

File Formats

Johan van der Knijff
Digital Access to Cultural Heritage 2016

Digital data: bits & bytes

0|1|0|0|1|0|1|1|0|1|0|0|0|0|1|0|...

astronau.txt - Notepad

File Edit Format View Help

"6_3_2_16.TXT" (8108 bytes) was created on 07-08-92

REQUIREMENTS FOR BECOMING AN ASTRONAUT

ANNOUNCEMENT NO. 3ACS-83

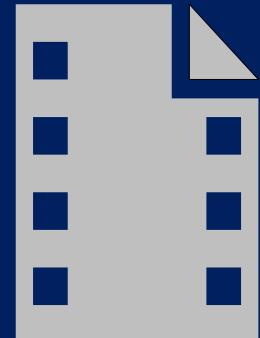
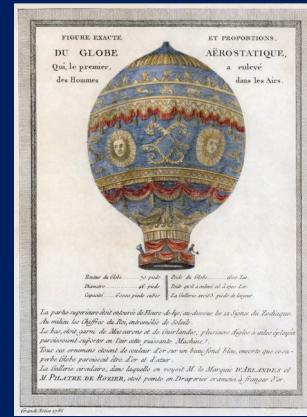
FOR

MISSION SPECIALIST & PILOT ASTRONAUT CANDIDATES

ASTRONAUT CANDIDATE PROGRAM

MISSION SPECIALIST ASTRONAUT

Mission specialist astronauts, working with the commander and pilot, have overall responsibility for the coordination of Shuttle operations in the areas of crew activity planning, consumables usage, and other Shuttle activities affecting experiment operations. Thus, mission specialists are proficient in payload operations and are required to have a detailed knowledge of the Shuttle systems as well as the operational characteristics, mission requirements and objectives, and supporting systems and equipment for each of the experiments that will be conducted on their assigned mission. Mission specialists participate in extravehicular activities, perform special payload handling or maintenance operations using a remote manipulator system, and assist in specific experiment operation at the discretion of the experiment sponsor.

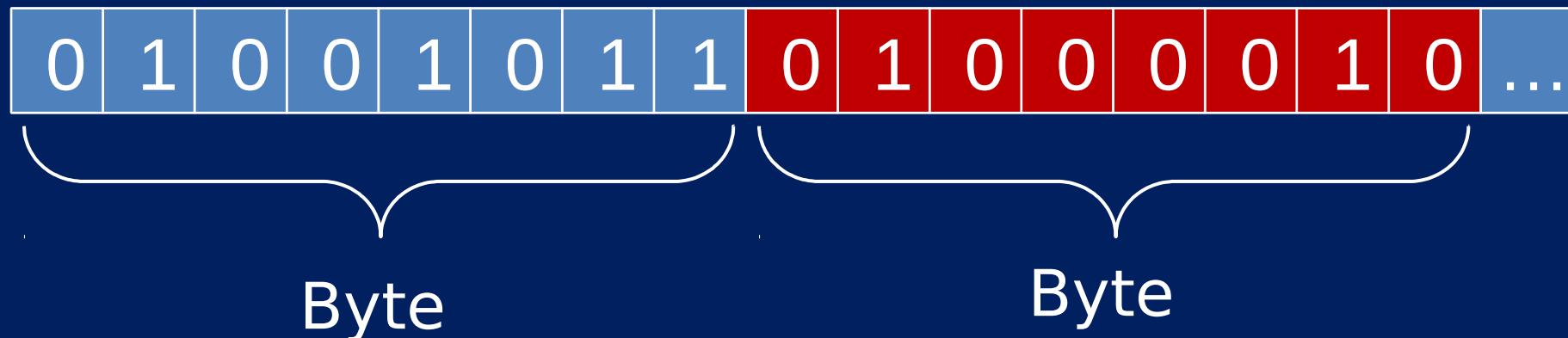


0	1	0	0	1	0	1	1	0	1	0	0	0	0	1	0	...
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	-----

