

Was ist digitale Langzeitarchivierung und was bedeutet sie für Museen?



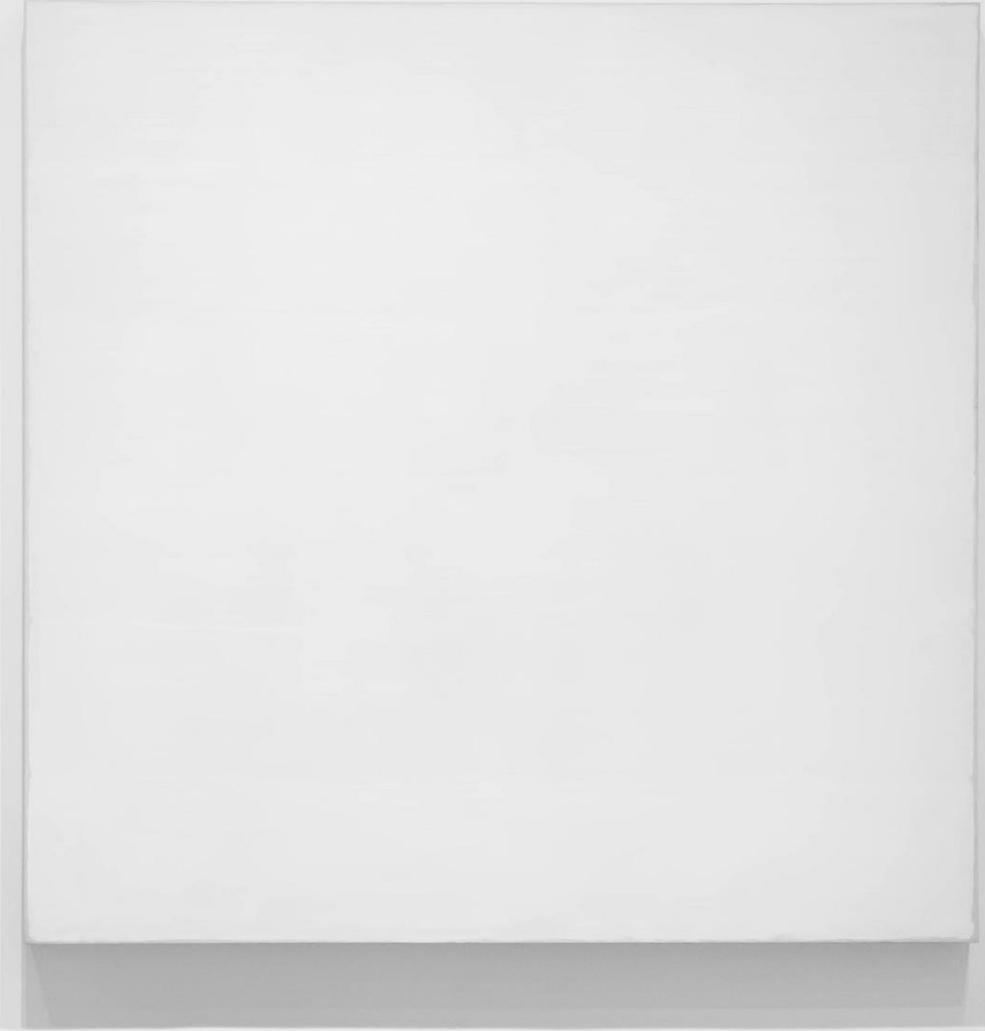
Worum es geht ...



Es geht nicht darum
ETWAS *auf digitale Weise* für lange Zeit zu archivieren

Es geht darum
DAS DIGITALE für lange Zeit zu archivieren

Vom Wesen des Digitalen ...



Vom Wesen des Digitalen ...



Schon in analoger Zeit wurden Fotos erstellt ...

Vom Wesen des Digitalen ...



... das setzt sich in digitaler Zeit fort



↓ Das Ergebnis des Fotografierens ist anders ↓





Eine analoge Kamera erzeugte ...

... eine analoge **Fotografie**.

Etwas Be-greifbares, das sich

- mit bloßem Auge sehen
- mit Händen anfassen und
- in Archivkisten verstauen lies.

