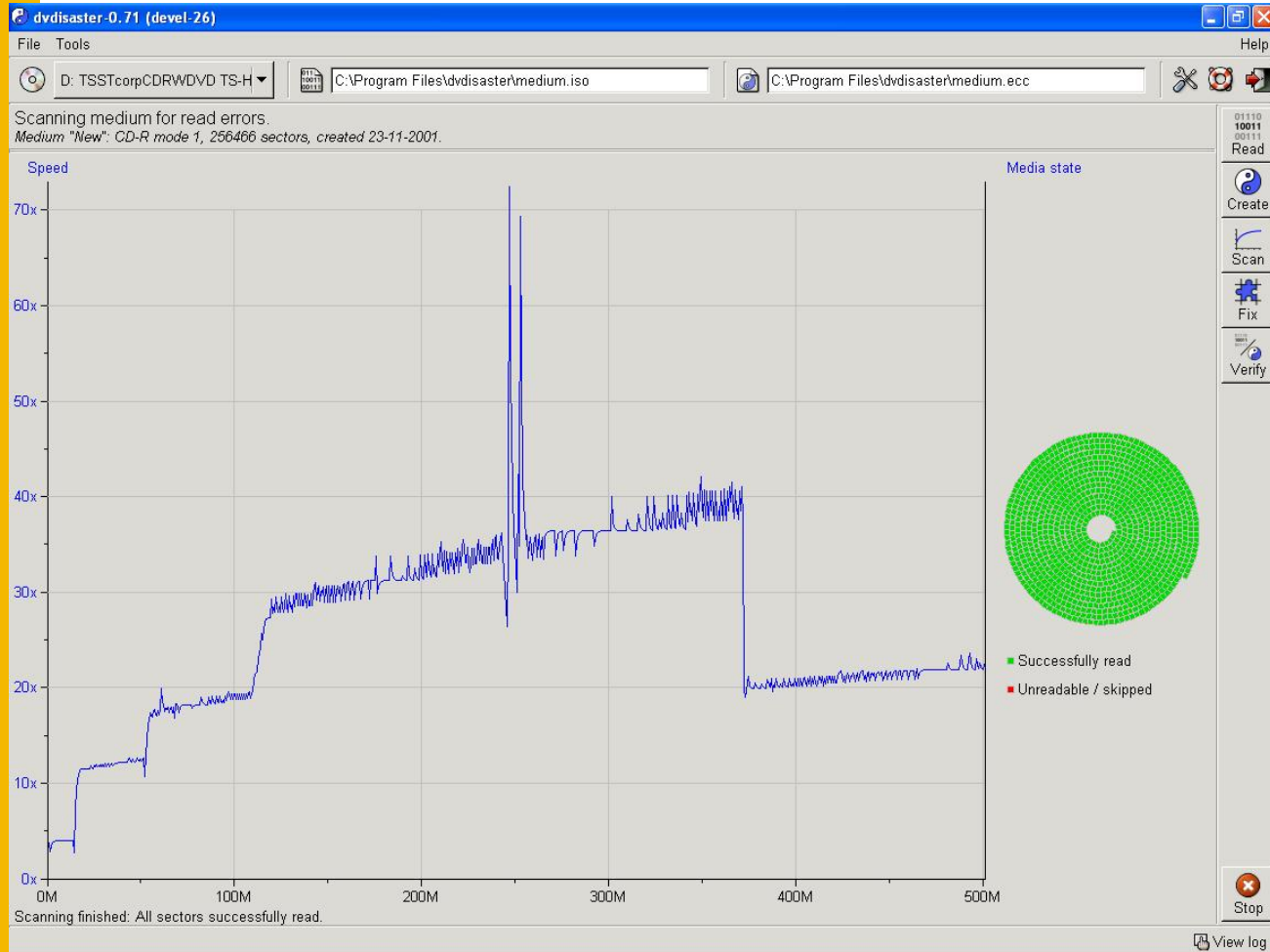
Successful results





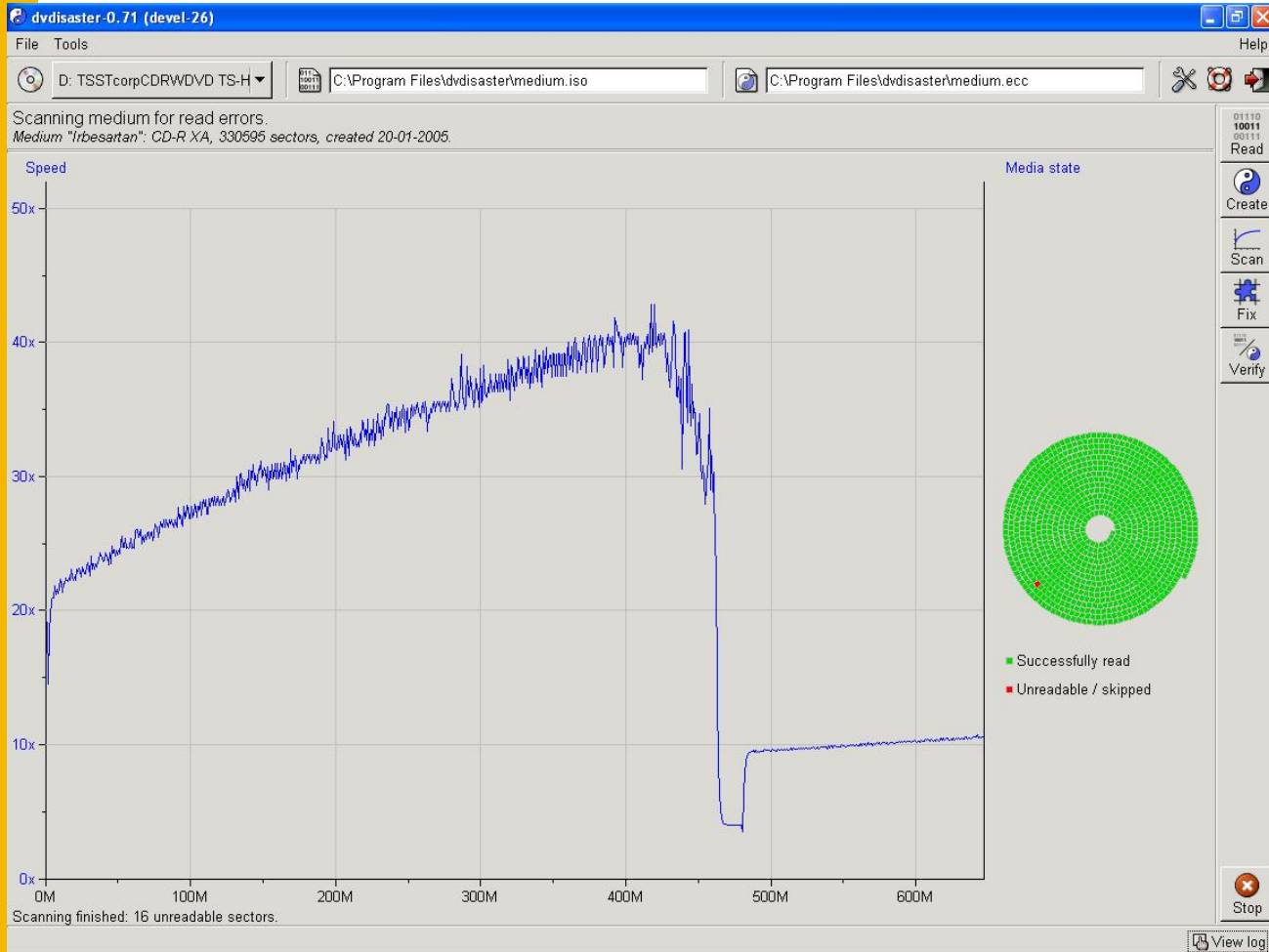
Results showing potential problems