Bit: 1/0

Byte

Group of (usually 8) bits:



Decimal

10 symbols: 0, 1, 2, 3, 4, 5, 6,
7, 8, 9

Binary

2 symbols: 0, 1

75, decimal

10^1	10^0
7	5

75, decimal

10^1	10^0
7	5



least significant
digit, multiples of
 $10^0 = 1$

75, decimal

10^1	10^0
7	5

highest significant
digit, multiples of
 $10^1 = 10$

75, decimal

10^1	10^0
7	5

$$7 \times 10^1 + 5 \times 10^0$$

$$= 70 + 5 = 75$$

75, binary

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
0	1	0	0	1	0	1	1

75, binary

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
0	1	0	0	1	0	1	1



least significant
digit, multiples of
 $2^0 = 1$

75, binary

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
0	1	0	0	1	0	1	1



highest significant
digit, multiples of
 $2^7 = 128$

75, binary

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
0	1	0	0	1	0	1	1

$$\begin{aligned}1 &\times 2^6 + 1 \times 2^3 + 1 \times 2^1 + 1 \times 2^0 \\&= 64 + 8 + 2 + 1 \\&= 75\end{aligned}$$

8 bits: 256 (2^8) possible values

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
0	0	0	0	0	0	0	0

0

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
1	1	1	1	1	1	1	1

255

$$(2^7 + 2^6 + 2^5 + 2^4 + 2^3 + 2^2 + 2^1 + 2^0)$$

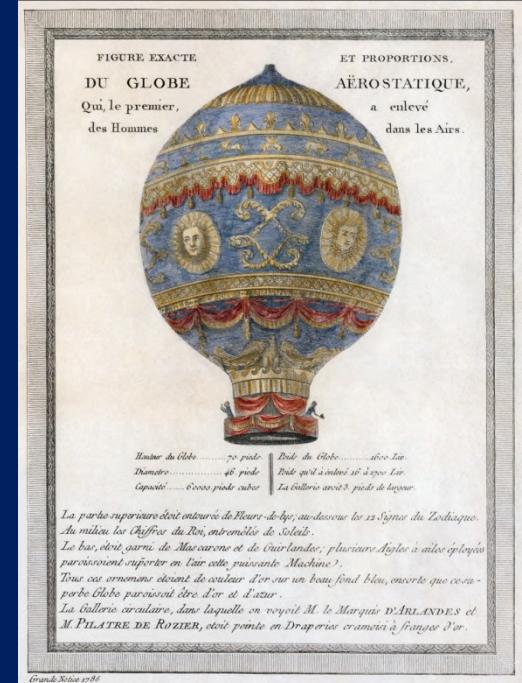
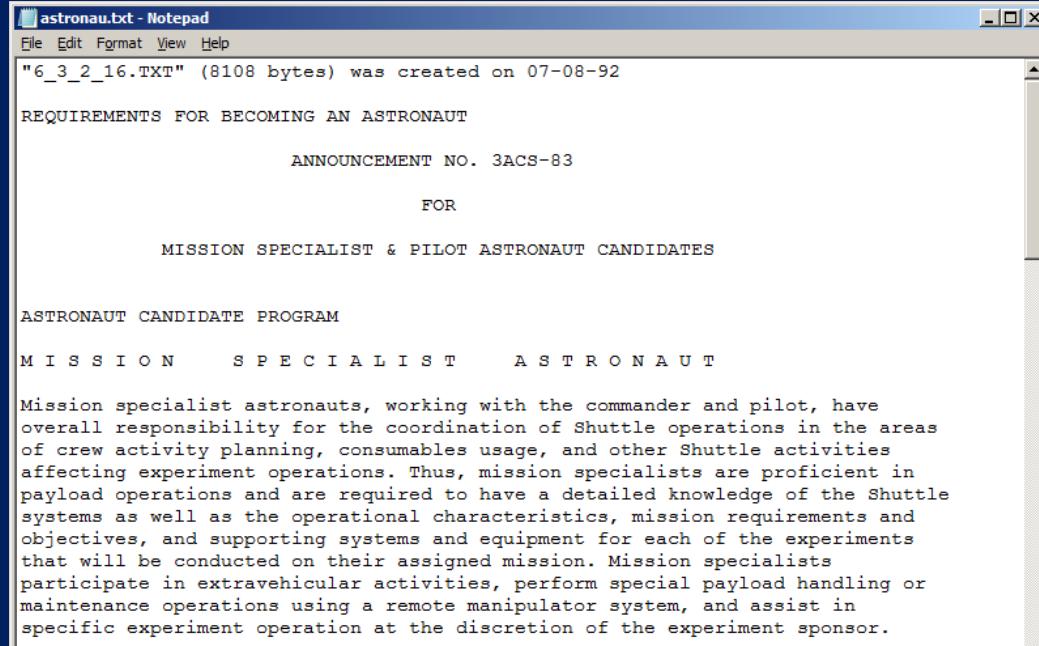
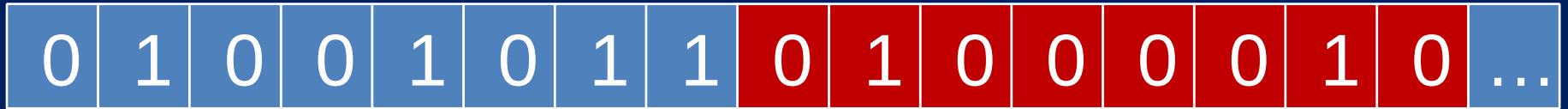
Two Bytes Are Better Than One