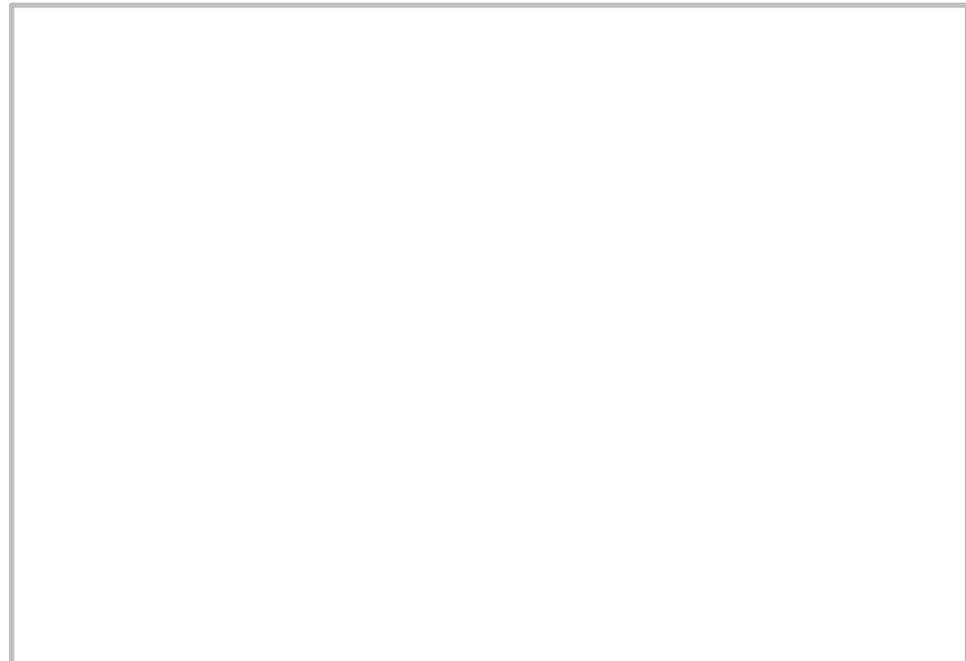
Eine Digitalkamera erzeugt ...

... eine **Bilddatei**.

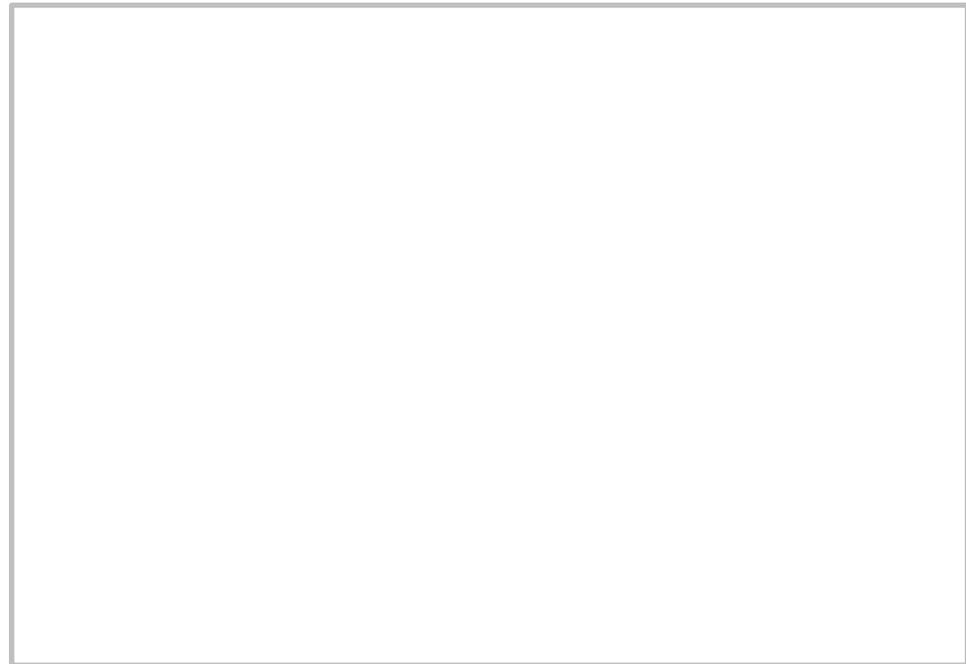


Etwas Nicht-greifbares, das sich

- weder mit bloßem Auge sehen
- noch mit Händen anfassen und sich
- nicht in Archivkisten verstauen lässt.



Natürlich ist eine Bilddatei nicht ganz ohne Struktur ...



Eine Bilddatei besteht aus Einsen und Nullen ...

```
100101101100011001010111011000100110
110010110110001100101011101100010011
011001011011000110010101110110001001
101100101101010001100101011101100010
011010010110110001100101011101100010
011011001011001000110010101110110001
100101101100011001010111011000100110
110010110110001001010111011000110010
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
```



Analoge Fotografien verändern sich

1:1 Kopien sind unmöglich

Nachteil



```
100101101100011001010111011000100110
110010110110001100101011101100010011
011001011011000110010101110110001001
101100101101010001100101011101100010
011010010110110001100101011101100010
011011001011001000110010101110110001
100101101100011001010111011000100110
110010110110001001010111011000110010
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
```