digital preservation Europe



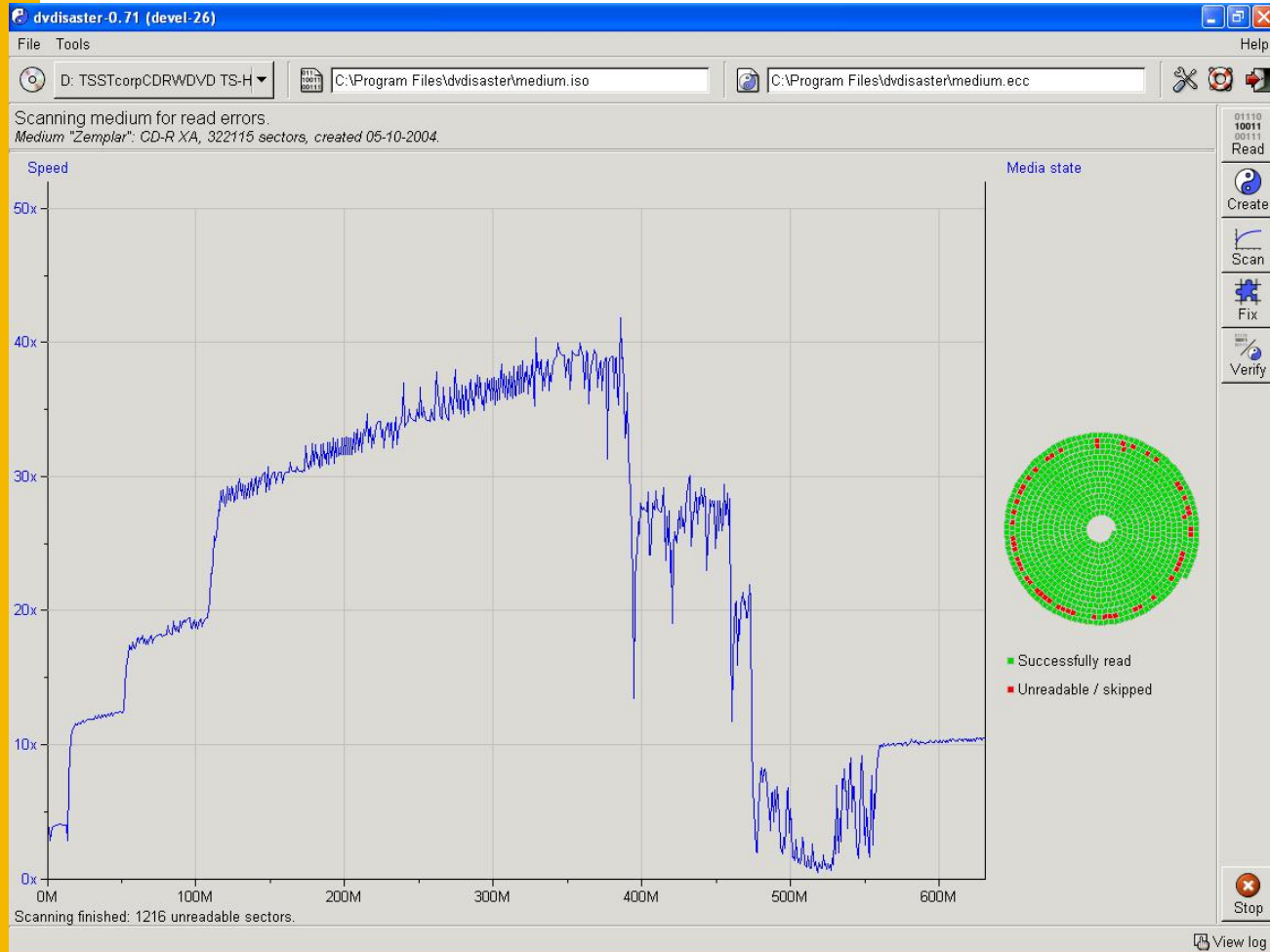


Results showing problems



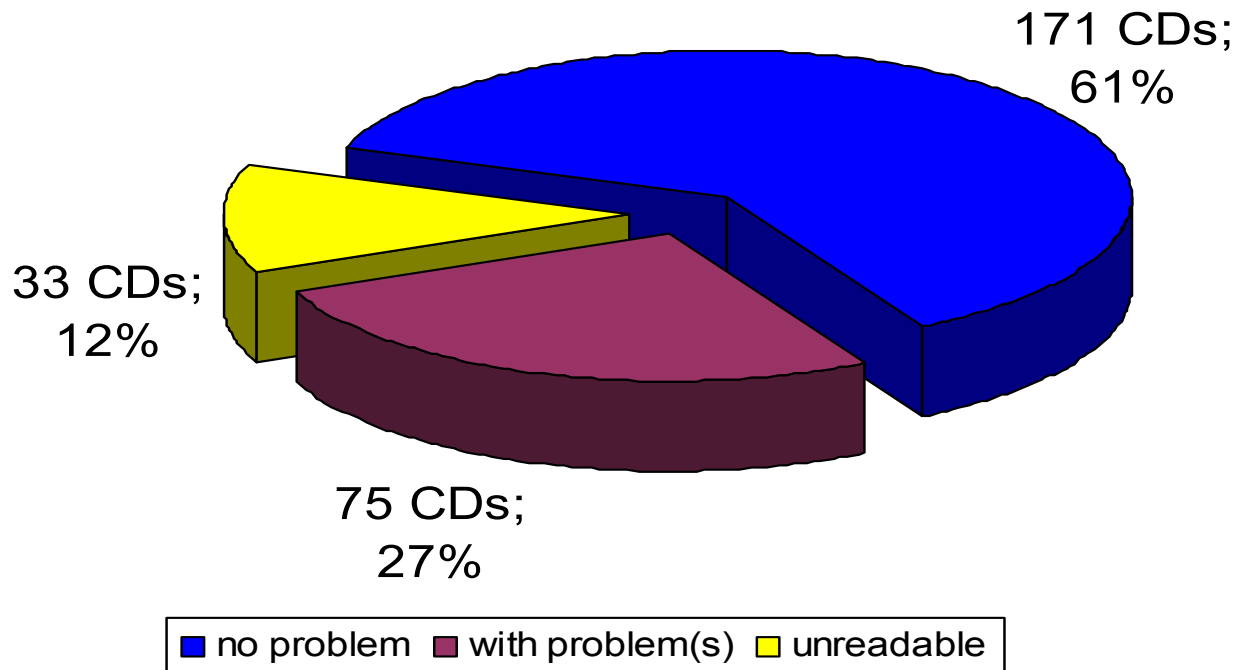


Results showing severe problems



Optical media degradation – research ...

- Overall percentage of detected problems (279 CDs)



Most of the problematic media were 4-5 years old!

What if ...

- ... we have a disc that could last for 1,150 years?

DATA TRESOR DISC
For how long?

