imaging & media lab

Long term preservation of digital content in analog supports

PD Dr. Lukas Rosenthaler

1st International Workshop on Standards and Technologies in Multimedia Archives and Records (STAR)

26th-27th April 2010

ecma INTERNATIONAL

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

550 JAHRE WISSEN BEWEGT UNS

UNI BASEL

The digital Universe

- 2008: Amount of data existing:
  ~3‘892‘179‘868‘480‘350‘000‘000 Bits
  ~4‘865‘220‘000 Terabytes
  IDC Digital Universe Whitepaper, 2008

- Every minute, 13 hours of video are beeing uploaded to YouTube
  Doug Garland, Google, 2008

- How large us your “digital shadow?
  (mine: ca. 1 TB)

The Digital Dark Ages

digital information lasts forever

—

or five years,
whichever comes first!

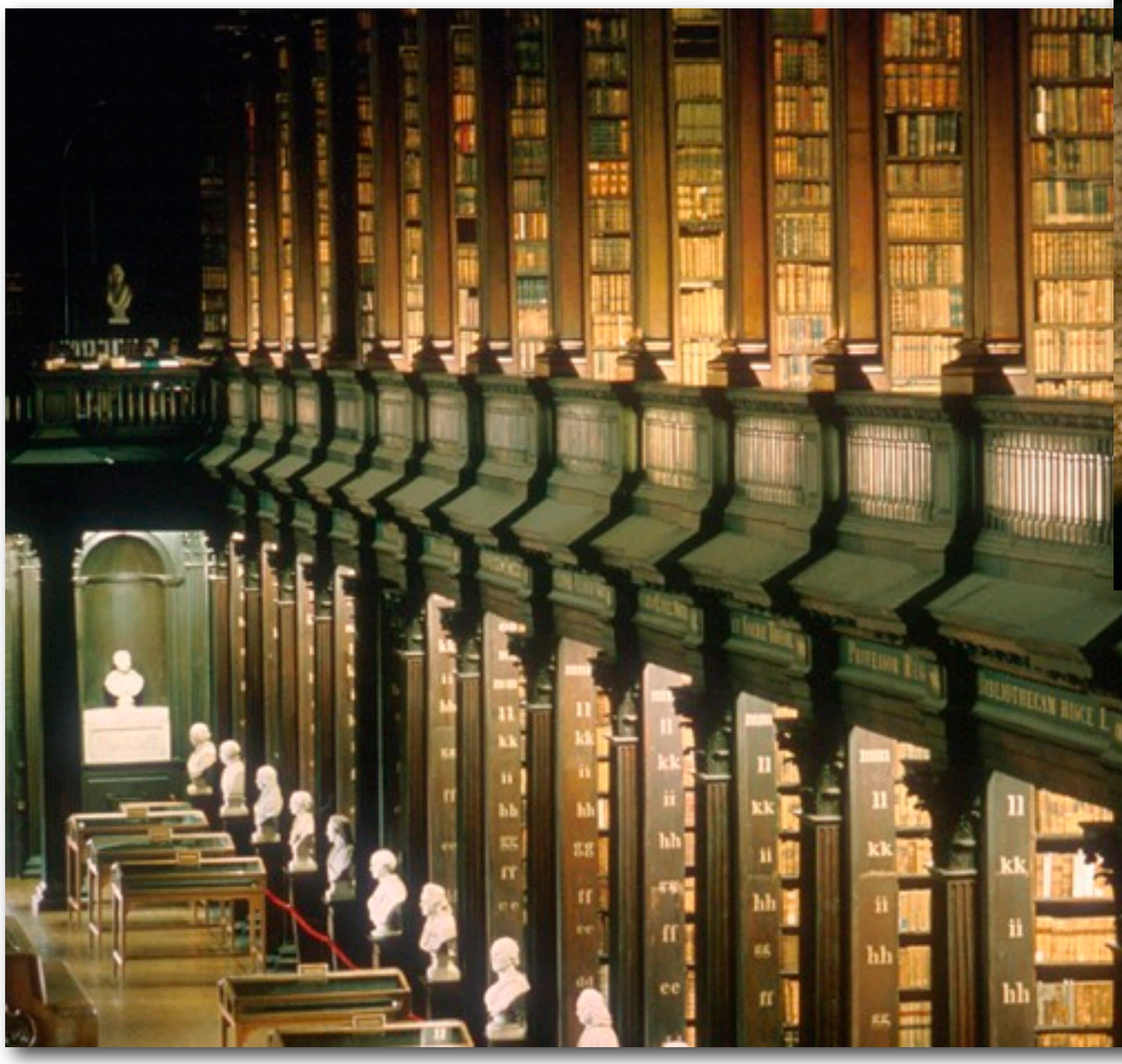
Jeff Rothenberg (1999)

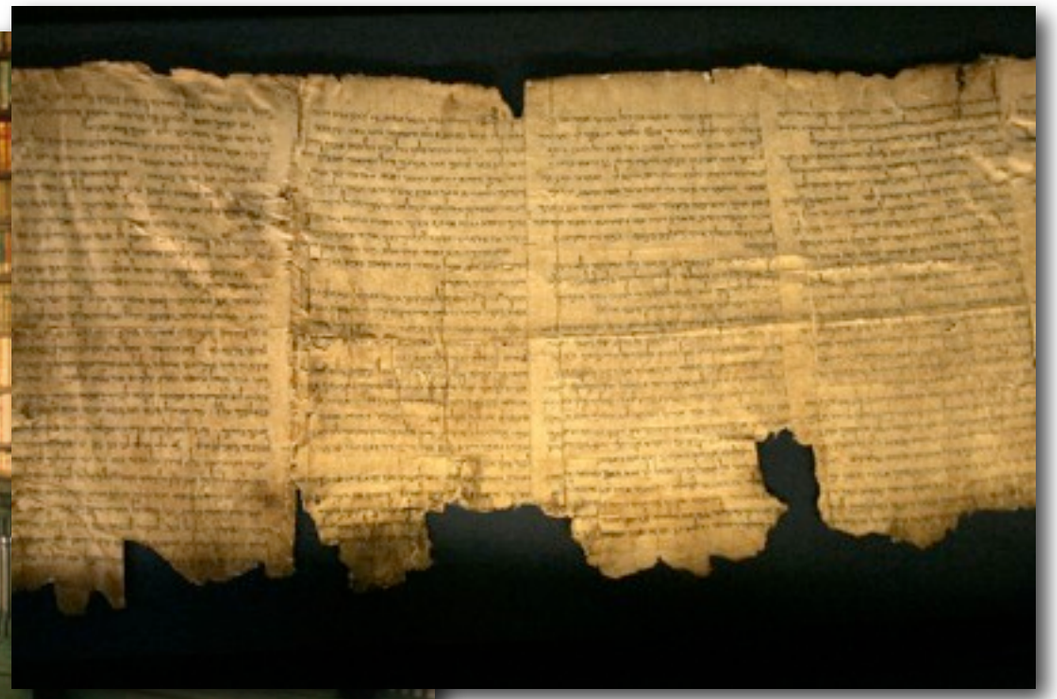
2010: It's still a problem...

- Why is long-term preservation of digital data still a problem?
  - **Obsolescence of technology**
    within 2-5 years storage systems are replaced by newer generation with limited compatibility
  - **Longevity of media**
    It's even worth than You think (e.g. CD-R: 10% loss after 10 years)
  - **Dataformats, Software-versions, DRM, undisclosed proprietary technologies, etc.**

We should know it better!

We should know it better!

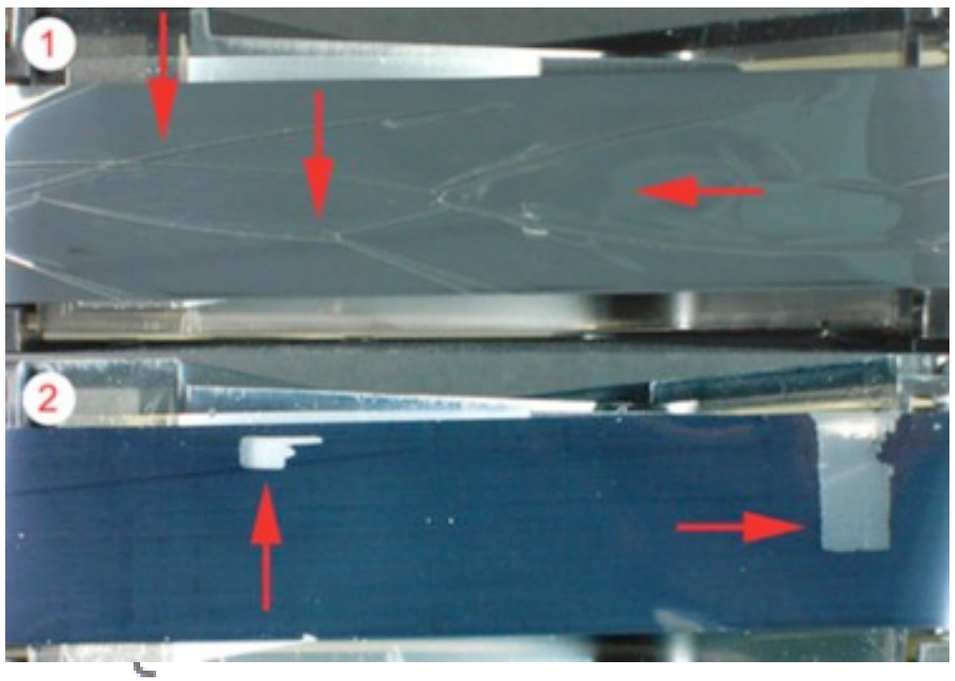




Obsolescence...