Von Bilddateien lassen sich identische Kopien erstellen

```
100101101100011001010111011000100110
110010110110001100101011101100010011
011001011011000110010101110110001001
101100101101010001100101011101100010
011010010110110001100101011101100010
011011001011001000110010101110110001
100101101100011001010111011000100110
110010110110001001010111011000110010
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
1000110010
1100010011
0110001001
1101100010
1101100010
1101100010
1110110001
1000100110
1000110010
11011000110010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
```

Vorteil



```
100101101100011001010111011000100110
110010110110001100101011101100010011
011001011011000110010101110110001001
101100101101010001100101011101100010
011010010110110001100101011101100010
011011001011001000110010101110110001
100101101100011001010111011000100110
110010110110001001010111011000110010
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
```

Eine Bilddatei besteht aus Einsen und Nullen ...



Wo ist das Problem ?

```
100101101100011001010111011000100110
110010110110001100101011101100010011
011001011011000110010101110110001001
101100101101010001100101011101100010
011010010110110001100101011101100010
011011001011001000110010101110110001
100101101100011001010111011000100110
110010110110001001010111011000110010
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
```



Eine Bilddatei besteht aus Einsen und Nullen ...

Eine Tondatei besteht aus Einsen und Nullen ...

Wo ist das Problem ?

```
100101101100011001010111011000100110
110010110110001100101011101100010011
011001011011000110010101110110001001
101100101101010001100101011101100010
011010010110110001100101011101100010
011011001011001000110010101110110001
100101101100011001010111011000100110
110010110110001001010111011000110010
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
```



Eine Bilddatei besteht aus Einsen und Nullen ...

Eine Tondatei besteht aus Einsen und Nullen ...

Eine Textdatei besteht aus Einsen und Nullen ...

Wo ist das Problem ?

```
100101101100011001010111011000100110
110010110110001100101011101100010011
011001011011000110010101110110001001
101100101101010001100101011101100010
011010010110110001100101011101100010
011011001011001000110010101110110001
100101101100011001010111011000100110
110010110110001001010111011000110010
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
```



Eine Bilddatei besteht aus Einsen und Nullen ...

Eine Tondatei besteht aus Einsen und Nullen ...

Eine Textdatei besteht aus Einsen und Nullen ...

Kommunikationsdaten sind Einsen und Nullen ...

```
100101101100011001010111011000100110
110010110110001100101011101100010011
011001011011000110010101110110001001
101100101101010001100101011101100010
011010010110110001100101011101100010
011011001011001000110010101110110001
100101101100011001010111011000100110
110010110110001001010111011000110010
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
```

Wo ist das Problem ?



Eine Bilddatei besteht aus Einsen und Nullen ...

Eine Tondatei besteht aus Einsen und Nullen ...

Eine Textdatei besteht aus Einsen und Nullen ...

Kommunikationsdaten sind Einsen und Nullen ...

Datenbankinhalte sind Einsen und Nullen ...

...

```
100101101100011001010111011000100110
110010110110001100101011101100010011
011001011011000110010101110110001001
101100101101010001100101011101100010
011010010110110001100101011101100010
011011001011001000110010101110110001
100101101100011001010111011000100110
110010110110001001010111011000110010
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
```

Wo ist das Problem ?

Das Problem:

```
100101101100011001010111011000100110  
110010110110001100101011101100010011  
011001011011000110010101110110001001  
101100101101010001100101011101100010  
011010010110110001100101011101100010  
011011001011001000110010101110110001  
100101101100011001010111011000100110  
110010110110001001010111011000110010  
110110001100101011101100010011011001  
011011000110010101010110001100101101  
100011001010111011000100110110010110
```

?
?
?
?
?



Die gleichen Einsen und Nullen können alles oder nichts bedeuten !

Das Problem:

```
100101101100011001010111011000100110  
110010110110001100101011101100010011  
011001011011000110010101110110001001  
101100101101010001100101011101100010  
011010010110110001100101011101100010  
011011001011001000110010101110110001  
100101101100011001010111011000100110  
110010110110001001010111011000110010  
110110001100101011101100010011011001  
011011000110010101010110001100101101  
100011001010111011000100110110010110
```

Bildprogramm <->



Audioprogramm <->



Textprogramm <->



Kommunikationsprogramm <->



Datenbankprogramm <->



„1 und 0“ müssen **zwingend** interpretiert werden, dafür sind **passende** Programme nötig.

Programm und Bilddatei müssen aufeinander abgestimmt sein !

Bildprogramm <- ->



Ohne passendes Programm ist der Inhalt der Datei nicht zu rekonstruieren !

Hinter diesem Problem liegt noch ein anderes ...

Programm und Bilddatei müssen aufeinander abgestimmt sein !

Bildprogramm <- ->



Ohne passendes Programm ist der Inhalt der Datei nicht zu rekonstruieren !

