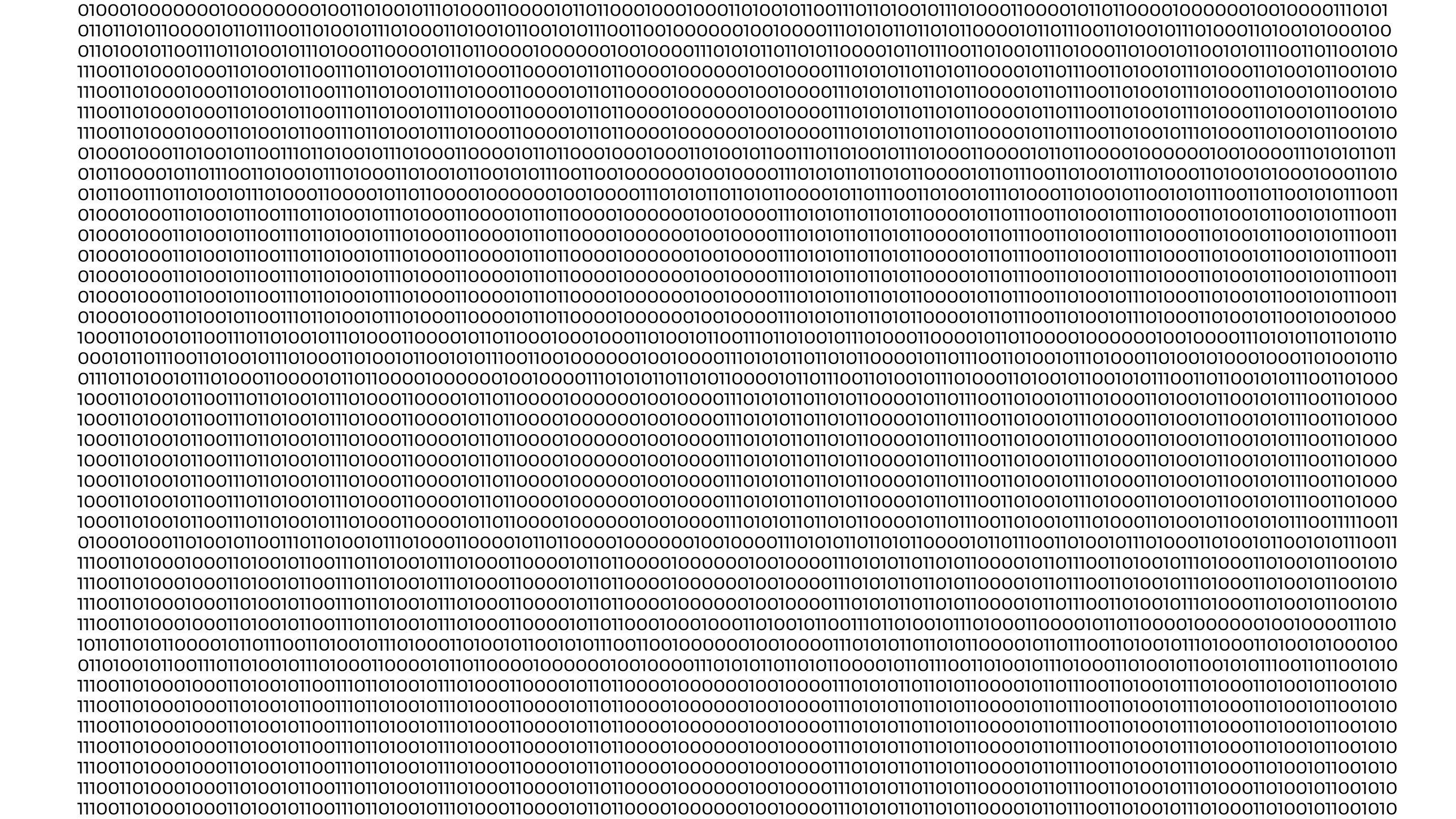
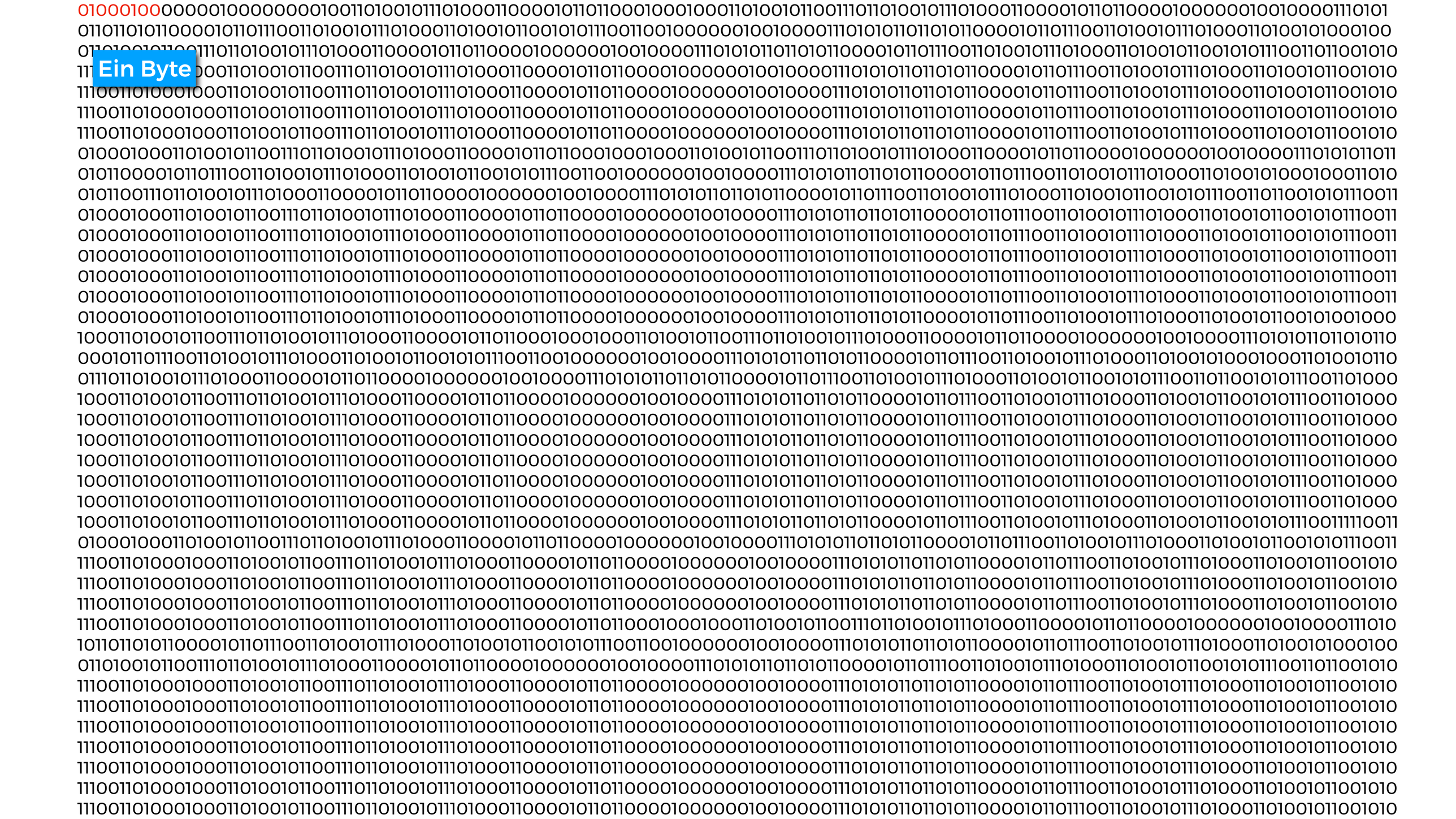
Universität Basel

Kurzer Abriss über eingebettete Metadaten, Primärdaten und deren Archivierung aus der Sicht von OAIS

Peter Fornaro
Digital Humanities Lab
Universität Basel

Das fotografische Kulturerbe im digitalen Zeitalter
15.–16. November 2018 | IDHEAP, Universität Lausanne




Ein Byte

Universität Basel

Image File Header

A TIFF file begins with an 8-byte image file header, containing the following information:

Bytes 0-1: The byte order used within the file. Legal values are:

"II" (4949.H)

"MM" (4D4D.H)

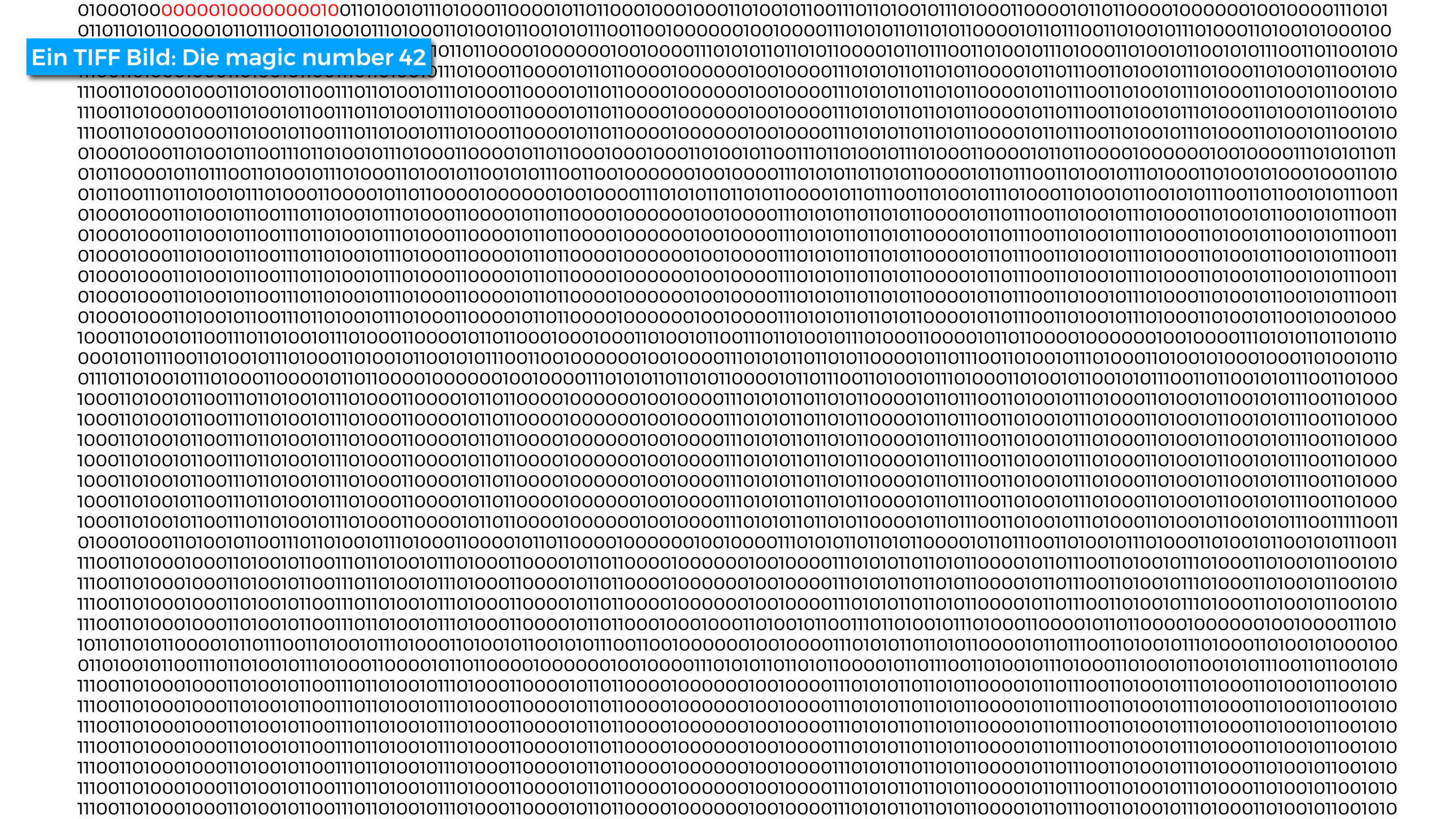
In the "II" format, byte order is always from the least significant byte to the most significant byte, for both 16-bit and 32-bit integers This is called *little-endian* byte order. In the "MM" format, byte order is always from most significant to least significant, for both 16-bit and 32-bit integers. This is called *big-endian* byte order.

Bytes 2-3 An arbitrary but carefully chosen number (42) that further identifies the file as a TIFF file.

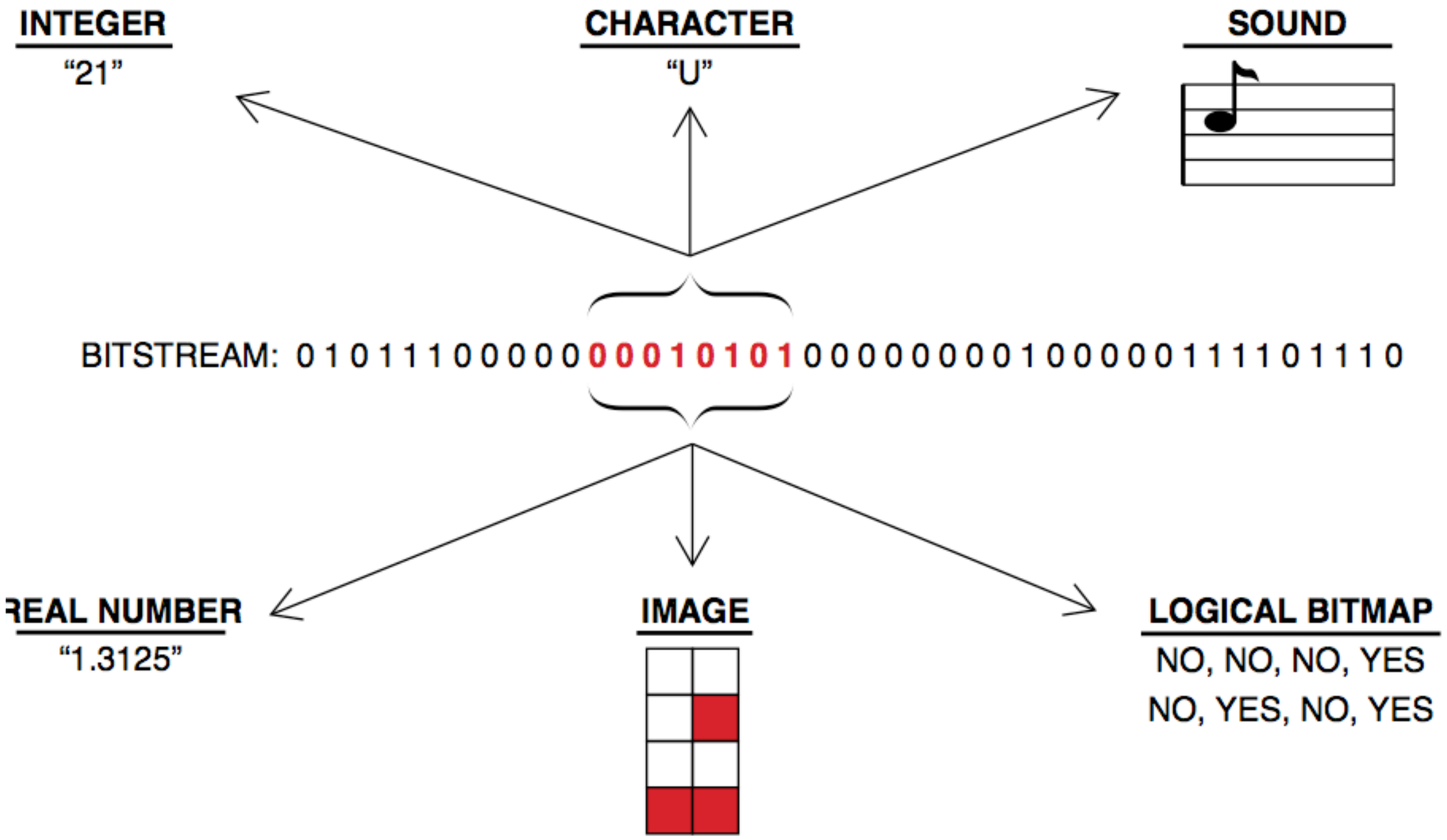
The byte order depends on the value of Bytes 0-1.