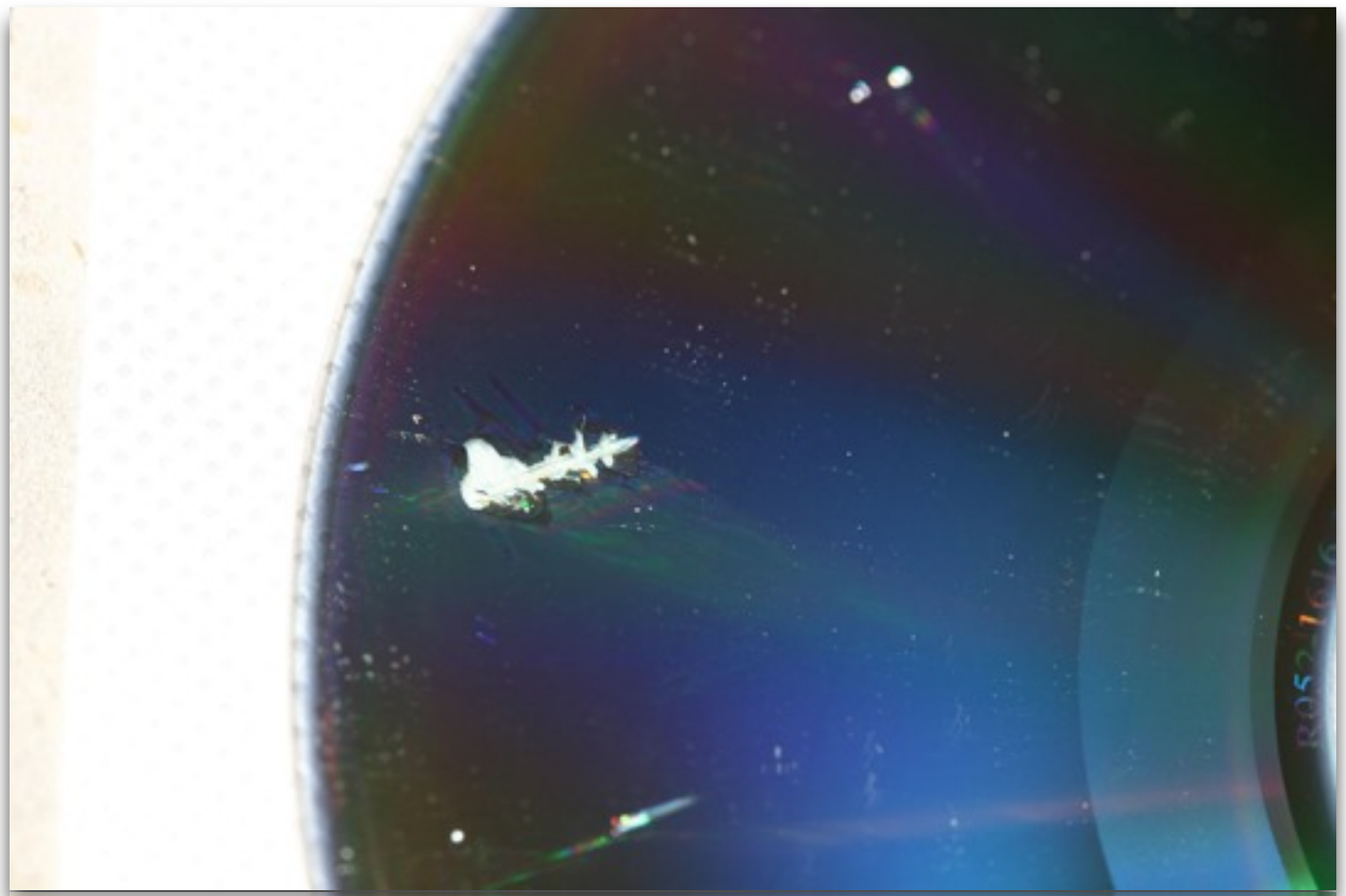
Longevity of storage media

HD after headcrash



Damaged magnetic tape

Damaged CD-R



Longevity of storage media

- **Magnetic Tape**: 5-30 years
- **CD-R**: 1-50 years

- **Harddisk**
  - **spinning**: 3-5 years (worn out, failure rate goes up)
  - **on-shelf**: 5-8 years (sticky bearings)

What are the common options?

- **Do nothing!** ➳ "digital archeology"
- **Computermuseum** ➳ archive media, machines, documentation and *service technician* (last item poses most problems...)
- **Emulation** ➳ Simulate old computer on modern machine (but how can a ½" Tape be read on a modern computer?)
- **Migration** ➳ copy, copy, copy,.....
- **Permanent medium** ➳ carve it in stone...

One step back....

- What do we really **want** to archive?

- What do we really **do** archive?

One step back....

- What do we really **want** to archive?

"Digital objects", which can be rendered in way that they **are accessible to the human senses** or can be **processed by an information processing machine** aka "computer"

- What do we really **do** archive?

One step back....

- What do we really **want** to archive?

"Digital objects", which can be rendered in way that they **are accessible to the human senses** or can be **processed by an information processing machine** aka "computer"

- What do we really **do** archive?

A physical object like a CD-R, a Magtape, HD,...

What is "digital data"?

- Any information recorded in a code based on a limited set of symbols

$$S = \{s_1, s_2, s_3, \ldots, s_n\}; n \geq 2$$

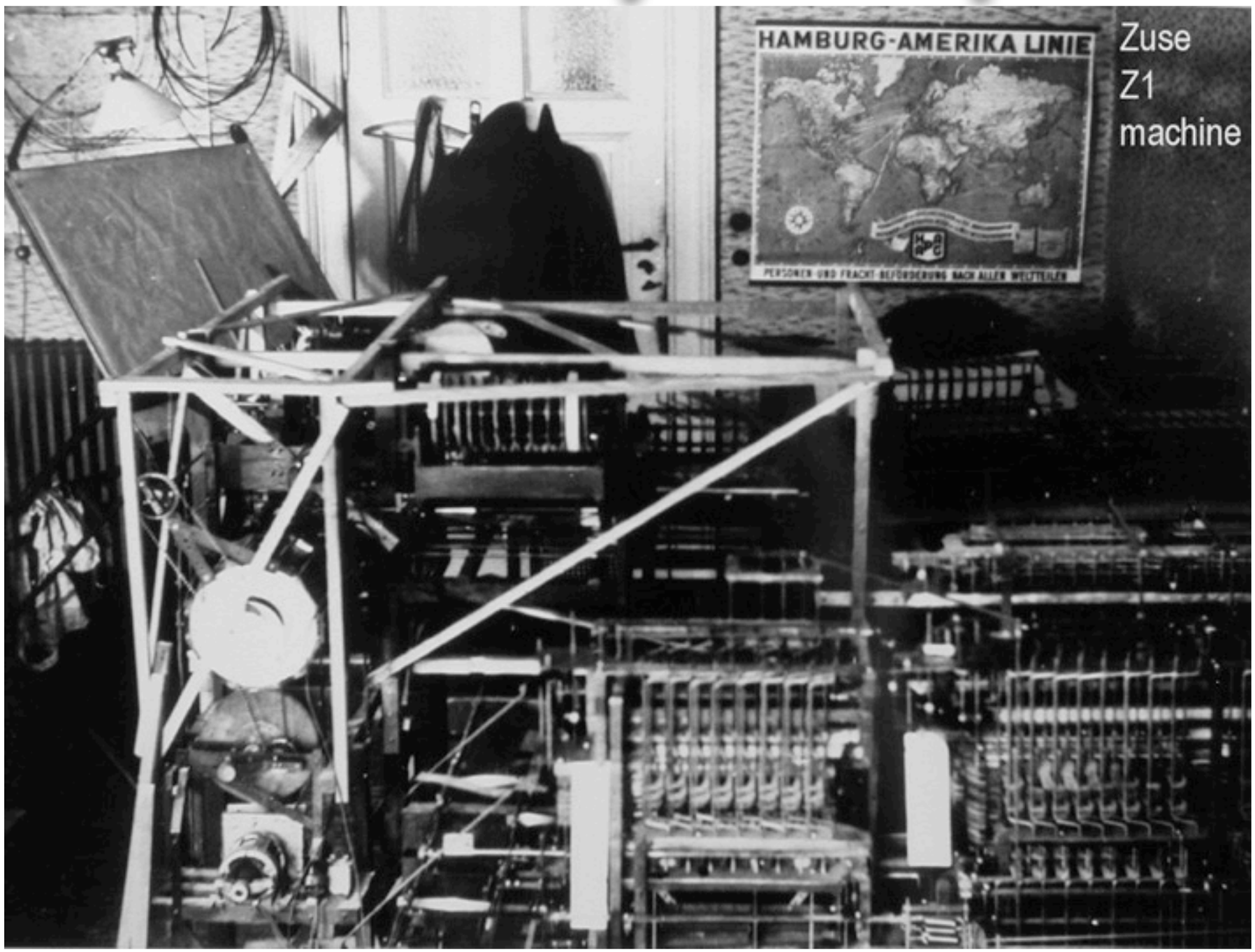
- Most of it converted from an analog signal as a function of location and/or time:

$$F(x,y,z,t) \rightarrow s_i, s_k, \ldots \in S$$

"Binary" Data?

- A special case of "digital data"
  - S={0,1}
  - S={TRUE, FALSE}
  - S={+5V,-5V}
  - S={↑, ↓}
- Most simple set of symbols
- Easy to implement

First binary computer



Konrad Zuse's Z1 (1936)

First binary computer (electro-mechanical) built in the living room of his parents

Problem: binary data is not self-explanatory!!

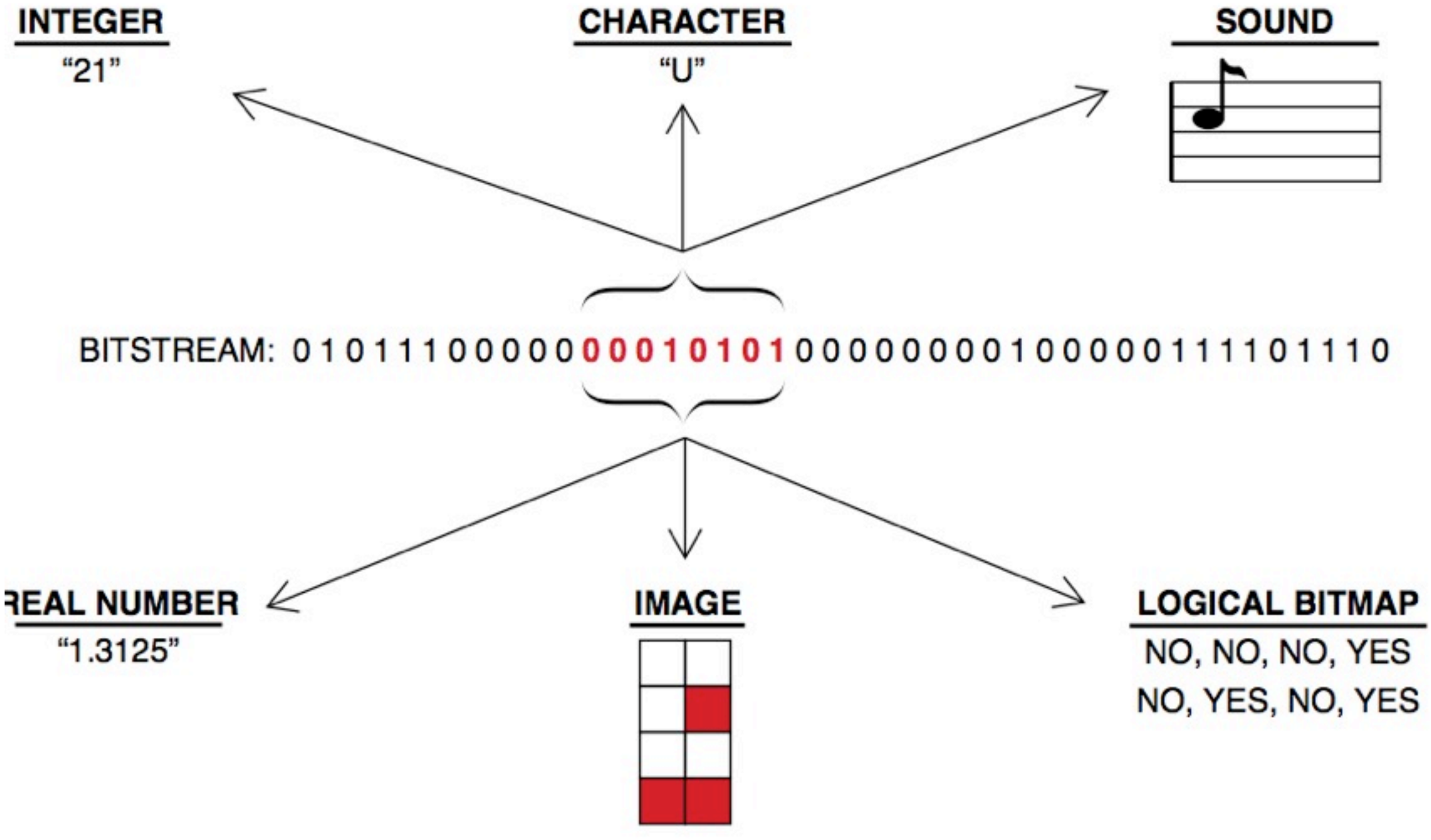
➳ rendering needs metadata

(➳*file format = **knowledge** about the meaning of the bits*)

Problem: binary data is not self-explanatory!!