Hinter diesem Problem liegt noch ein anderes ...

Betriebssystem <- -> Bildprogramm

Ohne passendes Betriebssystem wird das passende Programm nicht laufen !

Betriebssystem und Bildprogramm müssen aufeinander abgestimmt sein !

Betriebssystem <- -> Bildprogramm

Ohne passendes Betriebssystem wird das passende Programm nicht laufen !

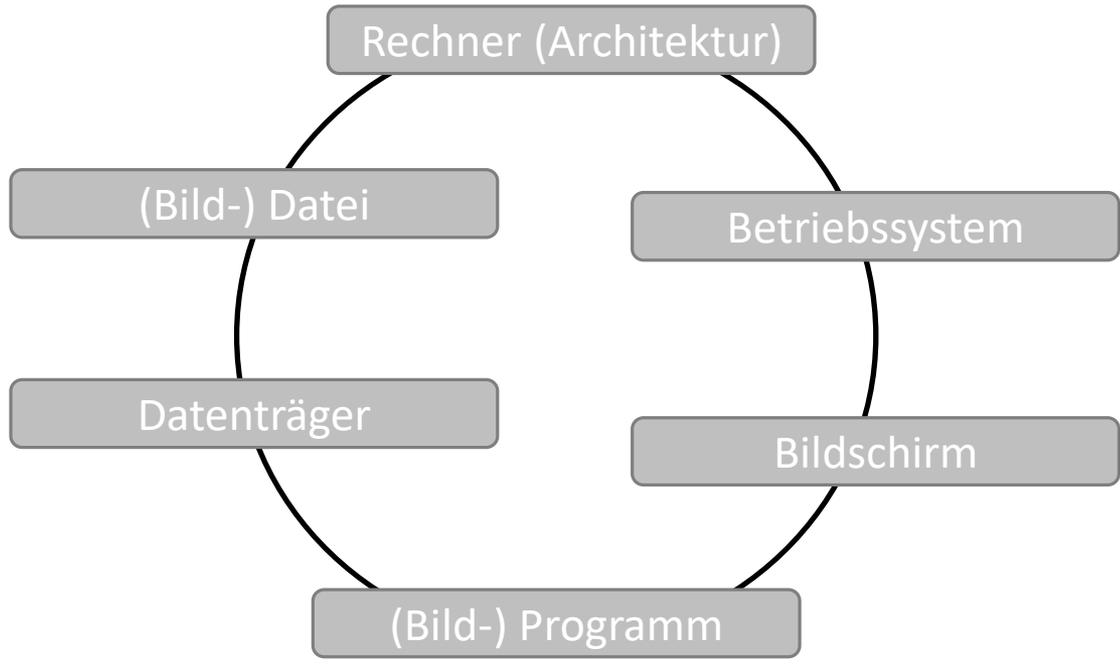
Und hinter diesem Problem liegt noch ein weiteres ...

Rechnerarchitektur <- -> Betriebssystem

Ohne passende Rechnerarchitektur wird das passende Betriebssystem nicht laufen !



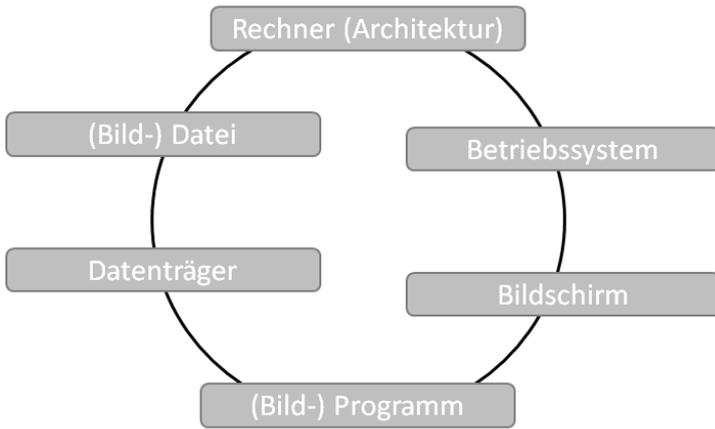
```
100101101100011001010111011000100110  
110010110110001100101011101100010011  
011001011011000110010101110110001001  
101100101101010001100101011101100010  
011010010110110001100101011101100010  
011011001011001000110010101110110001  
100101101100011001010111011000100110  
110010110110001001010111011000110010  
110110001100101011101100010011011001  
0110110001100101010110001100101101  
100011001010111011000100110110010110
```



Die Rückgewinnung von Inhalten klappt nur, wenn **ALLE** Komponenten zusammenarbeiten !