Larger numbers?
More bytes!

- **1 byte:** 2^8 values
[0 - 255]
- **2 bytes:** 2^{16} values
[0 - 65,535]
- **4 bytes:** 2^{32} values
[0 - 4,294,967,295]

Text-based vs binary formats



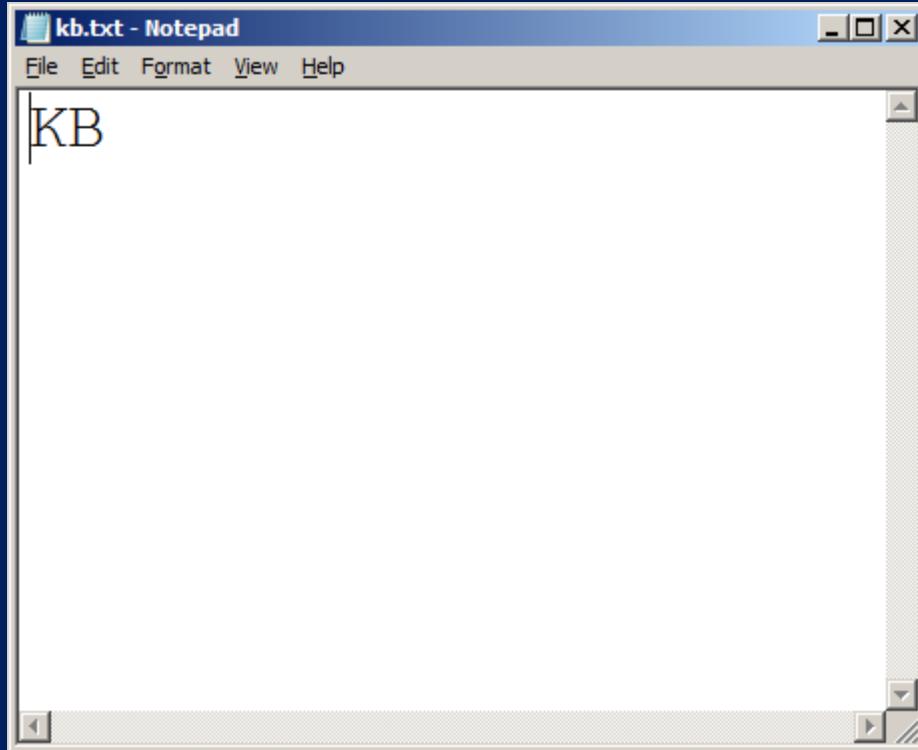
MINGW64:f/Johan/DACH2016

Bytes to text: ASCII

- * 128 characters, 7 bits
- * numbers: 0-9
- * letters: a-z; A-Z