➳ rendering needs metadata

(➳*file format = **knowledge** about the meaning of the bits*)



Dis-/Advantages of binary data

- most simple encoding
- natural code for information processing machines
- lossless copy possible!
  - there is no "digital original", only "clones"
- not for humans:

Dis-/Advantages of binary data

Can You read this?

```
01001100010101010100101101000001010100110010000001010010
01001111010100110100010101001110010101000100100001000001
0100110001000101010010
```

Dis-/Advantages of binary data

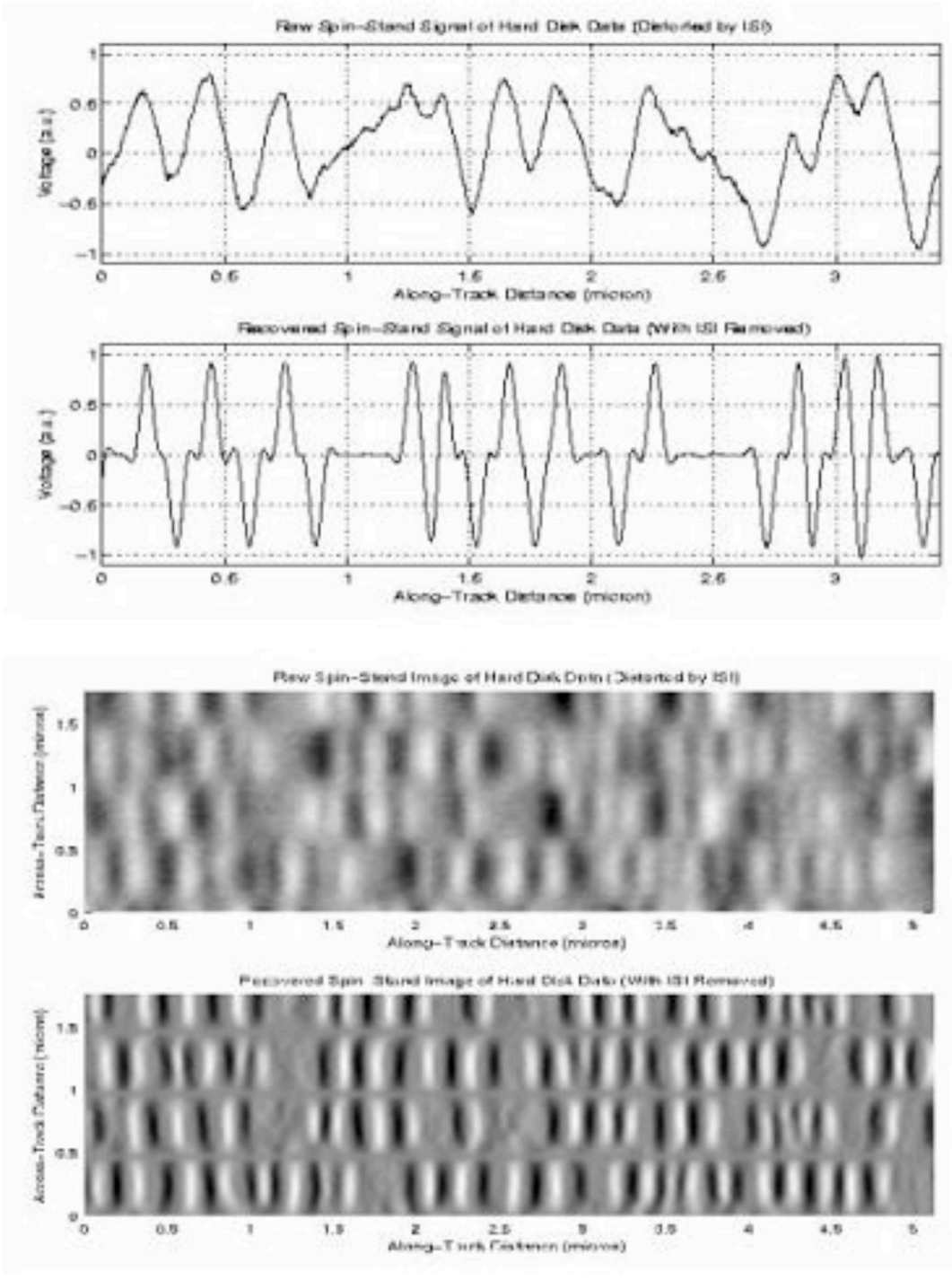
Can You read this?

```
01001100010101010100101101000001010100110010000001010010
01001111010100110100010101001110010101000100100001000001
0100110001000101010010
```

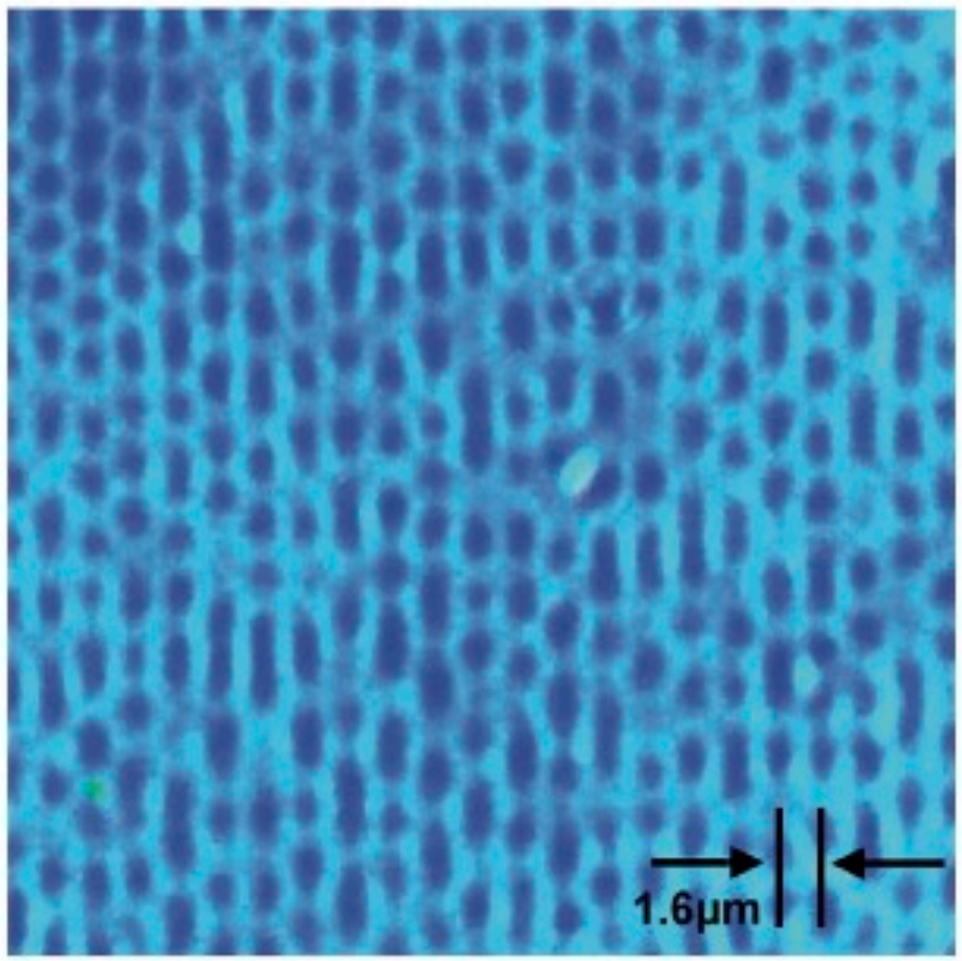
LUKAS ROSENTHALER

But: digital data do not exist!

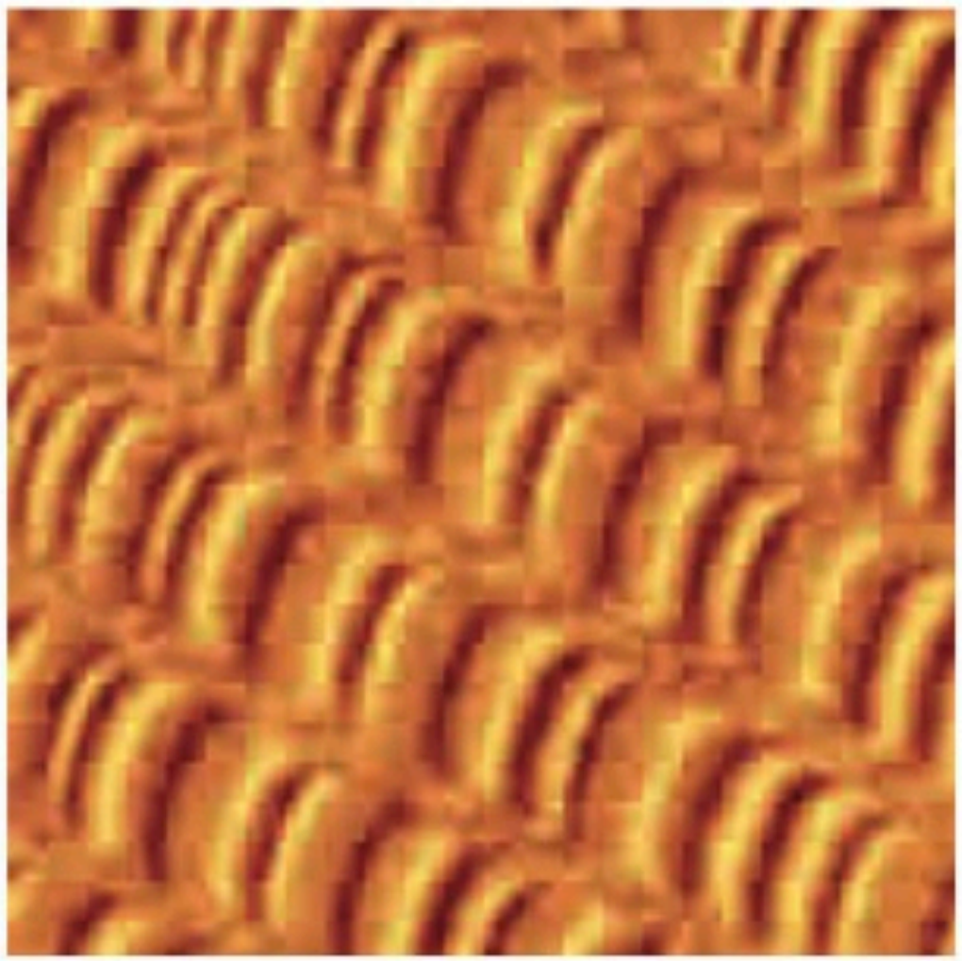
- "digital data" is an **"immateriel, logical concept"**
- all material physical recordings are **always analogue**
- to go to "digital data", a **decision process** is required (e.g. thresholding)
- The decision process is **error-prone**



But: digital data do not exist!