Bytes 4-7 The offset (in bytes) of the first IFD. The directory may be at any location in the file after the header but *must begin on a word boundary*. In particular, an Image File Directory may follow the image data it describes. Readers must follow the pointers wherever they may lead.

The term *byte offset* is always used in this document to refer to a location with respect to the beginning of the TIFF file. The first byte of the file has an offset of 0.


Ein TIFF Bild: Die magic number 42

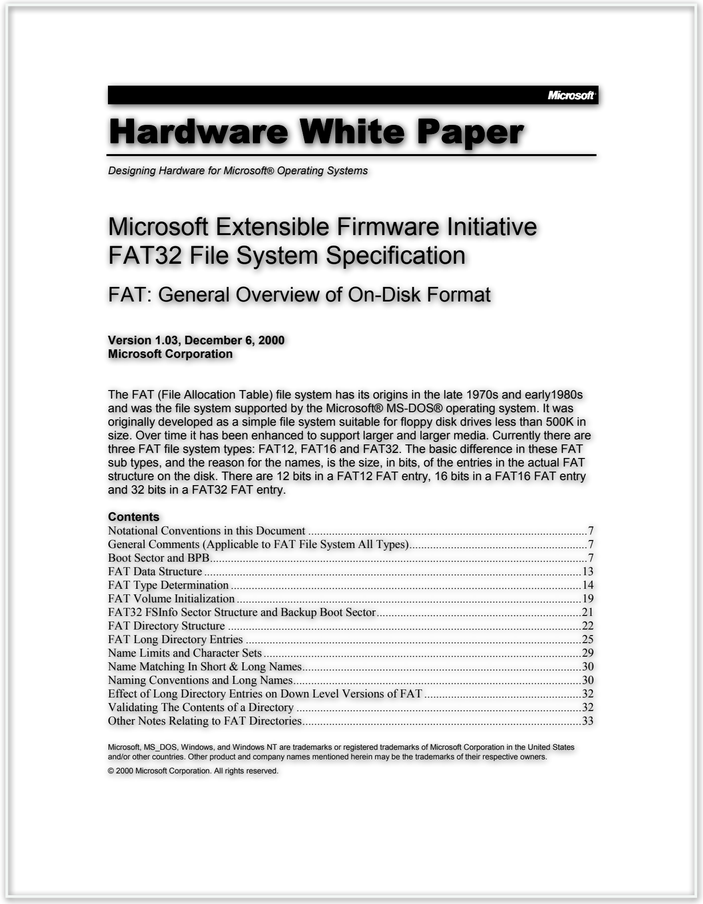

Ein RGB Wert eines Bildes


INTEGER
“21”
CHARACTER
“U”
SOUND
BITSTREAM: 0 1 0 1 1 1 0 0 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 1 1 0 1 1 1 0
REAL NUMBER
“1.3125”
IMAGE
LOGICAL BITMAP
NO, NO, NO, YES
NO, YES, NO, YES

Universität Basel

Metadaten

- Erklärung eines Bits und Bytes
- Erklärung des Dateisystems (z. B FAT32)
- Erklärung des Dateiformats (z.B. TIFF)
- Erklärung des Metadatenformats (z.B. EXIF)
- Die erweiterte Information über den Bildinhalt und dessen Entstehung (IPTC, EXIF)



Microsoft

Hardware White Paper

Designing Hardware for Microsoft® Operating Systems

Microsoft Extensible Firmware Initiative FAT32 File System Specification

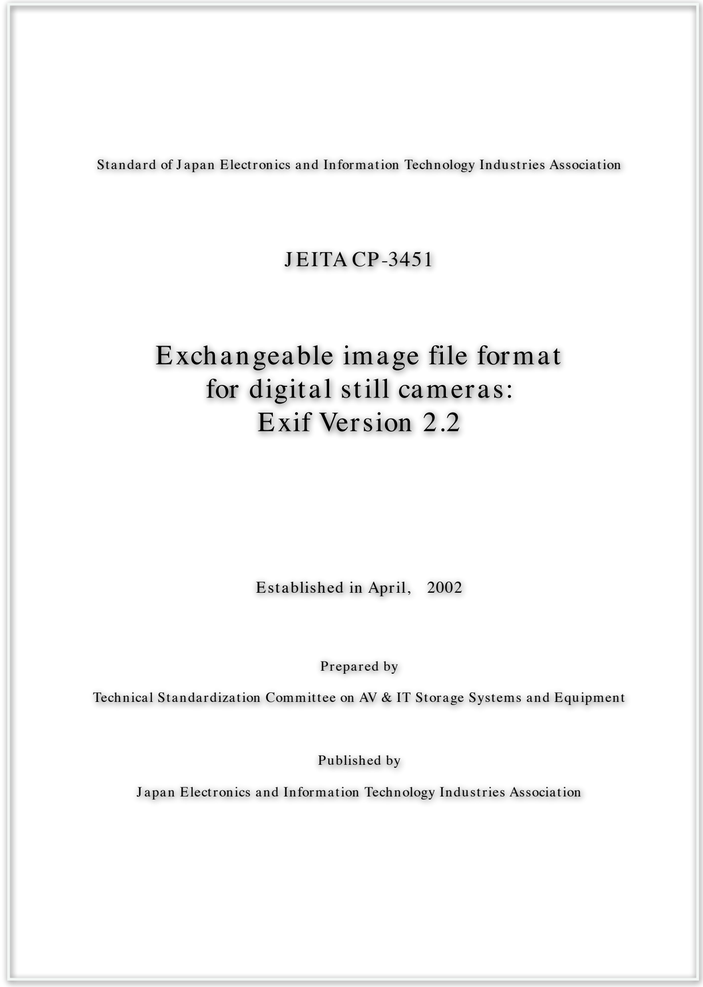
FAT: General Overview of On-Disk Format

Version 1.03, December 6, 2000
Microsoft Corporation

The FAT (File Allocation Table) file system has its origins in the late 1970s and early1980s and was the file system supported by the Microsoft® MS-DOS® operating system. It was originally developed as a simple file system suitable for floppy disk drives less than 500K in size. Over time it has been enhanced to support larger and larger media. Currently there are three FAT file system types: FAT12, FAT16 and FAT32. The basic difference in these FAT sub types, and the reason for the names, is the size, in bits, of the entries in the actual FAT structure on the disk. There are 12 bits in a FAT12 FAT entry, 16 bits in a FAT16 FAT entry and 32 bits in a FAT32 FAT entry.