All diese Komponenten sind in ständigem Wandel begriffen

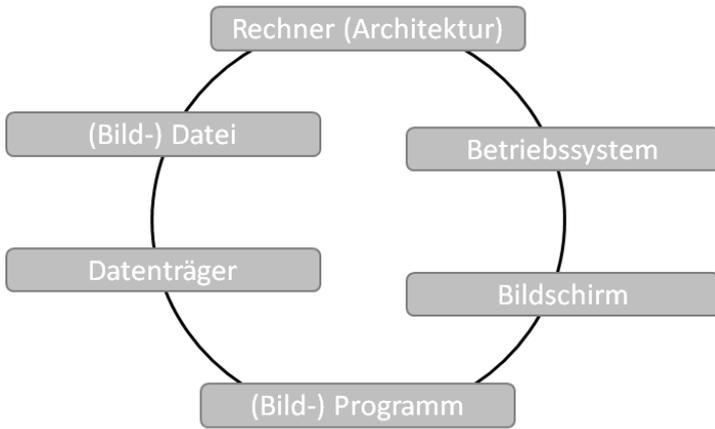
```
100101101100011001010111011000100110  
110010110110001100101011101100010011  
011001011011000110010101110110001001  
101100101101010001100101011101100010  
011010010110110001100101011101100010  
011011001011001000110010101110110001  
100101101100011001010111011000100110  
110010110110001001010111011000110010  
110110001100101011101100010011011001  
011011000110010101010110001100101101  
100011001010111011000100110110010110
```



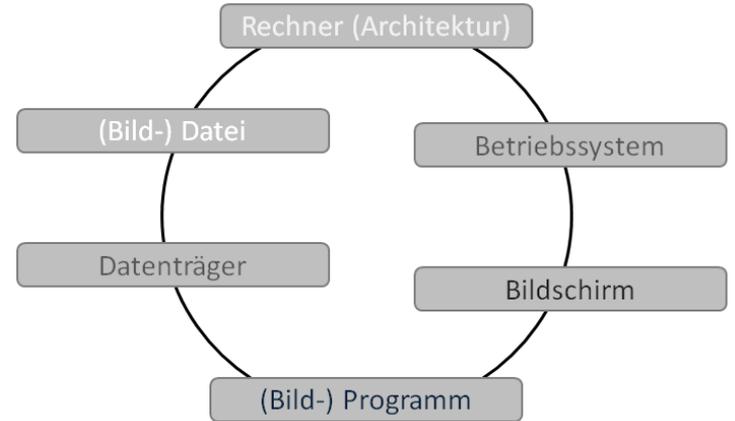
Die Umgebung bei Erstellung der Daten ...

All diese Komponenten sind in ständigem Wandel begriffen

```
100101101100011001010111011000100110
110010110110001100101011101100010011
011001011011000110010101110110001001
101100101101010001100101011101100010
011010010110110001100101011101100010
011011001011001000110010101110110001
100101101100011001010111011000100110
110010110110001001010111011000110010
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
```



```
100101101100011001010111011000100110
110010110110001100101011101100010011
011001011011000110010101110110001001
101100101101010001100101011101100010
011010010110110001100101011101100010
011011001011001000110010101110110001
100101101100011001010111011000100110
110010110110001001010111011000110010
110110001100101011101100010011011001
011011000110010101010110001100101101
100011001010111011000100110110010110
```



Die Umgebung bei Erstellung der Daten ... ist nicht die Umgebung bei Nutzung der Daten

Gedankenexperiment ... Angenommen jemand wäre es gelungen eine Datei auf einem haltbaren Medium über 100 Jahre zu bewahren ...

- 
1. Möglicherweise wird es für das Medium kein Lesegerät mehr geben
 2. Möglicherweise wird es für das Dateiformat kein passendes Programm mehr geben
 3. Möglicherweise wird es für das Programm kein Betriebssystem mehr geben

... ..



1. Möglicherweise wird es für das Medium kein Lesegerät mehr geben

1. Möglicherweise wird es für das Medium kein Lesegerät mehr geben



! Vorbeugen

Kopieren von Datenträger (Generation X) auf einen Datenträger der (Generation Y)

Das Kopieren muss rechtzeitig erfolgen, bevor die Datenträger und Lesegeräte „verschwinden“. Deshalb muss man den „Markt“ der Datenträger ständig im Auge behalten und braucht dafür festgelegte Verantwortliche.