CD recording pattern



HD recording pattern

What is "digital" storage really?

Analogue marks on a physical medium

- these marks have to be interpreted correctly (recognized) in order to create the immaterial digital information
  - **material**: analogue marks
  - **immaterial**: "digital" information

Whats the objective?

- we want to retrieve the archived information at some point ***far*** (whatever this means for You!) **in the future**

- 2 levels of preservation required

1. **Keep the bits**
2. **Keep the knowledge about the meaning of the bits**

Keep the bits...

- Migration:
  - Periodically (every 3-5 years) copy the bits to new media
  - proove (with checksum, bitwise comparison etc.) that the copy is identical to the "original"
- "Eternal" media
  - use a long-lasting medium ("eternal" my be difficult to achieve)
  - use a recording method which is independent of a specialized technology (no technological obsolescence)

Keep the **meaning** of the bits

- use open, well documented, widespread and simple formats
- Emulation? UVC?

Keeping the bits: sequel

- we do both at the IML....
  - **DISTARNET**: automated migration using a distributed, self-organizing P2P-architecture
  - **PEVIAR: PErmanent VIsual ARchive**
    - use a **well known**, **long lasting** medium:
      Photographic film, Microfilm
    - use a **technology-independent**, **simple** recording method:
      ~~optical~~ visual recording using a 2D ~~bar~~ dot-code

Keeping the bits: sequel

- we do both at the IML....
  - **DISTARNET**: automated migration using a distributed, self-organizing P2P-architecture
  - **PEVIAR: PErmanent VIsual ARchive**
    - use a **well known**, **long lasting** medium:
      Photographic film, Microfilm
    - use a **technology-independent**, **simple** recording method:
      ~~optical~~ visual recording using a 2D ~~bar~~ dot-code

Nothing new...

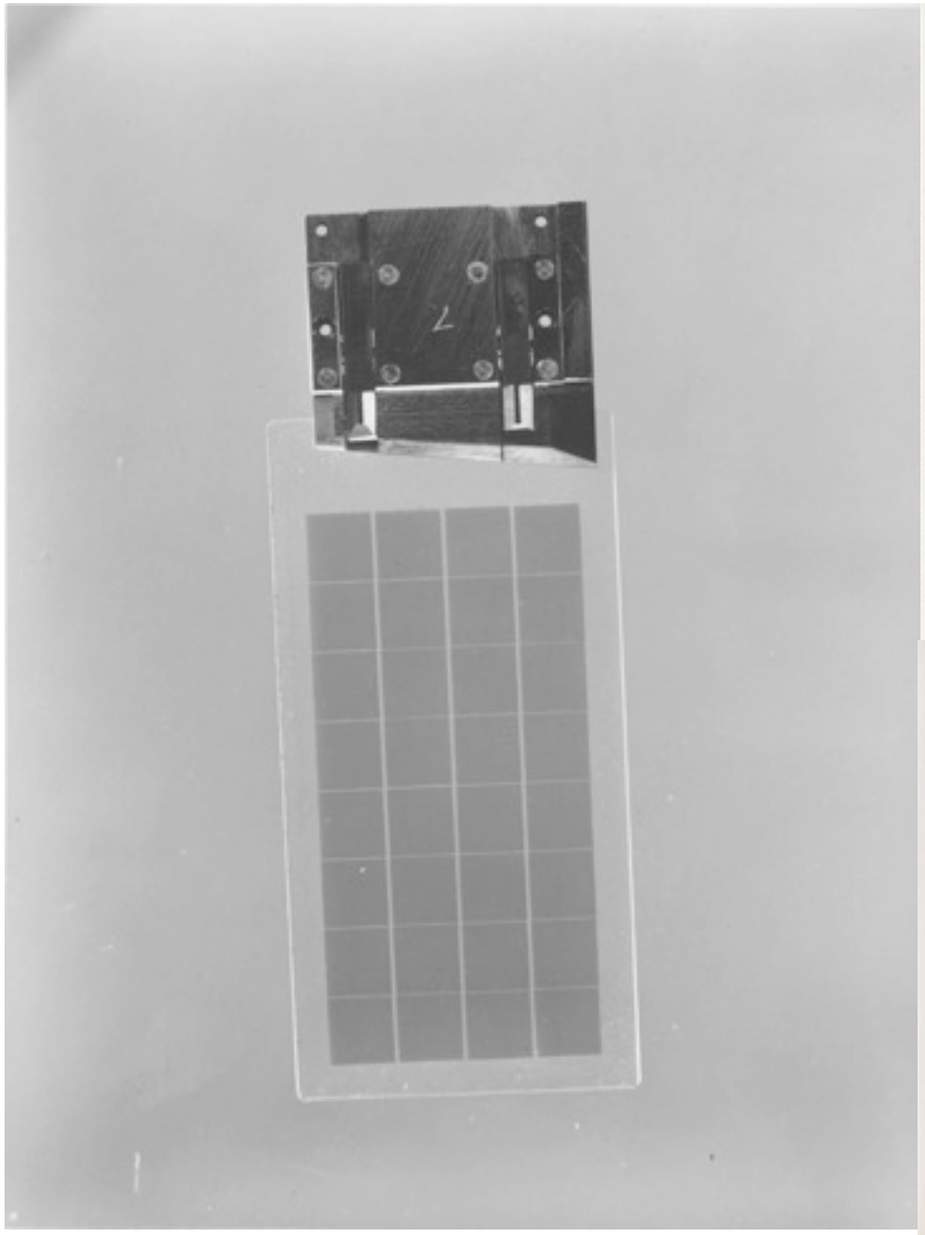
- **1967: IBM Photo-Digital Storage System**
  - 1 Terabit
  - Recorded on photographic film using an electron-beam writer
  - reading back using a microscope and a TV-camera

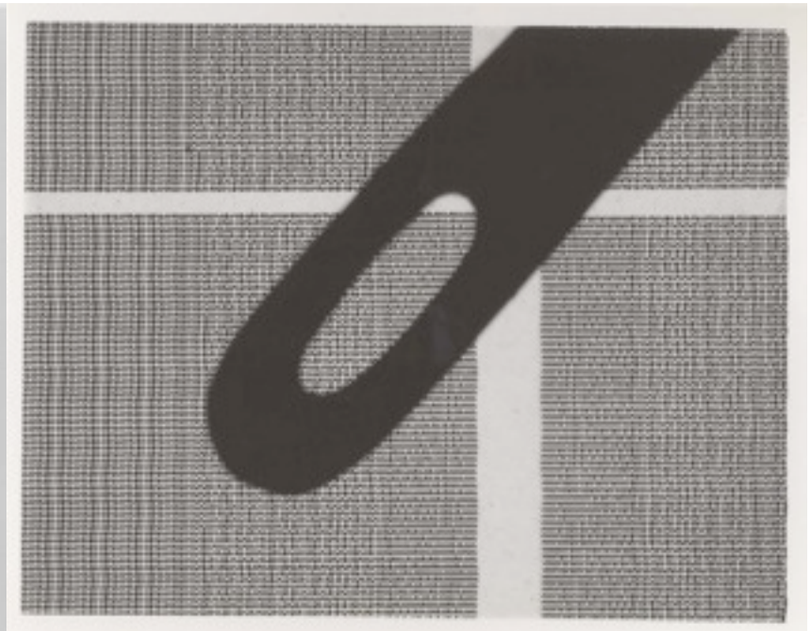
Nothing new...

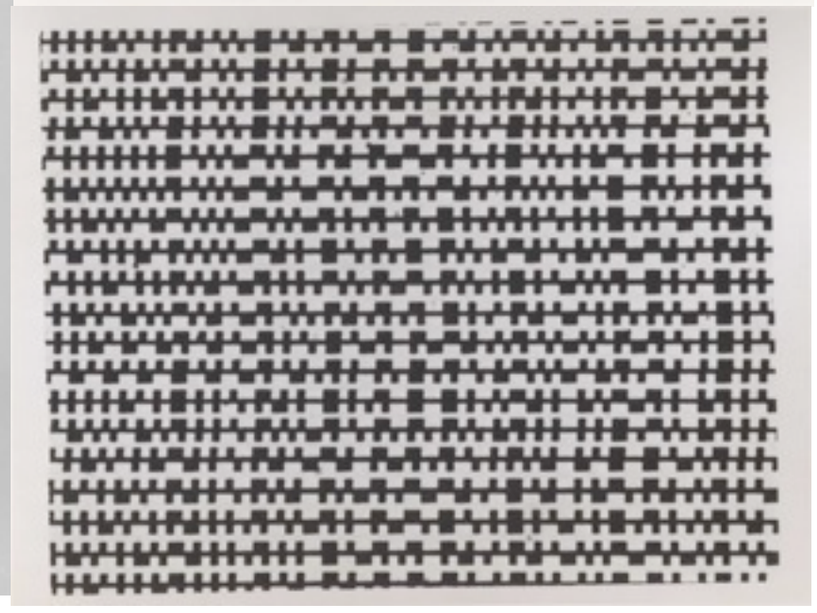
- **1967: IBM Photo-Digital Storage System**

Nothing new...

- **1967: IBM Photo-Digital Storage System**







PEVIAR

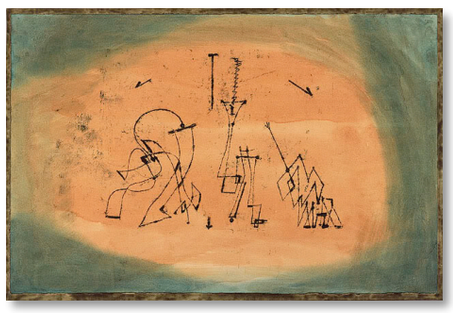
- use "normal" film or microfilm (both b/w and color) as medium
- record bitpatterns using off-the shelf film recorder (e.g. ARRI Laser, ArchiveLaser, ...)