Contents

Filesystem => **32 Seiten**
technische Metainformation



Standard of Japan Electronics and Information Technology Industries Association

JEITA CP-3451

Exchangeable image file format for digital still cameras: Exif Version 2.2

Established in April, 2002

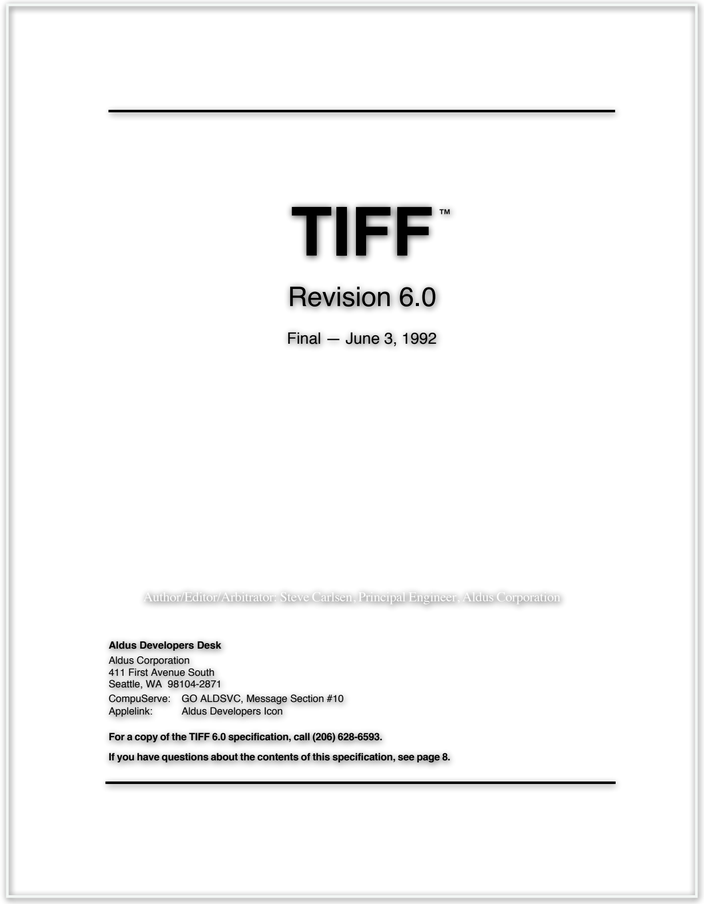
Prepared by

Technical Standardization Committee on AV & IT Storage Systems and Equipment

Published by

Japan Electronics and Information Technology Industries Association

Metadaten => **148 Seiten**
technische Metainformation



TIFF™

Revision 6.0

Final — June 3, 1992

Author/Editor/Arbitrator: Steve Carlsen, Principal Engineer, Aldus Corporation

Aldus Developers Desk
Aldus Corporation
411 First Avenue South
Seattle, WA 98104-2871
CompuServe: GO ALDSVC, Message Section #10
Applelink: Aldus Developers Icon

For a copy of the TIFF 6.0 specification, call (206) 628-6593.

If you have questions about the contents of this specification, see page 8.

Dateiformat => **121 Seiten**
technische Metainformation

Datum und Uhrzeit
Orientierung (Bilddrehung um 90, 180 bzw. 270°)
Brennweite
Belichtungszeit
Blendenzahl
...

Farbraum
Farbtransformation
Lichtquelle
Rendering Intent

Urheberrechtsvermerke,
den Namen des Erstellers,
eine Überschrift oder Stich-/Schlagwörter

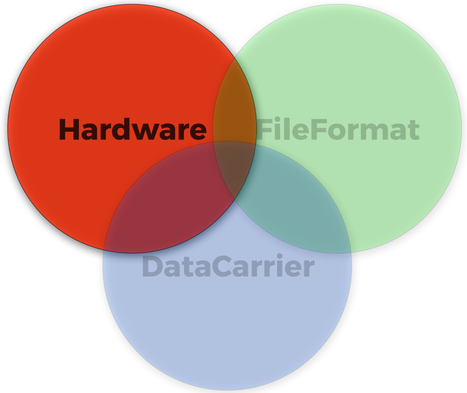
Universität Basel

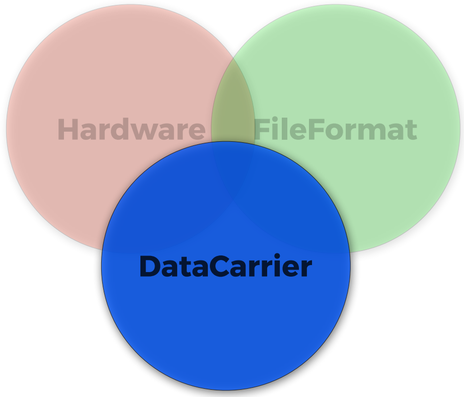
XMP (von Adobe)

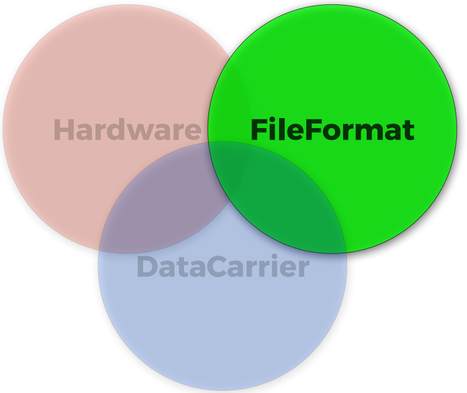
XMP basiert auf offenen Standards und bettet die vom World Wide Web Consortium veröffentlichte formale Sprache **RDF** (Resource Description Framework) in Binärdaten ein. Damit sollen Metadaten in verschiedenen Applikationen nach einem einheitlichen Schema integriert werden, um **von anderen Programmen** gelesen werden zu können.

Interoperabilität

Universität Basel







Migration heisst den bitstream **kopieren**.

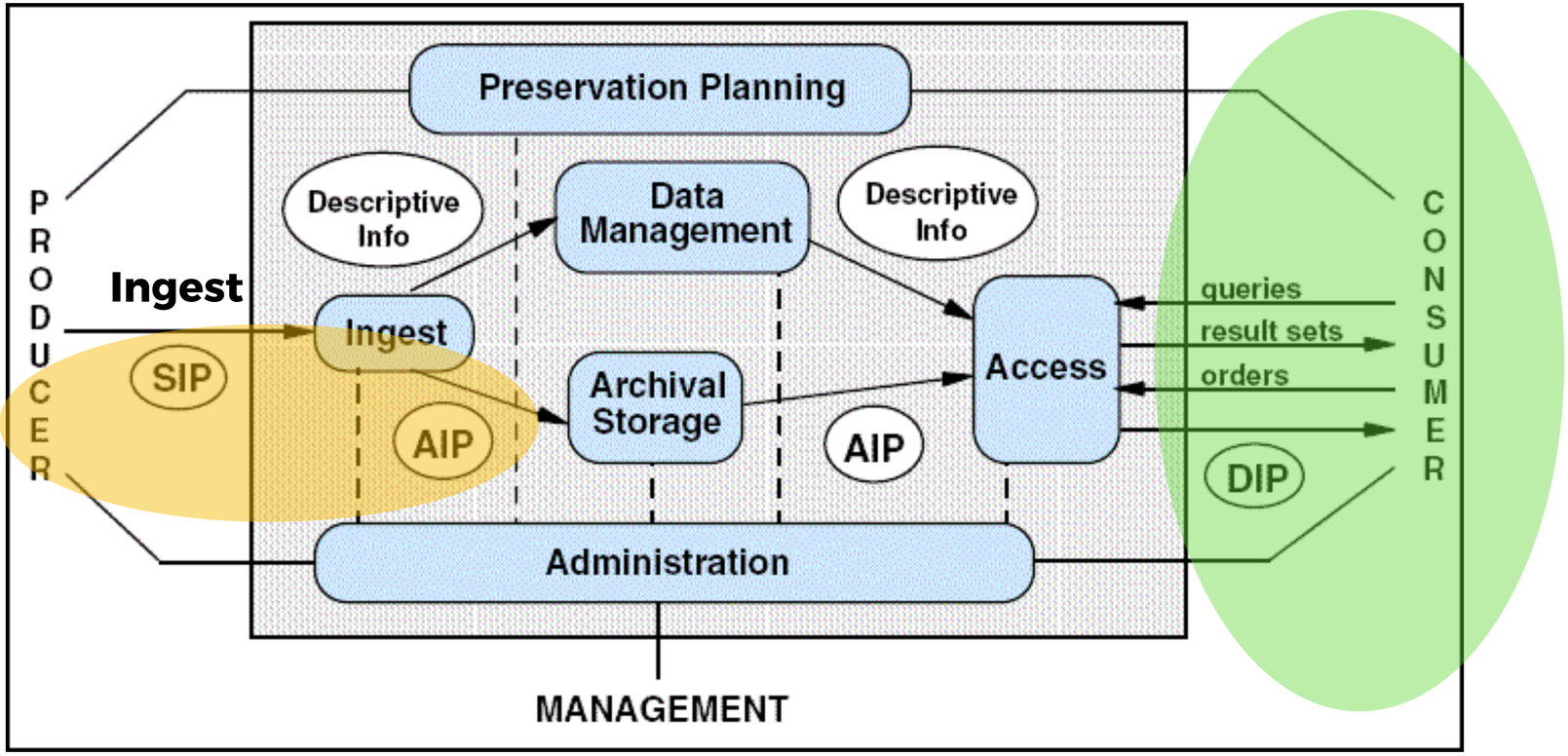
Migration heisst den bitstream **kopieren**.

Migration heisst den bitstream **transcodieren**.