2. Möglicherweise wird es für das Dateiformat kein passendes Programm mehr geben

2. Möglicherweise wird es für das Dateiformat kein passendes Programm mehr geben



! Vorbeugen

Ein Dateiformat wählen,

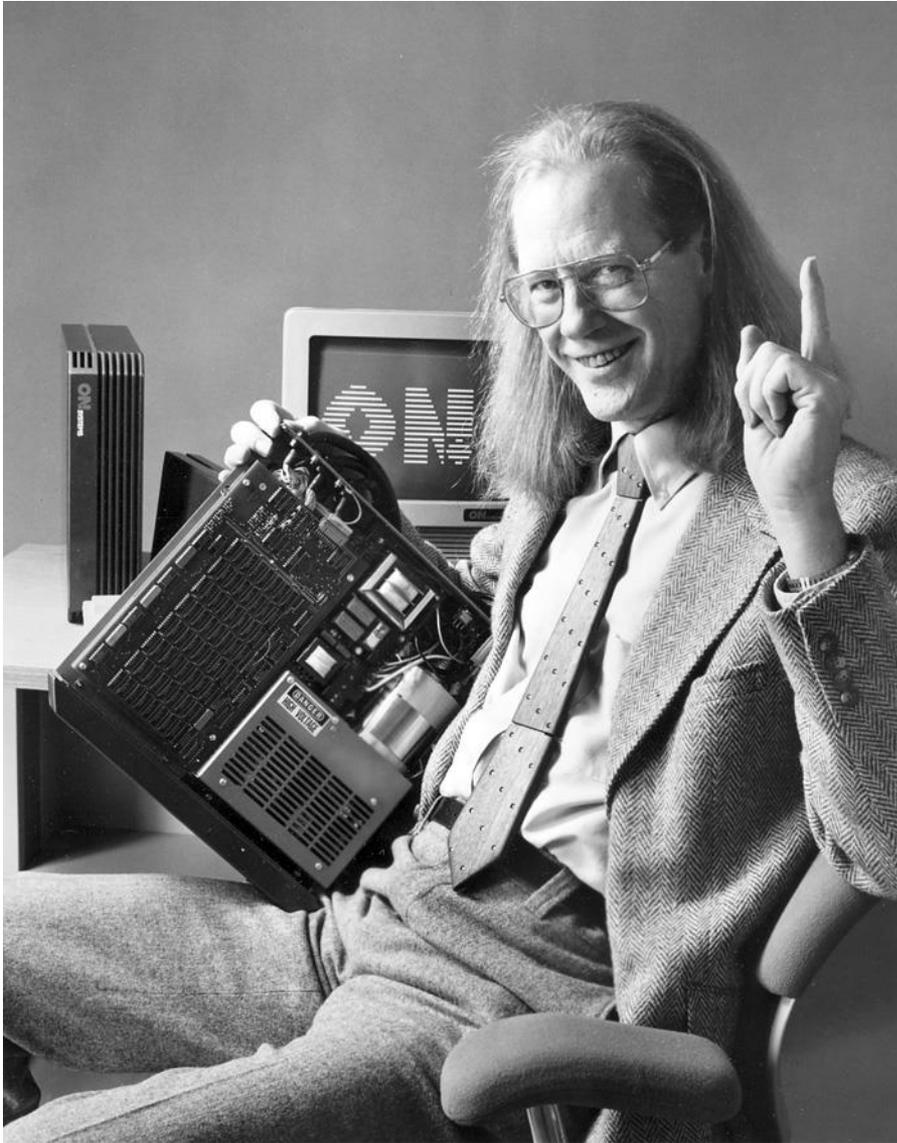
- das aktuell weit verbreitet ist
- für das es aktuell viele Programme gibt
- das offengelegt ist

Das „Dateiformat der Wahl“ kann sich in 100 oder mehr Jahren ändern. Deshalb muss man den „Markt“ der Programme ständig im Auge behalten und braucht dafür festgelegte Verantwortliche.



3. Möglicherweise wird es für das Programm kein Betriebssystem mehr geben

3. Möglicherweise wird es für das Programm kein Betriebssystem mehr geben



Das ist die Stunde der Nerds !

Die „nerds“ können passende Programme entwickeln, denn

- die Dateien sind vollständig vorhanden (weil ständig kopiert wurde)
- das Dateiformat ist offen definiert (weil nur solches verwendet wurde)
- sie kennen das aktuelle Betriebssystem.



Verantwortlichen definieren ...

. Von ihm Speichermedien beobachten lassen ...

.. Wenn nötig kopieren lassen

Verantwortlichen definieren ...

. Von ihm Dateiformate beobachten lassen ...

.. Wenn nötig umformatieren lassen



Ist das alles, was Museen unternehmen sollten?



Die wohl wichtigste Frage ist:

WAS soll überhaupt für lange Zeit aufbewahrt werden

- Es macht z.B. wenig Sinn alle Versionen aller Objektfotos aufzubewahren
- Es macht dagegen viel Sinn Mindestqualitäten für die Übernahme ins Archiv zu definieren
- Man sollte nur überprüfte Dateien (heißt „TIF“ oder ist „TIF“?) ins Archiv übernehmen

Die Entscheidung, was ins Archiv übernommen werden soll, ist festzuschreiben und von Zeit zu Zeit zu überprüfen.