- Use **any** scanner or dig. camera with enough resolution

PEVIAR:Visual encoding

.pdf

.wav



.j2k

encode and
record on film

Any file type can be encoded and recorded as 2D dotcode.
Metadata, deocding information etc. and other visual information
can be record in a human readable fahsion

PEVIAR: Example

UNIVERSITÄT BASEL

recorded by: mikrosave

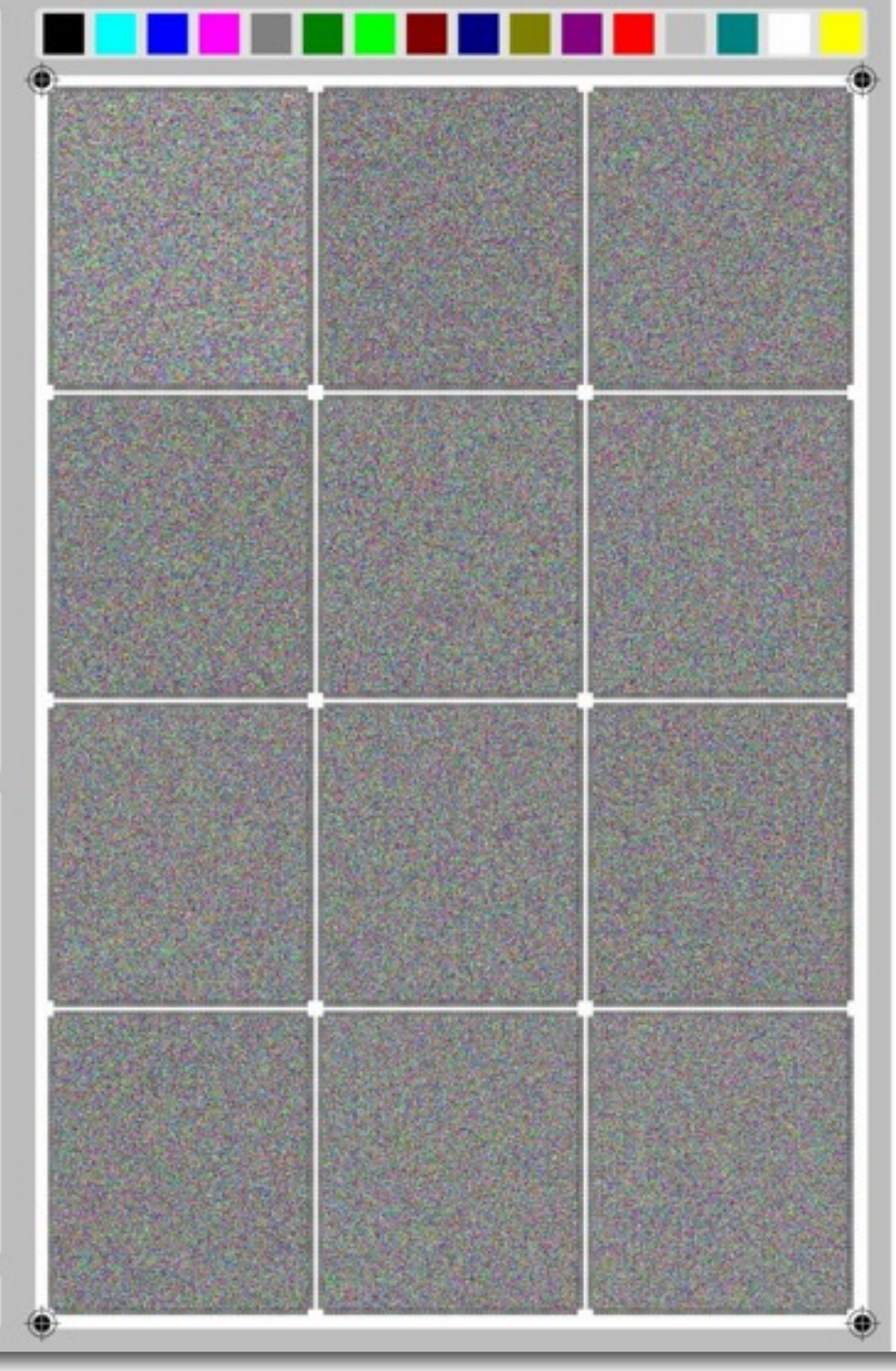
imaging & media lab

FLS_20070220_MKBA_0001

TITLE	Mona Lisa
PUBLISHER	Graf Library Ltd.
CREATOR	Graf, F. Otto
DESCRIPTION	Foto of 'Mona Lisa'
DATE	2006-08-05T12:00:00
TYPE	Image
FORMAT	image/tiff
FORMAT	Colordepth 8 bit
FORMAT	Colorspace Adobe RGB
IDENTIFIER	CHBS_LIB99_AX01
SOURCE	FRIDF_L01_A001
RIGHTS	(c) Graf, F. Otto
RIGHTS	All rights reserved

```
<?xml version="1.0"?>
METS_Profile xmlns="http://www.loc.gov/METS_Profile/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/METS_Profile/
http://www.peviar.ch/mets/mets.profile.v1-2.xsd">
  <URI LOCTYPE="URL">http://www.iml.unibas.ch/Peviar/Mets-Profile</URI>
  <title>Peviar Mets Profile</title>
  <abstract>
  This is the METS Profile used to describe digital objects archived with Peviar.
  </abstract>
  <date>2006-01-01T12:00:00</date>
  <contact>
    <name>N.N.</name>
    <institution>Imaging and Media Lab, University of Basel</institution>
    <address>Bernoullistrasse 32, 4051 Basel</address>
    <phone>+41610000000</phone>
    <email>peviar@iml.unibas.ch</email>
  </contact>
  <registration_info regDate="" regURI=""/>
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>Digitalisierung von "Mona Lisa"</dc:title>
  <dc:creator>Graf, F. Otto</dc:creator>
  <dc:publisher>My Digital Library</dc:publisher>
  <dc:type>image</dc:type>
  <dc:format>image-tiff</dc:format>
  <dc:format>Colordepth: 8 bit</dc:format>
  <dc:format>Color Space: Adobe RGB</dc:format>
  <dc:identifier>CHBS_LIB99_AX01</dc:format>
  <dc:source>FRIDF_L01_A001</dc:source>
  <dc:rights>Copyright Graf, F. Otto</dc:rights>
  </metadata>
</METS_Profile>
```

```
/**
 * PEVIAR - Imaging & Media Lab,
 * University of Basel, Switzerland
 */

/** Example readback program */
int[][] matrix = getImageMatrix();
int[] bitstream = getBitstream(matrix);
int[] decoded = getDecodedBitstream(bitstream);
int[] imageMetaData = getImageMetaData(decoded);
int[] imageData = getImageData(decoded);
TIFF img = new TIFF(imageMetaData);
img.setData(imageData);

/** Method Detail */
getImageMatrix():
  // pseudocode for getImageMatrix()

getBitstream(int[][] matrix):
  // pseudocode for getBitstream()

getDecodedBitstream(int[] bitstream):
  // pseudocode for getDecodedBitstream()

getImageMetaData(int[] decoded):
  // pseudocode for getImageMetaDta()

getImageData(int[] decoded):
  // pseudocde for getImageData()
```