Medienstandards sind wichtig: Dies beinhaltet auch detaillierte Informationen über verwendete Daten- und Metadatenformate. So kann eine aufwändige Formatmigration vermieden werden

Universität Basel

OAIS: Open Archival Information System



Universität Basel

Webzugriff

Webzugriff

Webzugriff

Nur eine “Masterdatei“ (JPEG2000) in entsprechender Auflösung ist auf dem Bildserver gespeichert.

Alle Derivate werden beim Aufruf erzeugt und nicht gespeichert.

Universität Basel

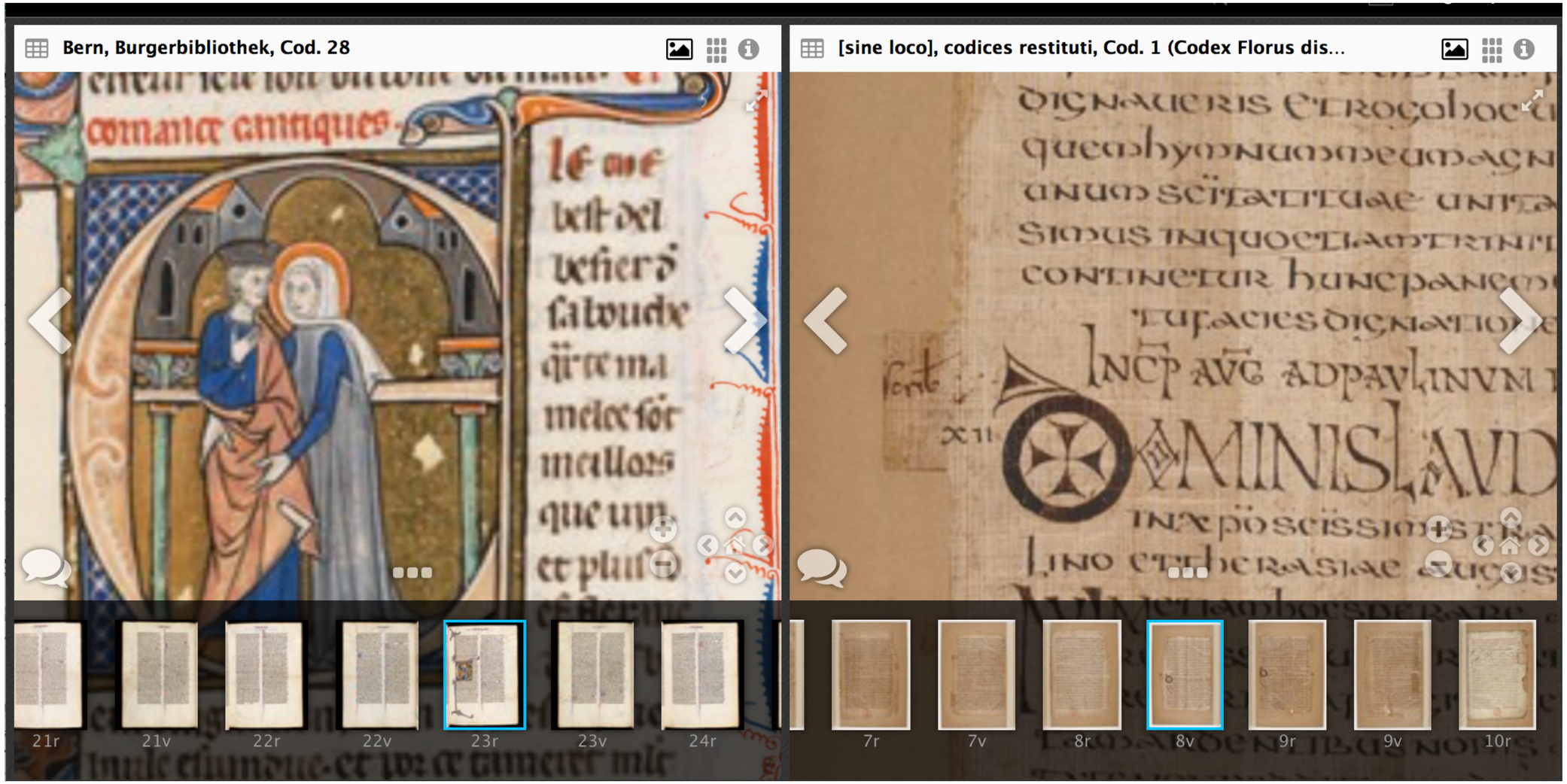
International Image Interoperability Framework

http://iiif.io

- Standardisierter Zugriff auf **Bild Ressourcen** mit HTTP:
  - **IIIF Image API 2.1**
    - dynamic resolution, region of interest, rotation, format
  - **IIIF Presentation API 2.1**
    - structured information (semantic metadata) about an object (image, collection)
    - based on JSON-LD (Linked data)
  - **IIIF Authentication API 1.0**
    - based on JSON-LD (implementation could be web-token)
  - **IIIF Search API 1.0**
    - based on JSON-LD

Universität Basel

e-codices: Swiss virtual Library of Medieval Manuscripts



Universität Basel

Interoperabilität

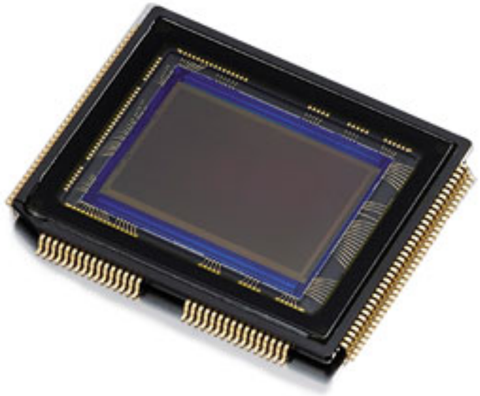
- Heisst Vereinheitlichung der Daten und vor allem der Metadaten, um Vergleichbarkeit zu schaffen.
- Das Wissen aus den vergangen Jahren fliesst in moderne Technologien ein: JPEG2000, RDF.
- Es steht und fällt mit dem “community building“ und Beispielen, die laufen und die gezeigt werden können.

Universität Basel

Informationsdichte in digitaler und analoger Fotografie

Universität Basel

Analog



Electronic≠Digital

Universität Basel

Niedrigempfindliches Material

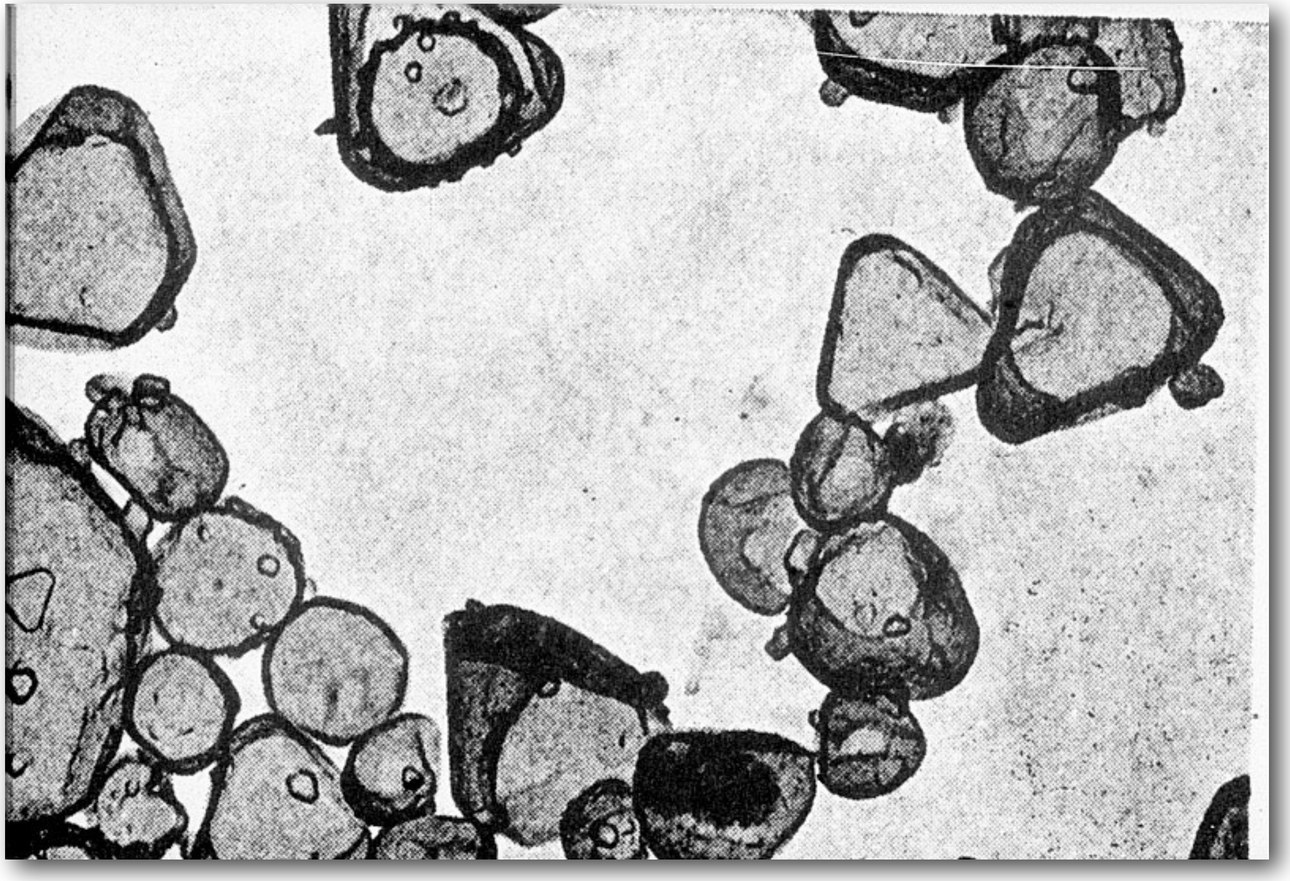
(Kodak Technical Pan in Neofin Doku entwickelt)