Ebenfalls wichtig ...

Es ist gut, mit einer Datei, die man aufbewahrt auch **alle zugehörigen Informationen** zu bewahren.

- *inhaltliche Informationen*
(Bsp: Bei Foto eine Inhaltsbeschreibung, ...)
- *rechtliche Informationen*
(Bsp: Rechte am abgebildeten Objekt und an Abbildung)
- *technischen Informationen über die Datei*
(Das ermöglicht u.U. eine automatische Überprüfung der Datei auf Vollständigkeit und Lesbarkeit in Intervallen)

Es ist zu entscheiden, welche Informationen mit den jeweiligen Dateien gespeichert werden und wie sichergestellt wird, dass Datei und Information später zueinander geordnet werden können. Die Entscheidung ist von Zeit zu Zeit zu überprüfen.

Was also ist zu tun?



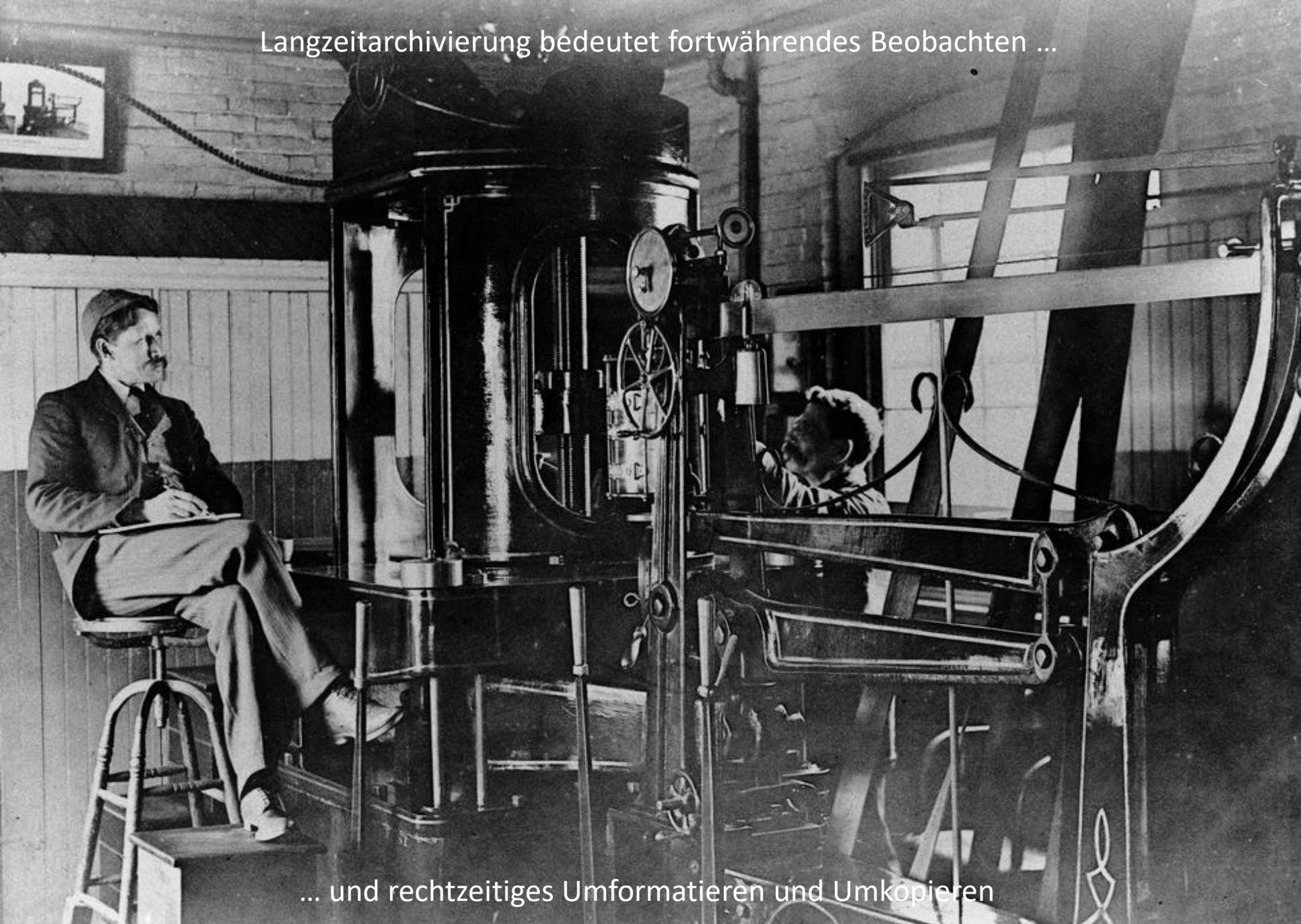


Entscheidungen treffen !