PEVIAR: Example

UNIVERSITÄT BASEL

recorded by: mikrosave

imaging & media lab

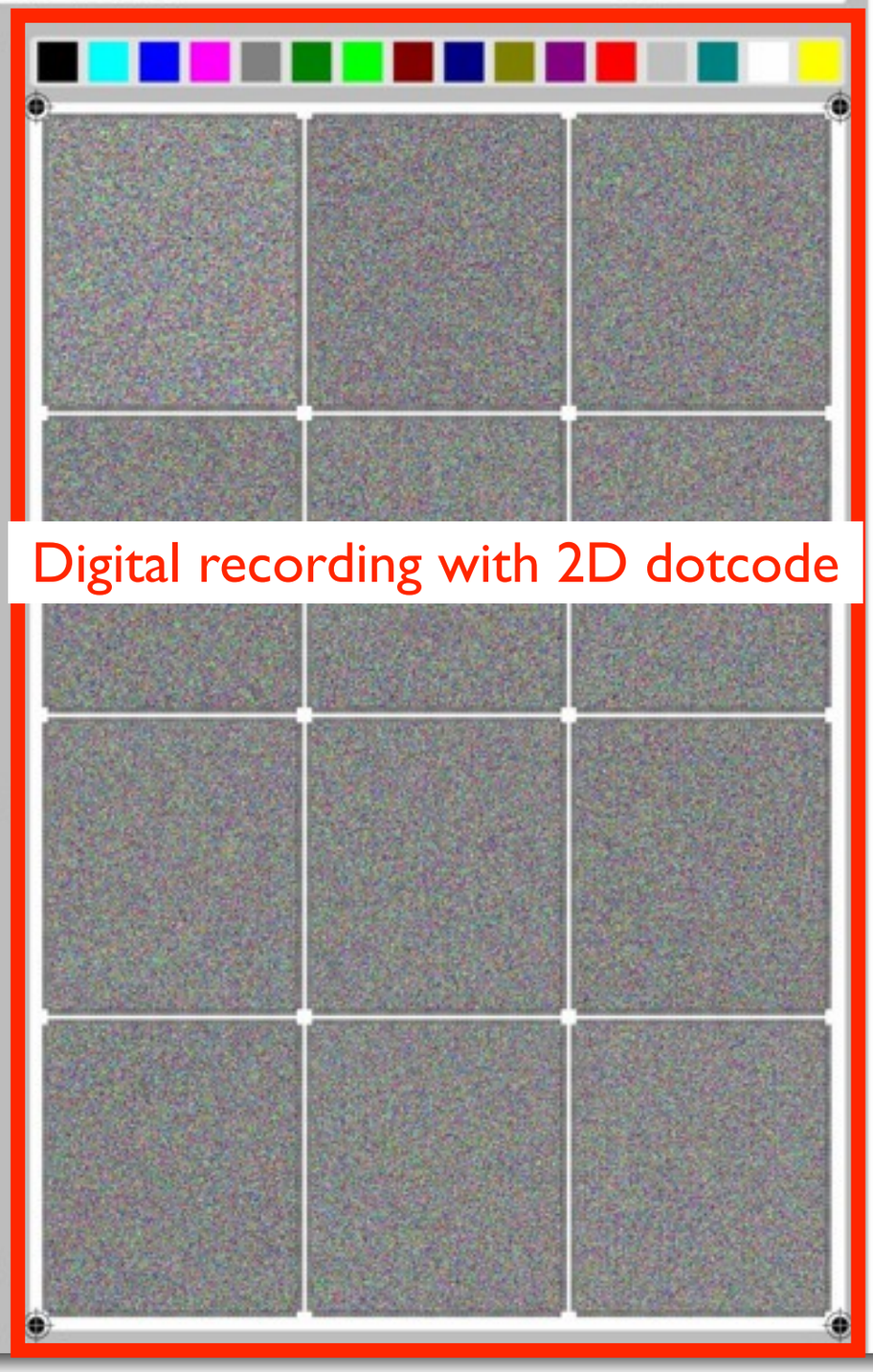
FLS_20070220_MKBA_0001

TITLE	Mona Lisa
PUBLISHER	Graf Library Ltd.
CREATOR	Graf, F. Otto
DESCRIPTION	Foto of 'Mona Lisa'
DATE	2006-08-05T12:00:00
TYPE	Image
FORMAT	image/tiff
FORMAT	Colordepth 8 bit
FORMAT	Colorspace Adobe RGB
IDENTIFIER	CHBS_LIB99_AX01
SOURCE	FRIDF_L01_A001
RIGHTS	(c) Graf, F. Otto
RIGHTS	All rights reserved

```
<?xml version="1.0"?>
METS_Profile xmlns="http://www.loc.gov/METS_Profile/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/METS_Profile/
http://www.peviar.ch/mets/mets.profile.v1-2.xsd">
  <URI LOCTYPE="URL">http://www.iml.unibas.ch/Peviar/Mets-Profile</URI>
  <title>Peviar Mets Profile</title>
  <abstract>
  This is the METS Profile used to describe digital objects archived with Peviar.
  </abstract>
  <date>2006-01-01T12:00:00</date>
  <contact>
    <name>N.N.</name>
    <institution>Imaging and Media Lab, University of Basel</institution>
    <address>Bernoullistrasse 32, 4051 Basel</address>
    <phone>+41610000000</phone>
    <email>peviar@iml.unibas.ch</email>
  </contact>
  <registration_info regDate="" regURI=""/>
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>Digitalisierung von "Mona Lisa"</dc:title>
  <dc:creator>Graf, F. Otto</dc:creator>
  <dc:publisher>My Digital Library</dc:publisher>
  <dc:type>image</dc:type>
  <dc:format>image-tiff</dc:format>
  <dc:format>Colordepth: 8 bit</dc:format>
  <dc:format>Color Space: Adobe RGB</dc:format>
  <dc:identifier>CHBS_LIB99_AX01</dc:format>
  <dc:source>FRIDF_L01_A001</dc:source>
  <dc:rights>Copyright Graf, F. Otto</dc:rights>
  </metadata>
</METS_Profile>
```

```
/**
 * PEVIAR - Imaging & Media Lab,
 * University of Basel, Switzerland
 */

/** Example readback program */
int[][] matrix = getImageMatrix();
int[] bitstream = getBitstream(matrix);
int[] decoded = getDecodedBitstream(bitstream);
int[] imageMetaData = getImageMetaData(decoded);
int[] imageData = getImageData(decoded);
TIFF img = new TIFF(imageMetaData);
img.setData(imageData);

/** Method Detail */
getImageMatrix():
  // pseudocode for getImageMatrix()

getBitstream(int[][] matrix):
  // pseudocode for getBitstream()

getDecodedBitstream(int[] bitstream):
  // pseudocode for getDecodedBitstream()

getImageMetaData(int[] decoded):
  // pseudocode for getImageMetaData()

getImageData(int[] decoded):
  // pseudocode for getImageData()
```

0 mm 1 2 3 4 5 6 7 8



PEVIAR: Example

UNIVERSITÄT BASEL

recorded by: mikrosave

imaging & media lab

FLS_20070220_MKBA_0001

analogue image

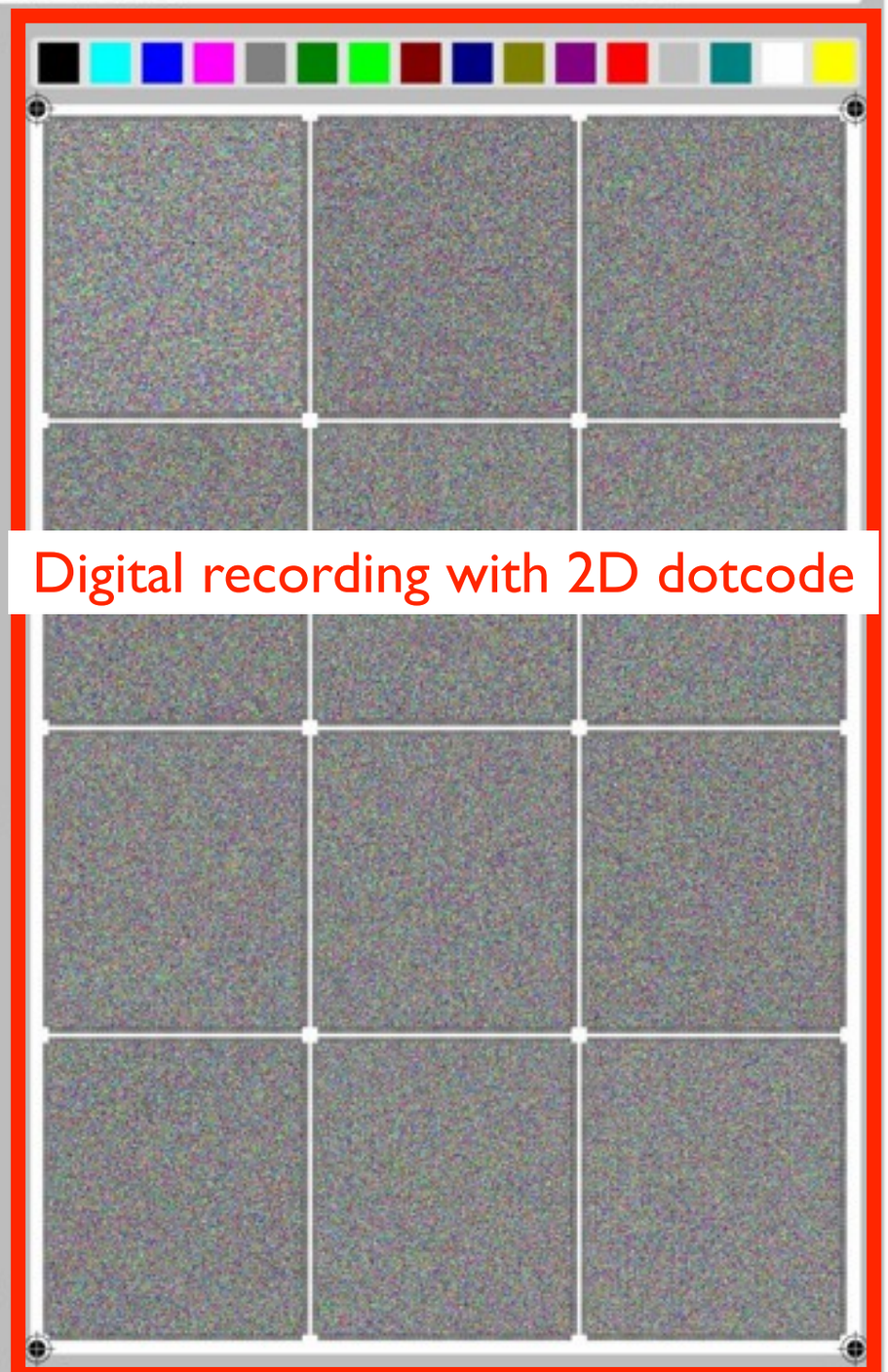
TITLE	Mona Lisa
PUBLISHER	Graf Library Ltd.
CREATOR	Graf, F. Otto
DESCRIPTION	Foto of 'Mona Lisa'
DATE	2006-08-05T12:00:00
TYPE	Image
FORMAT	image/tiff
FORMAT	Colordepth 8 bit
FORMAT	Colorspace Adobe RGB
IDENTIFIER	CHBS_LIB99_AX01
SOURCE	FRIDF_L01_A001
RIGHTS	(c) Graf, F. Otto
RIGHTS	All rights reserved

```
<?xml version="1.0"?>
METS_Profile xmlns="http://www.loc.gov/METS_Profile/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/METS_Profile/
http://www.peviar.ch/mets/mets.profile.v1-2.xsd">
  <URI LOCTYPE="URL">http://www.iml.unibas.ch/Peviar/Mets-Profile</URI>
  <title>Peviar Mets Profile</title>
  <abstract>
  This is the METS Profile used to describe digital objects archived with Peviar.
  </abstract>
  <date>2006-01-01T12:00:00</date>
  <contact>
    <name>N.N.</name>
    <institution>Imaging and Media Lab, University of Basel</institution>
    <address>Bernoullistrasse 32, 4051 Basel</address>
    <phone>+41610000000</phone>
    <email>peviar@iml.unibas.ch</email>
  </contact>
  <registration_info regDate="" regURI=""/>
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>Digitalisierung von "Mona Lisa"</dc:title>
  <dc:creator>Graf, F. Otto</dc:creator>
  <dc:publisher>My Digital Library</dc:publisher>
  <dc:type>image</dc:type>
  <dc:format>image-tiff</dc:format>
  <dc:format>Colordepth: 8 bit</dc:format>
  <dc:format>Color Space: Adobe RGB</dc:format>
  <dc:identifier>CHBS_LIB99_AX01</dc:format>
  <dc:source>FRIDF_L01_A001</dc:source>
  <dc:rights>Copyright Graf, F. Otto</dc:rights>
  </metadata>
</METS_Profile>
```

```
/**
 * PEVIAR - Imaging & Media Lab,
 * University of Basel, Switzerland
 */

/** Example readback program */
int[][] matrix = getImageMatrix();
int[] bitstream = getBitstream(matrix);
int[] decoded = getDecodedBitstream(bitstream);
int[] imageMetaData = getImageMetaData(decoded);
int[] imageData = getImageData(decoded);
TIFF img = new TIFF(imageMetaData);
img.setData(imageData);

/** Method Detail */
getImageMatrix():
  // pseudocode for getImageMatrix()

getBitstream(int[][] matrix):
  // pseudocode for getBitstream()

getDecodedBitstream(int[] bitstream):
  // pseudocode for getDecodedBitstream()

getImageMetaData(int[] decoded):
  // pseudocode for getImageMetaDta()

getImageData(int[] decoded):
  // pseudocode for getImageData()
```

0 mm 1 2 3 4 5 6 7 8

Digital recording with 2D dotcode



PEVIAR: Example

UNIVERSITÄT BASEL

recorded by: mikrosave

imaging & media lab

FLS_20070220_MKBA_0001

analogue image

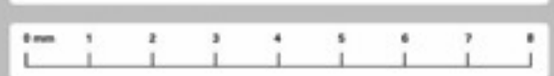
TITLE	Mona Lisa
PUBLISHER	Graf Library Ltd.
CREATOR	Graf, F. Otto
DESCRIPTION	Foto of 'Mona Lisa'
DATE	2006-08-05T12:00:00
TYPE	Image