OVER
1000
years*

DVD
up to 5yrs

HARD DRIVE
up to 7yrs

FLASH DRIVE
up to 8yrs

DTD
DataTresorDisc

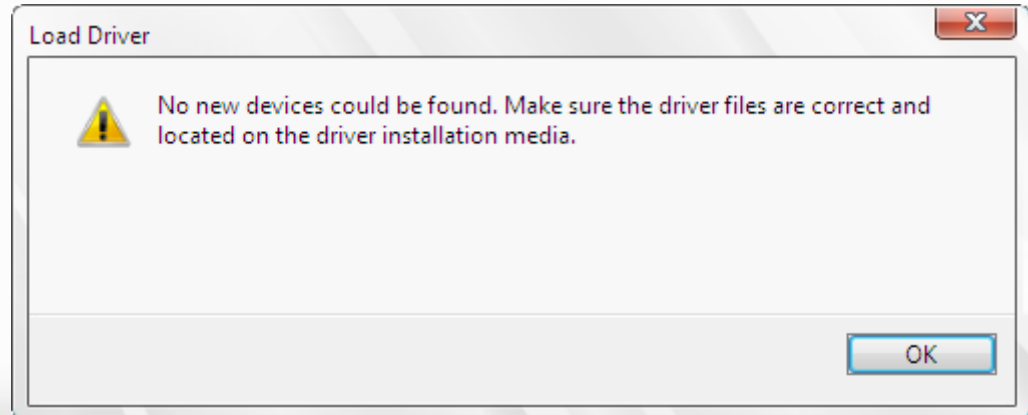
*Certified by the Federal center for conservation of the National Library of Russia



Warranted minimum
time of readability:
160 years

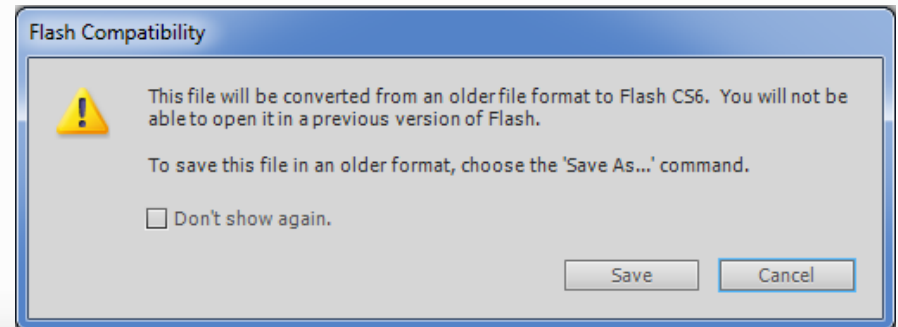
Media longevity question (let's go back in time)

- physical longevity vs. technological (un)obsolescence (5,25" floppy disk)
 - What problems do you see?



Media longevity ...

- Will the Data Tresor Disc (1,150 years) face the same problems in the future?
- How to approach the same problem in the networked world?
- Data has to be stored somewhere (even in the cloud solutions), on some kind of media!
- What about the file formats?



5. Migration ... yes, it is needed!

. Minimal sufficient requirement

- stability of the chosen new generation medium should be the same or greater than the approximate time until new migration, counting in the time needed for the migration procedure

$$SM \geq t_0 + \dots + t_{nm} + t_{mp}$$

S – stability

M – medium

t – time

nm – new migration

mp – migration procedure



6. Conclusion

- Proactive care is needed + the procedures of media
 - refreshment
 - migration
 - emulation / virtualization etc.
- + format migration
 - + consider the notion of authenticity where appropriate
- Close to ideal:
 - minimise uncontrolled impacts from outside
 - allow intentional and controlled changes
 - record enough metadata on the process

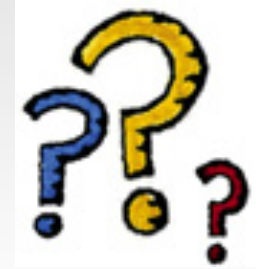
When you come home ...

- Check your 4 years or older CD / DVD collection.
- Try to copy the data to a hard drive.
- When you see me tomorrow – let me know if your valuable data is still there!

- Migrate as soon as possible!

THANK YOU!

**Data Migration:
Will 1,150 Years of Media
Longevity be Enough?**



Ph. D. **Hrvoje Stancic**, associate professor
Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences
University of Zagreb, Croatia
hrvoje.stancic@zg.t-com.hr