Analog

Electronic≠Digital

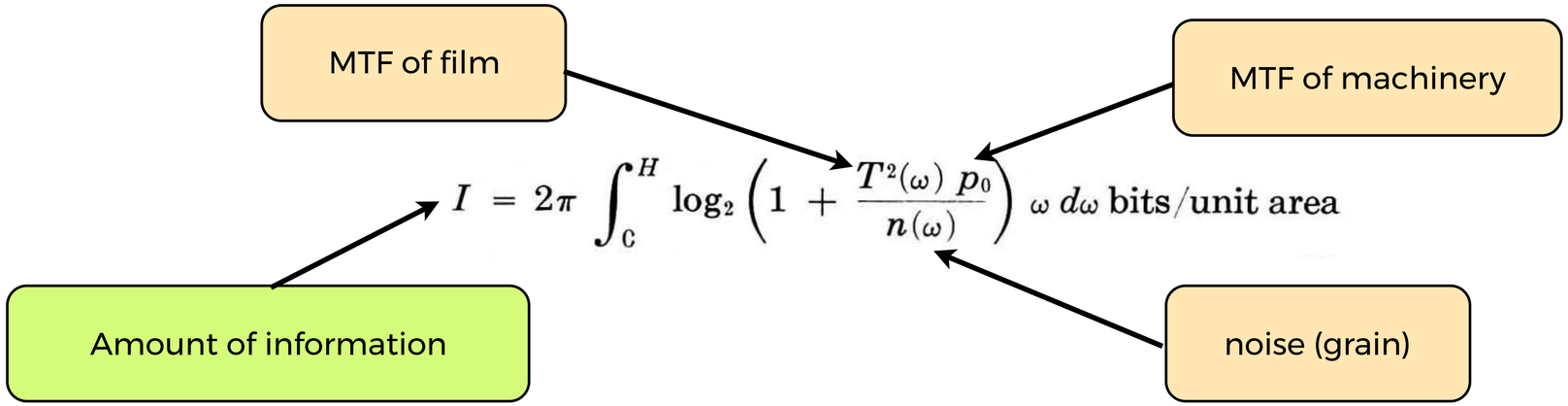
Universität Basel

Halbton, **hochempfindlich Emulsion**



Universität Basel

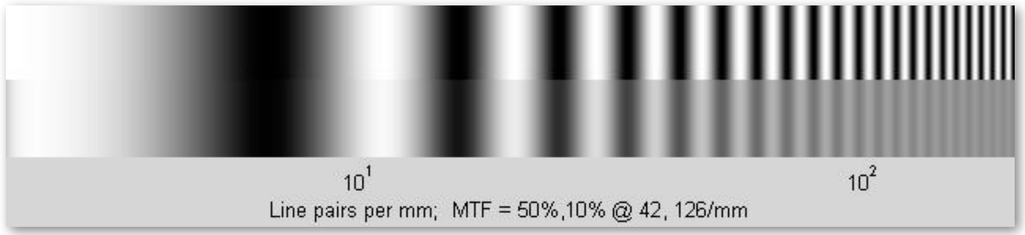
Film als Informationsträger, die Theorie



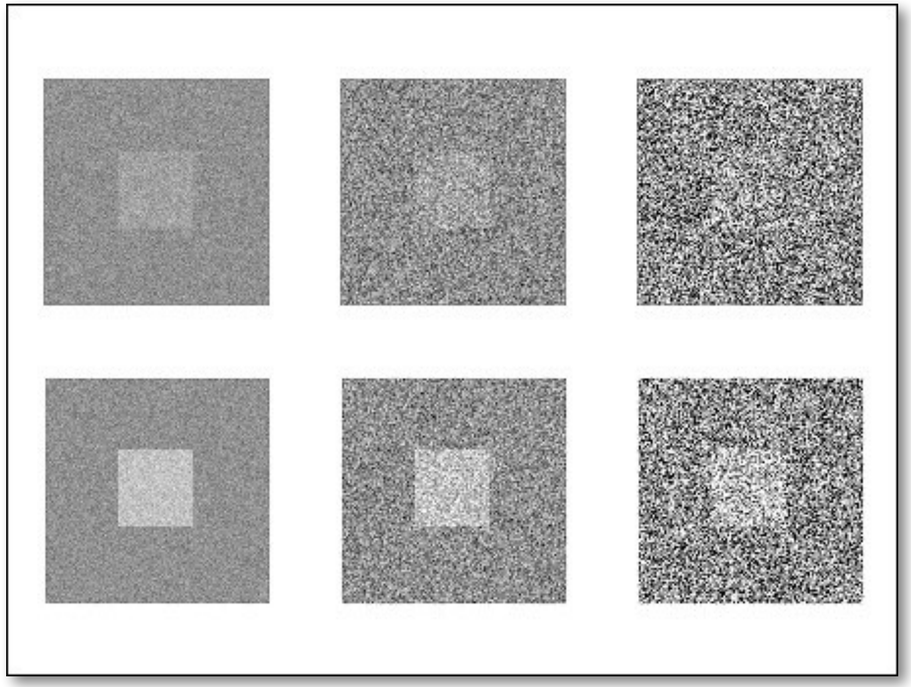
C. E. ***Shannon****, Bell System Tech.J ., 27: 379, 623 (1948).*

Universität Basel

Frequency Response, MTF



Signal to Noise Ratio, SNR



Maximaler Informationsgehalt
auf einem Kleinbild-Dia
700 - 800 KByte!

Universität Basel

50 MPixel

1 Bilddatei als TIFF, 16bit, RGB => **300 Mbyte**

Doch was ist der Informationsgehalt?