- 1. WAS soll aufbewahrt werden ?**
2. In welchen Intervallen wird geprüft ob die WAS-Entscheidung zu revidieren ist ?
- 3. Wie werden die Daten im Archiv strukturiert ?**
4. In welchen Intervallen wird die Struktur kritisch analysiert ?
- 5. WER überwacht das Archiv ?**
6. In welchen Intervallen werden (auf welche Art?) Kontrollen durchgeführt ?
- 7. Auf welchem Datenträger speichern ?**
8. In welchen Intervallen wird geprüft ob umkopiert werden muss ?
- 9. In welchem Dateiformat speichern ?**
10. In welchen Intervallen wird geprüft ob umformatiert werden muss ?

Langzeitarchivierung bedeutet fortwährendes Beobachten ...



... und rechtzeitiges Umformatieren und Umkopieren

Einige praktische Empfehlungen ... (nur heute gültig)

- Möglichst wenig aufheben (keine Kopien)
- Nur Hochwertiges aufheben (bei Bildern: nur hochaufgelöste)
- Sich für möglichst wenige Dateiformate entscheiden (bei Bild: Tiff, bei Text: PDF-A)
- Sich für leicht ersetzbare und schnelle Speichermedien entscheiden (z.B. Festplatten)
- Die Metadaten zum Objekt in eine XML-Datei schreiben
- Für jede Datei einen Ordner anlegen. In diesem Ordner Datei und Metadaten speichern
- Intervall für Medienprüfung (Obsolescenz), ½ Jahr
- Intervall für Formatprüfung (Obsolescenz), ½ Jahr
- Intervall für (stichprobenartige) Lesbarkeit der Dateien: 1 Jahr
- Kopien der Daten an zwei verschiedenen Orten sichern

... aber :

Das alles ist streng genommen noch kein wirkliches Langzeitarchiv (es fehlen vor allem die „Nerds“ (s.o.) und die Technik Überprüfungen, Kontrollen, Datenträgerersetzungen und Umformatierungen automatisch ablaufen zu lassen) ...



Langzeitarchive dieser Art können sich nur große Einrichtungen leisten.

Man kann – und sollte – sich Partner suchen, von denen man annehmen kann, dass Sie auch in 100 Jahren noch existieren und in deren Integrität (Umgang mit den Daten, Weitergabe nach außen, ...) man volles Vertrauen hat.

Kann man sich diese Partner nicht leisten oder für sich keinen solchen Partner finden, dann sollte man den gerade genannten Empfehlungen folgen und die gerade erwähnten Entscheidungen treffen und schriftlich fixieren („Strategie“, „Plan“, ...)



Ignorieren ist keine Lösung !

Vielen Dank

Credits

Georgie Pauwels, art understanding, <https://flic.kr/p/oENVog>
Emanuel Debrand, black and white.jpg, <https://flic.kr/p/rN9KrK>
Erik Forsberg, Nikon, <https://flic.kr/p/iXbgdf>
Krystian Olszanski, Fork in disguise, <https://flic.kr/p/8WbD4v>
Margo Toth, telefonbetyár, <https://flic.kr/p/qJ4Vjr>
Gauthier Delacroix, Generations, <https://flic.kr/p/J2uJzB>
Jason Trim, Room full of Jasons, <https://flic.kr/p/6FS58o>
Carmen Escobar Barrio, Untitled, <https://flic.kr/p/aWpqxZ>
Dave Conner, Inverness-shire Constabulary HMI Inspection 1967, <https://flic.kr/p/nKZHVM>
Manhhai, 1961 East Berlin Street with portraits of Communist Leaders, <https://flic.kr/p/qvfakL>
Nick dewolfe photo archive, 28 the ON! System, <https://flic.kr/p/huYRK>
Pasco Tomic, Man with two shadows, <https://flic.kr/p/hXQbMo>
Archives of Ontario, Demonstration of information technology, <https://flic.kr/p/oBSjDL>
William Warby, Julius Caesar, <https://flic.kr/p/ib3YS1>
Alexandre Dulaunoy, Pick one, <https://flic.kr/p/Lyfhgg>
Björn Rixmann, Something went wrong, <https://flic.kr/p/8FLkf>
Seier+seier, is he still looking at me, <https://flic.kr/p/hRNfvz>
Maria Arellano-Ongpin, Conversatione, <https://flic.kr/p/7DMgwT>
Kheel Center, Testing engineer at work, <https://flic.kr/p/93vfvG>
Nicolas Alejandro, Pointing at me, <https://flic.kr/p/dgfjLV>
Biblio Archives, Women worker replaces, <https://flic.kr/p/88qSeT>