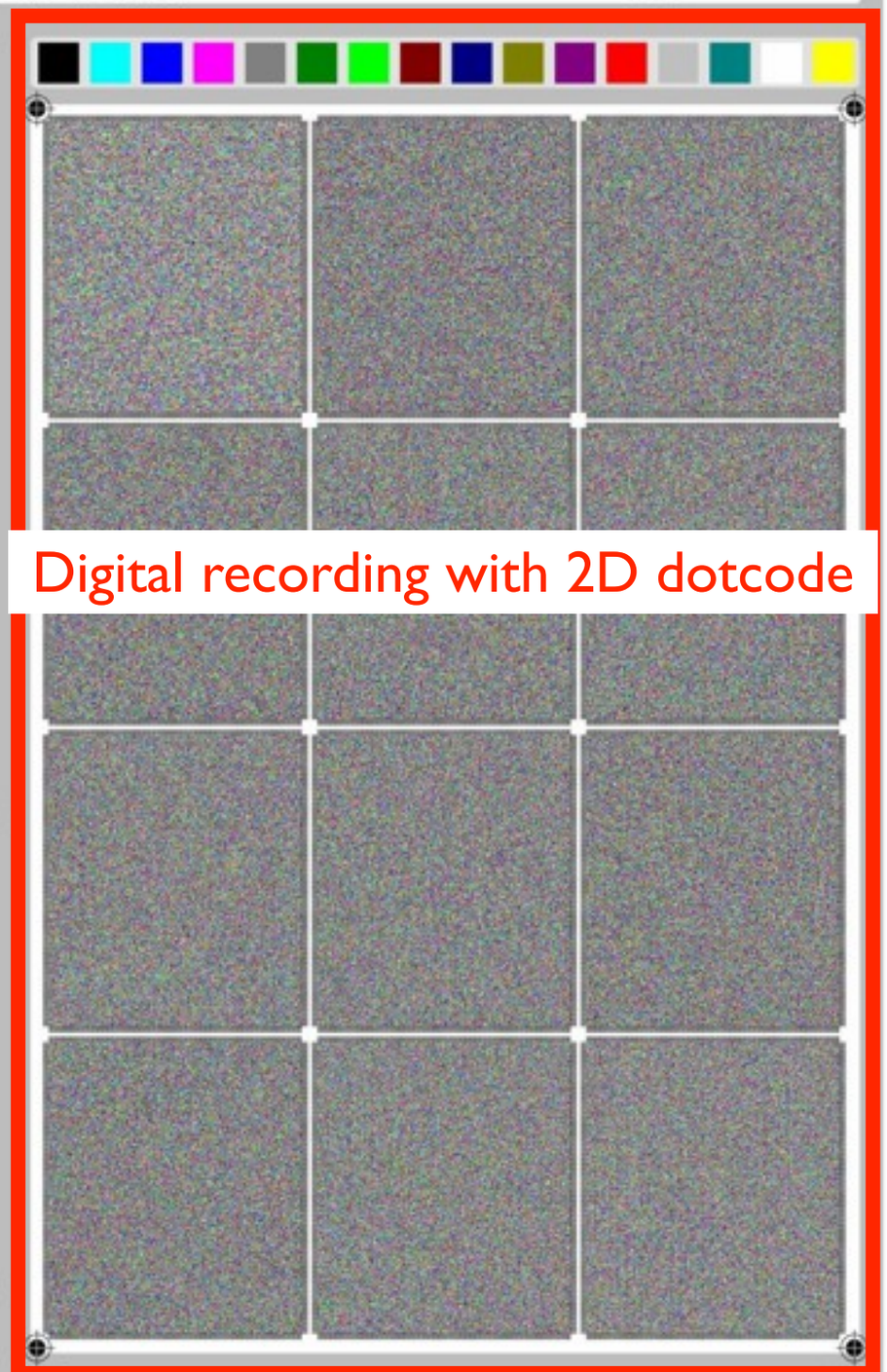
human readeable metadata

IDENTIFIER	CHBS_LIB99_AX01
SOURCE	FRIDF_L01_A001
RIGHTS	(c) Graf, F. Otto
RIGHTS	All rights reserved

```
<?xml version="1.0"?>
METS_Profile xmlns="http://www.loc.gov/METS_Profile/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/METS_Profile/
http://www.peviar.ch/mets/mets.profile.v1-2.xsd">
  <URI LOCTYPE="URL">http://www.iml.unibas.ch/Peviar/Mets-Profile</URI>
  <title>Peviar Mets Profile</title>
  <abstract>
  This is the METS Profile used to describe digital objects archived with Peviar.
  </abstract>
  <date>2006-01-01T12:00:00</date>
  <contact>
    <name>N.N.</name>
    <institution>Imaging and Media Lab, University of Basel</institution>
    <address>Bernoullistrasse 32, 4051 Basel</address>
    <phone>+41610000000</phone>
    <email>peviar@iml.unibas.ch</email>
  </contact>
  <registration_info regDate="" regURI=""/>
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>Digitalisierung von "Mona Lisa"</dc:title>
  <dc:creator>Graf, F. Otto</dc:creator>
  <dc:publisher>My Digital Library</dc:publisher>
  <dc:type>image</dc:type>
  <dc:format>image-tiff</dc:format>
  <dc:format>Colordepth: 8 bit</dc:format>
  <dc:format>Color Space: Adobe RGB</dc:format>
  <dc:identifier>CHBS_LIB99_AX01</dc:format>
  <dc:source>FRIDF_L01_A001</dc:source>
  <dc:rights>Copyright Graf, F. Otto</dc:rights>
  </metadata>
</METS_Profile>
```

```
/**
 * PEVIAR - Imaging & Media Lab,
 * University of Basel, Switzerland
 */

/** Example readback program */
int[][] matrix = getImageMatrix();
int[] bitstream = getBitstream(matrix);
int[] decoded = getDecodedBitstream(bitstream);
int[] imageMetaData = getImageMetaData(decoded);
int[] imageData = getImageData(decoded);
TIFF img = new TIFF(imageMetaData);
img.setData(imageData);

/** Method Detail */
getImageMatrix():
  // pseudocode for getImageMatrix()

getBitstream(int[][] matrix):
  // pseudocode for getBitstream()

getDecodedBitstream(int[] bitstream):
  // pseudocode for getDecodedBitstream()

getImageMetaData(int[] decoded):
  // pseudocode for getImageMetaDta()

getImageData(int[] decoded):
  // pseudocode for getImageData()
```





Digital recording with 2D dotcode

PEVIAR: Example

UNIVERSITÄT BASEL

recorded by: mikrosave

imaging & media lab

FLS_20070220_MKBA_0001

analogue image

TITLE	Mona Lisa
PUBLISHER	Graf Library Ltd.
CREATOR	Graf, F. Otto
DESCRIPTION	Foto of 'Mona Lisa'
DATE	2006-08-05T12:00:00
TYPE	Image

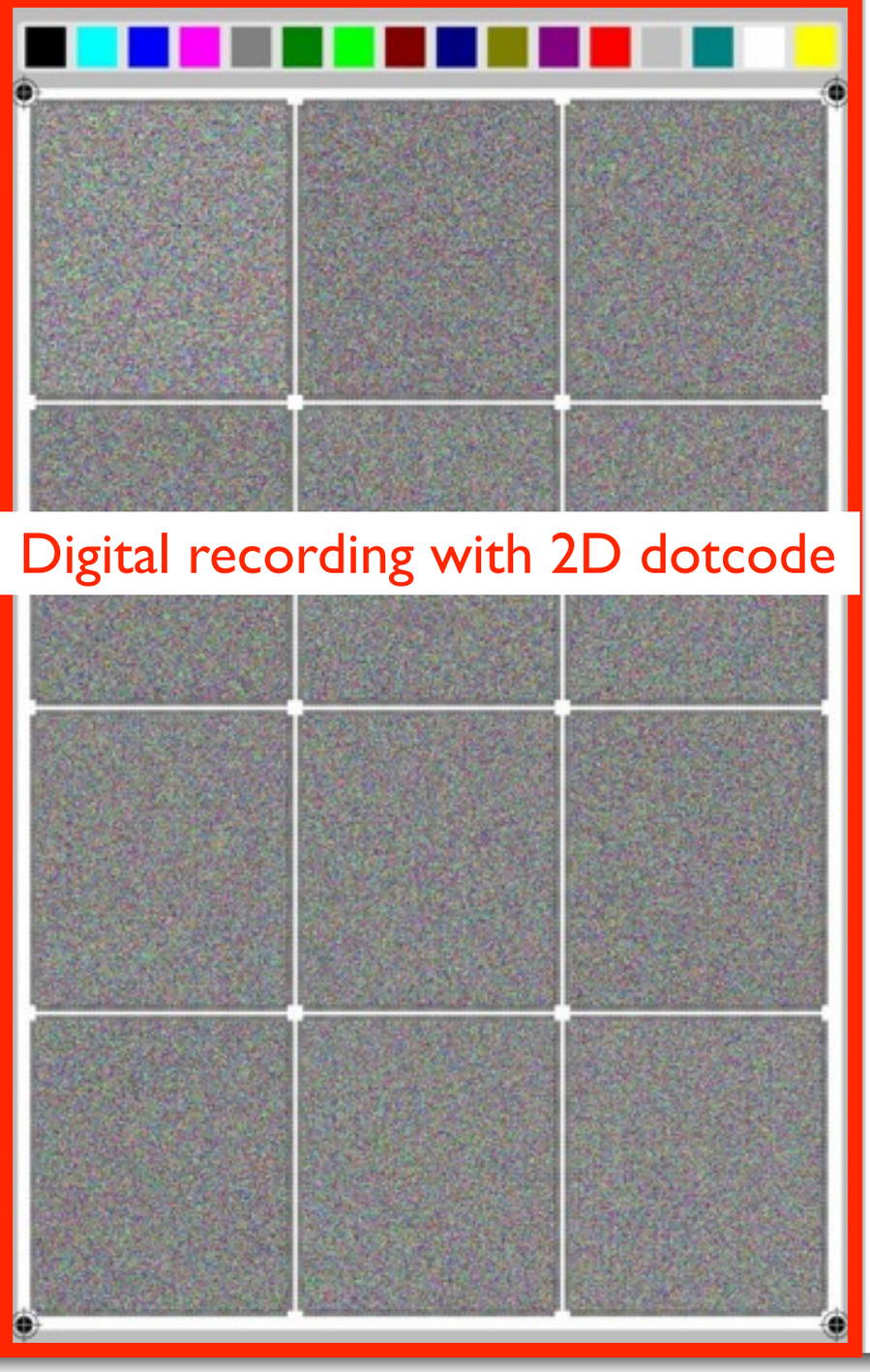
human readeable metadata

IDENTIFIER	CHBS_LIB99_AX01
SOURCE	FRIDF_L01_A001
RIGHTS	(c) Graf, F. Otto
RIGHTS	All rights reserved