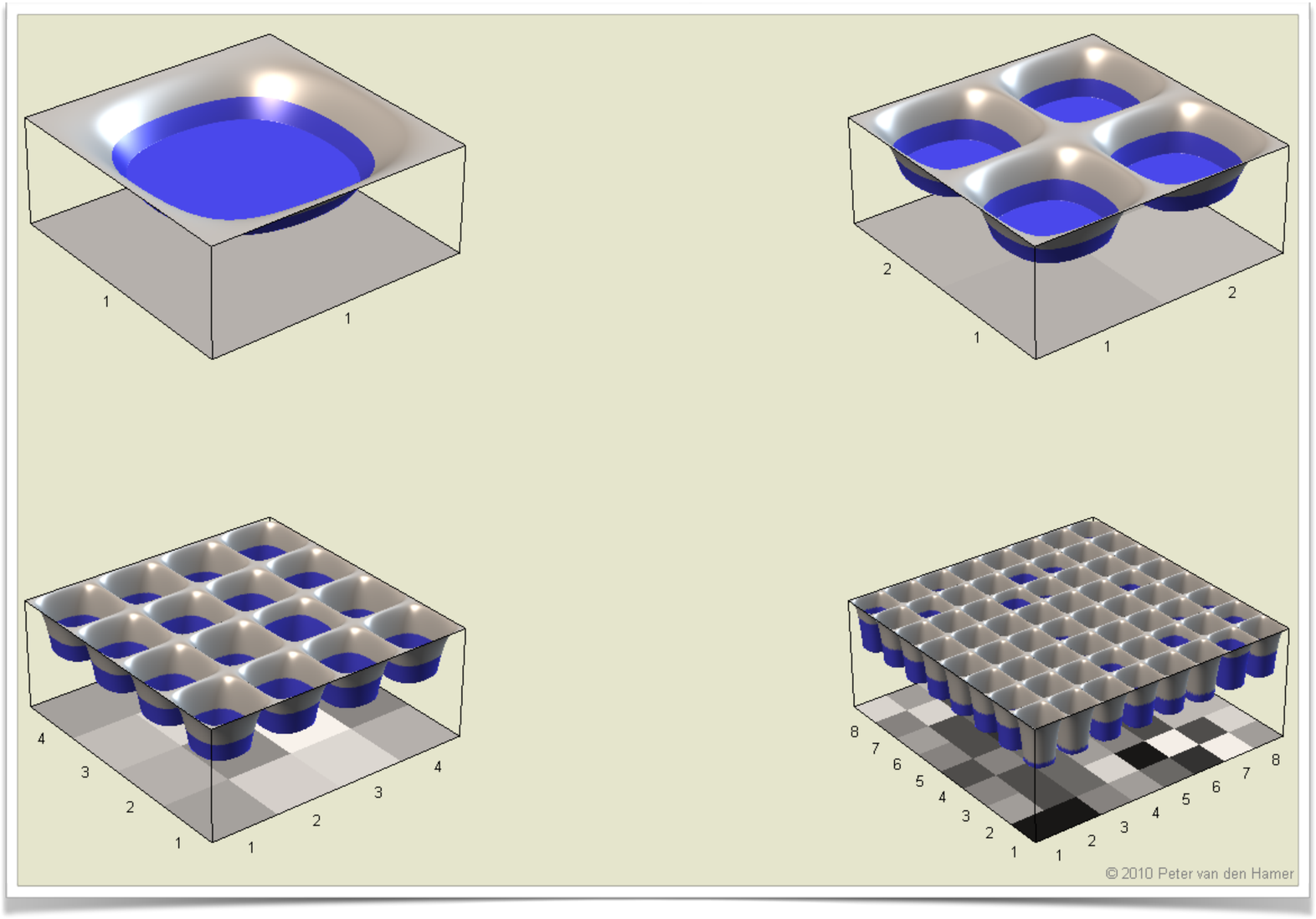
Universität Basel

400 MPixel

1 Bilddatei als TIFF, 16bit, RGB => **2,4 GByte**

Universität Basel

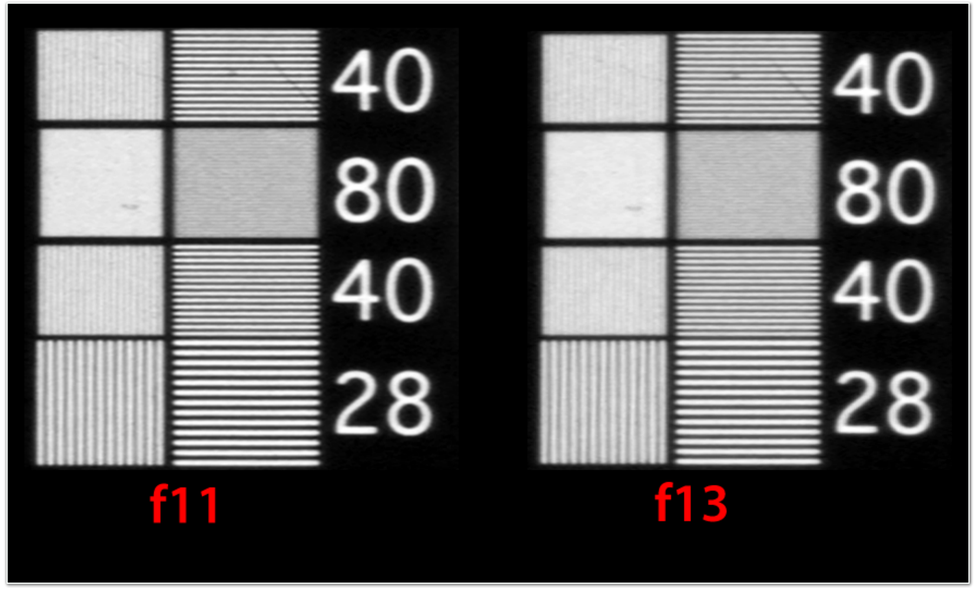
Pixeldichte

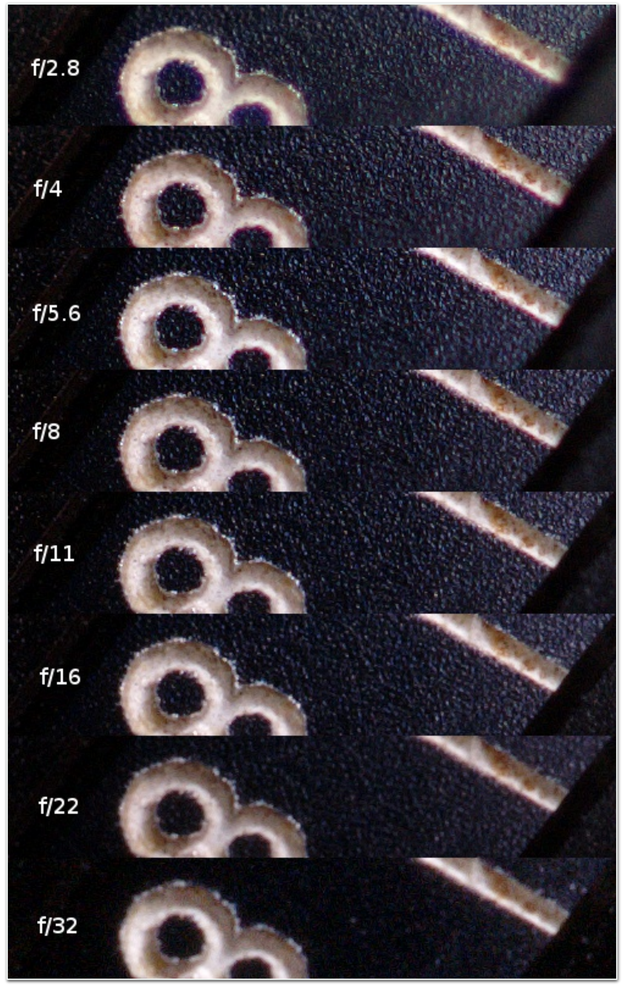


Universität Basel

Grenzen der Optik (Beugung)

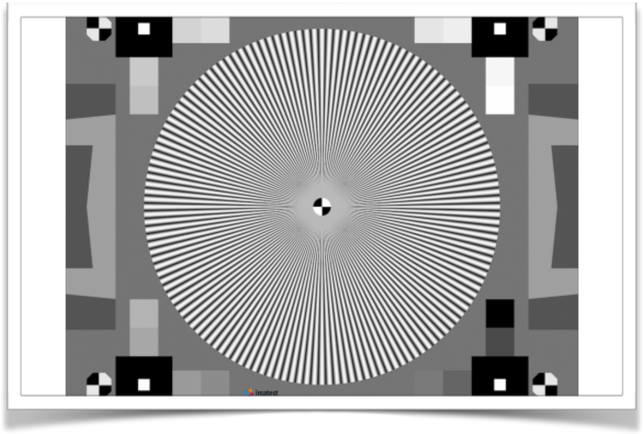
Bilder werden auf Grund optischer Gesetze unscharf!





Universität Basel

Mittelformat <=> Kleinbild



JPEG

2000

Universität Basel

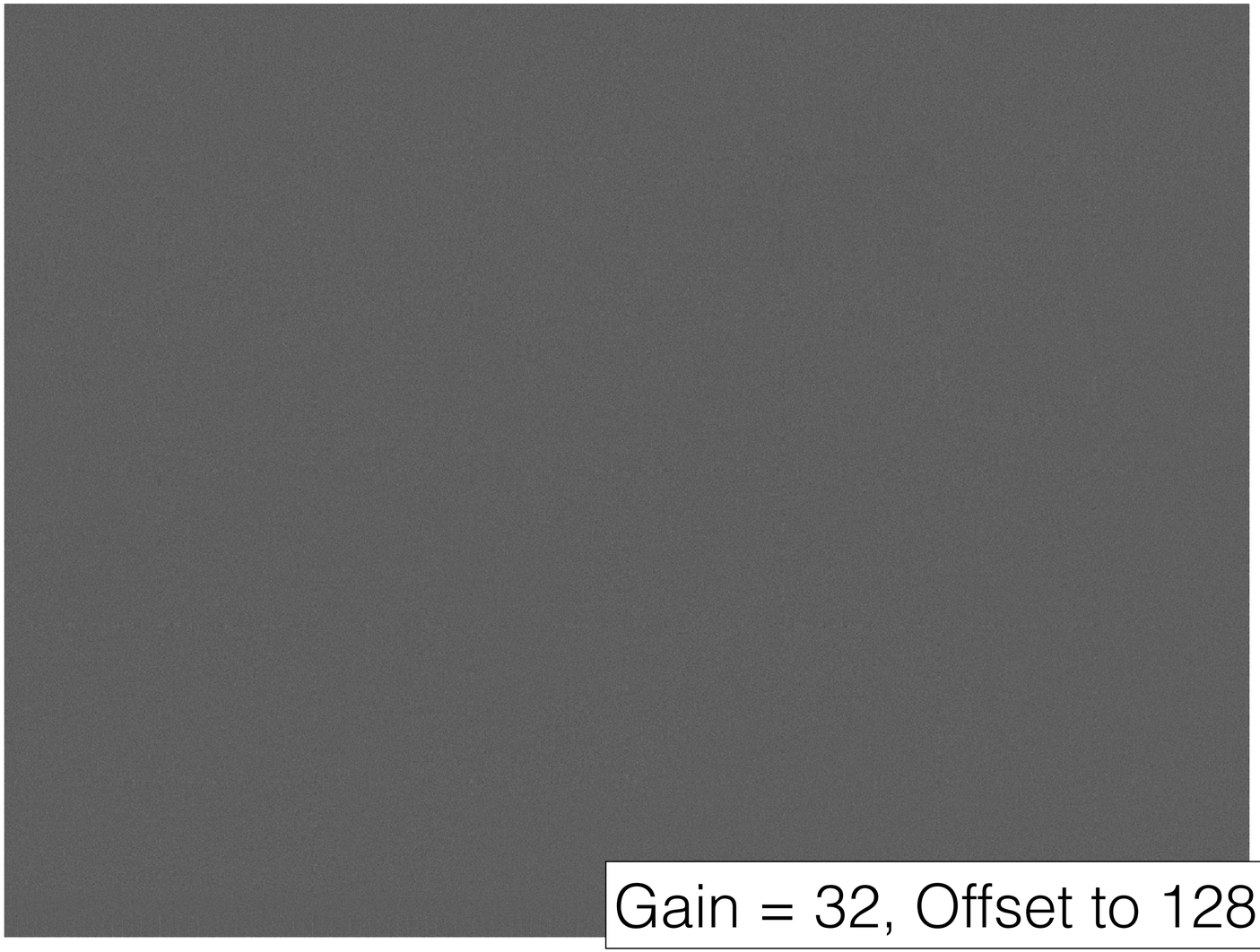
Kompressionswirkung

TIF 16 bit

JPEG2000 16 bit

Universität Basel

Differenz TIFF - J2K



Gain = 32, Offset to 128 (8bit)

Universität Basel

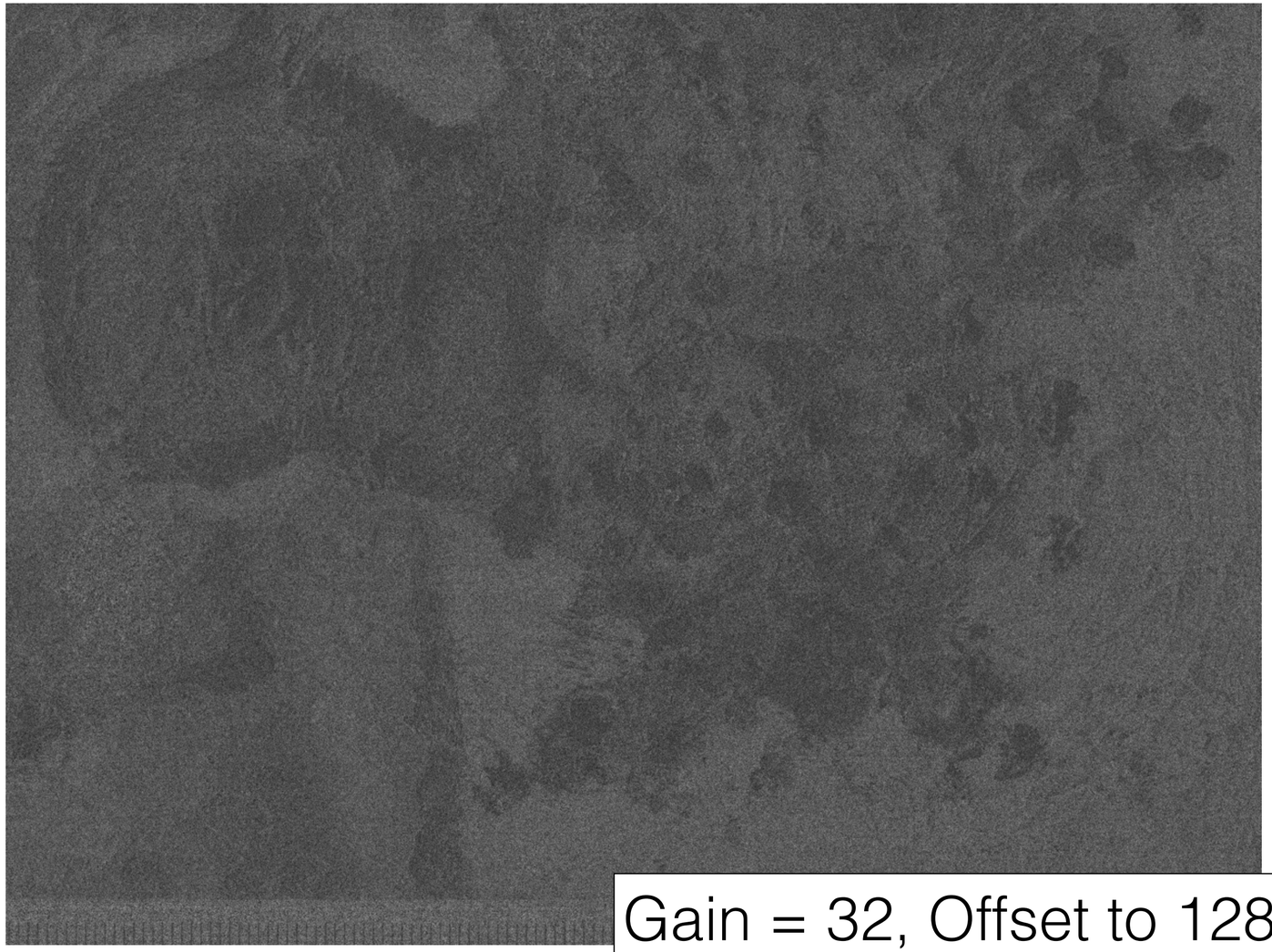
Hardwareeffekte

TIF 16 bit A

TIF 16 bit B (on same machine)

Universität Basel

Differenz TIFF A - TIFF B



Gain = 32, Offset to 128 (8bit)

Universität Basel

Vielen Dank!

peter.fornaro@unibas.ch