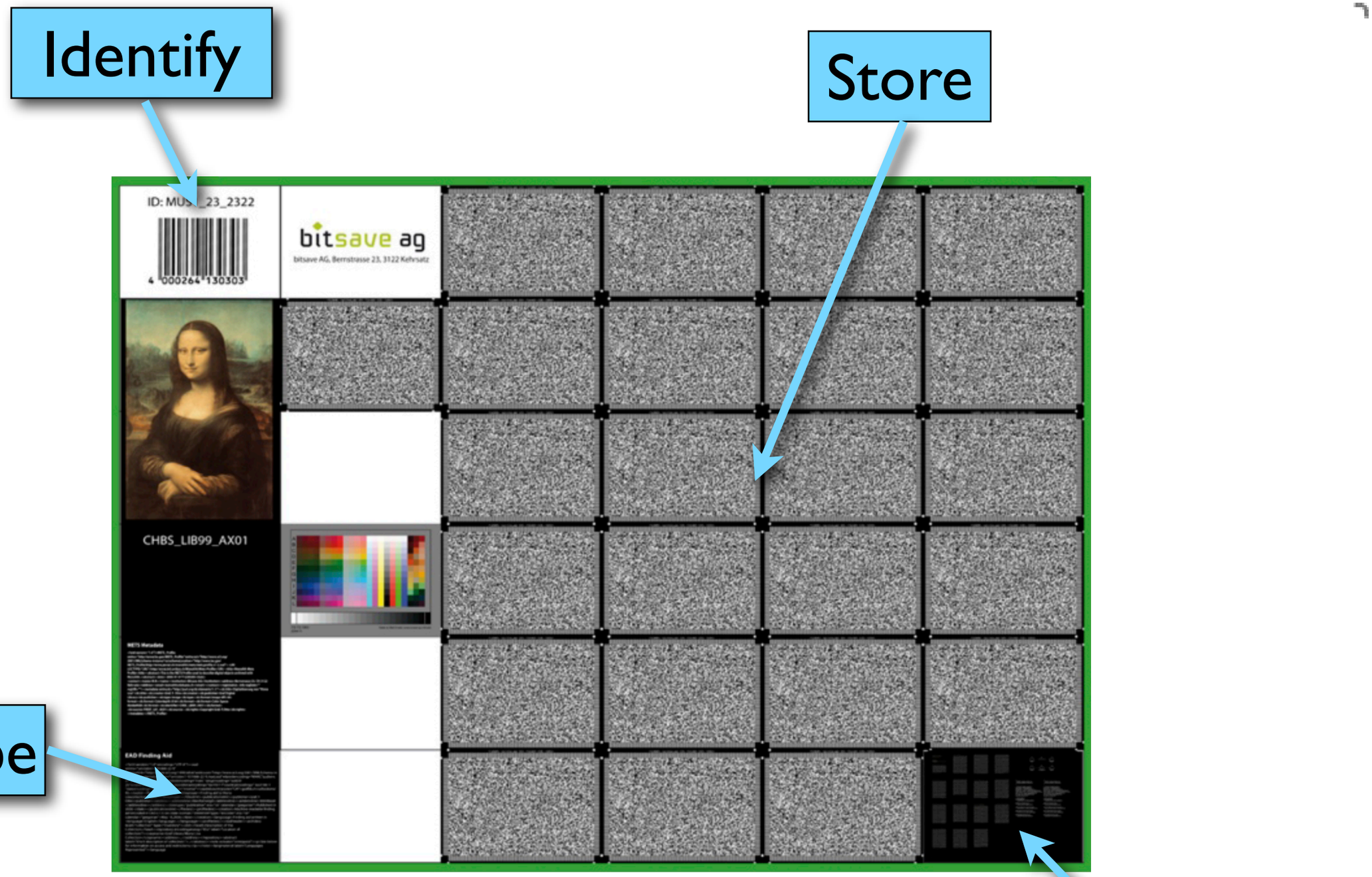
```
<?xml version="1.0"?>
METS_Profile xmlns="http://www.loc.gov/METS_Profile/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/METS_Profile/
http://www.peviar.ch/mets/mets.profile.v1-2.xsd">
  <URI LOCTYPE="URL">http://www.iml.unibas.ch/Peviar/Mets-Profile</URI>
  <title>Peviar Mets Profile</title>
  <abstract>
  This is the METS Profile used to describe digital objects archived with Peviar.
  </abstract>
  <date>2006-01-01T12:00:00</date>
  <contact>
    <name>N.N.</name>
    <institution>Imaging and Media Lab, University of Basel</institution>
    <address>Bernoullistrasse 32, 4051 Basel</address>
    <phone>+41610000000</phone>
    <email>peviar@iml.unibas.ch</email>
  </contact>
  <registration_info regDate="" regURI=""/>
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>Digitalisierung von "Mona Lisa"</dc:title>
  <dc:creator>Graf, F. Otto</dc:creator>
  <dc:publisher>My Digital Library</dc:publisher>
  <dc:type>image</dc:type>
  <dc:format>image-tiff</dc:format>
  <dc:format>Colordepth: 8 bit</dc:format>
```

technical metadata e.g. XML, decoding information format specification, etc.

```
 * University of Basel, Switzerland
 */

/** Example readback program */
int[][] matrix = getImageMatrix();
int[] bitstream = getBitstream(matrix);
int[] decoded = getDecodedBitstream(bitstream);
int[] imageMetaData = getImageMetaData(decoded);
int[] imageData = getImageData(decoded);
TIFF img = new TIFF(imageMetaData);
img.setData(imageData);

/** Method Detail */
getImageMatrix():
  // pseudocode for getImageMatrix()

getBitstream(int[][] matrix):
  // pseudocode for getBitstream()

getDecodedBitstream(int[] bitstream):
  // pseudocode for getDecodedBitstream()

getImageMetaData(int[] decoded):
  // pseudocode for getImageMetaData()

getImageData(int[] decoded):
  // pseudocode for getImageData()
```

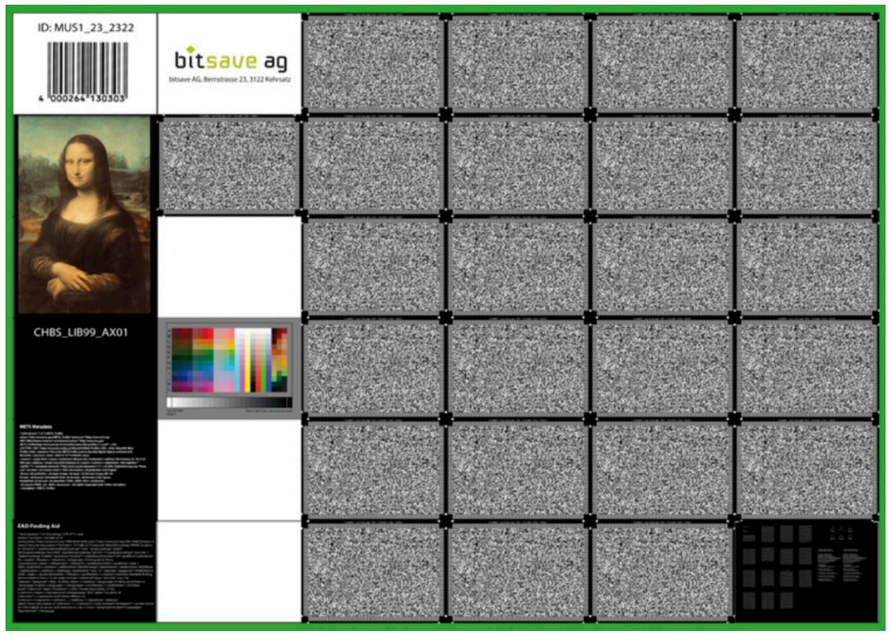


Digital recording with 2D dotcode



Monolith™ Datafilm example provided by bitsave ag

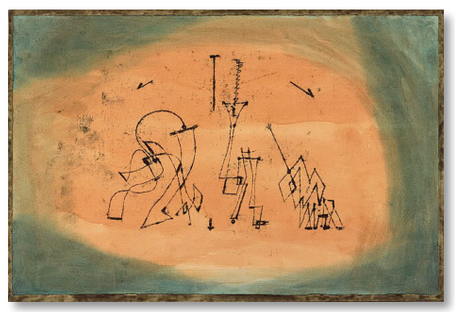
Decoding



scanning and
decoding

.pdf

.wav



.j2k

Recording method

- B/W: 1 dot = 1Bit
- Color: 1 dot = 3 Bit
- 3x3 dots = 1 Byte + Parity

- ReedSolomon: 255 Byte-blocks with 223 bytes payload and 32 parity bytes
- Bytes are locally distributed within a frame using a deterministic random number generator

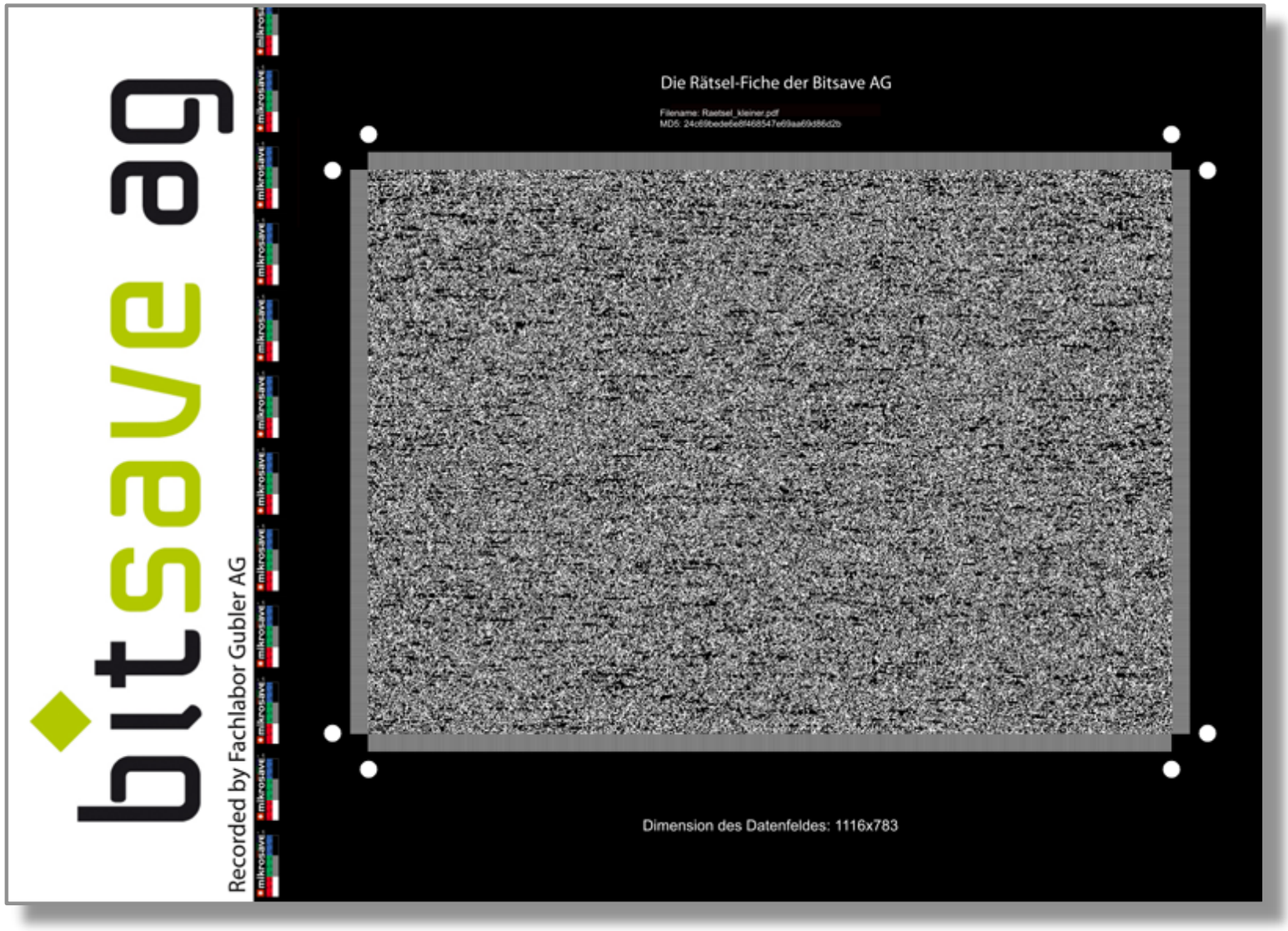
Properties of PEVIAR

- stable, independent of a specific technology and selfexplaining
- Capacity (with current technology):
  - on color microfilm, 10,5cm x14,8cm
    ca. 100MB max
  - on one movie-film sized 35mm frame:
    ca. 2-3MB max
- Longevity of medium: 500+ years

in other words....

- migrationless
- long-lasting
- authenticity guaranteed
  - electronically archived information can always be forged without leaving traces - digital signatures make it more difficult, but how to archive the signature?
  - any tampering with microfilm leaves visible traces!
  - It is “original” material (forensics possible)

Try it Yourself!!



Decoding-software and instructions found on
http://www.bitsave.ch/download/download.html