- * punctuation symbols, space
- * non-printable characters

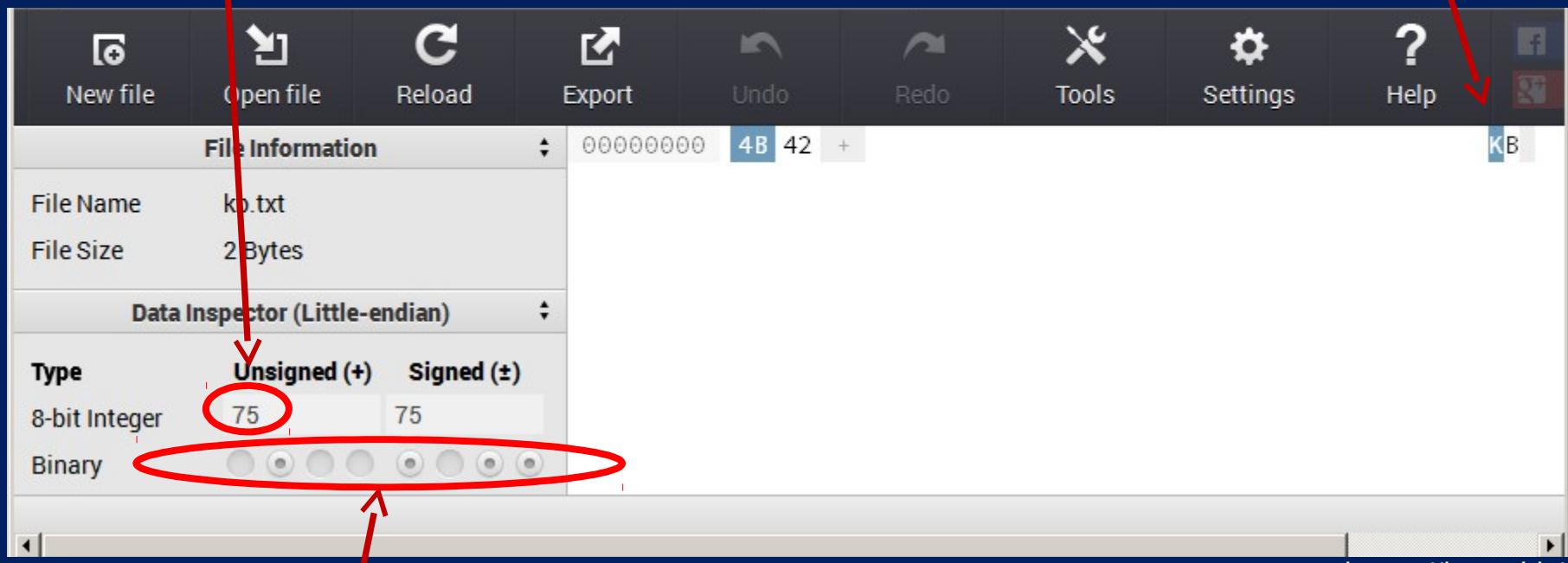
ASCII Table

Bin	Dec	Char
::	::	::
01000001	65	A
01000010	66	B
01000011	67	C
01000100	68	D
01000101	69	E
01000110	70	F
01000111	71	G
01001000	72	H
01001001	73	I
01001010	74	J
01001011	75	K
01001100	76	L
::	::	::



75

K



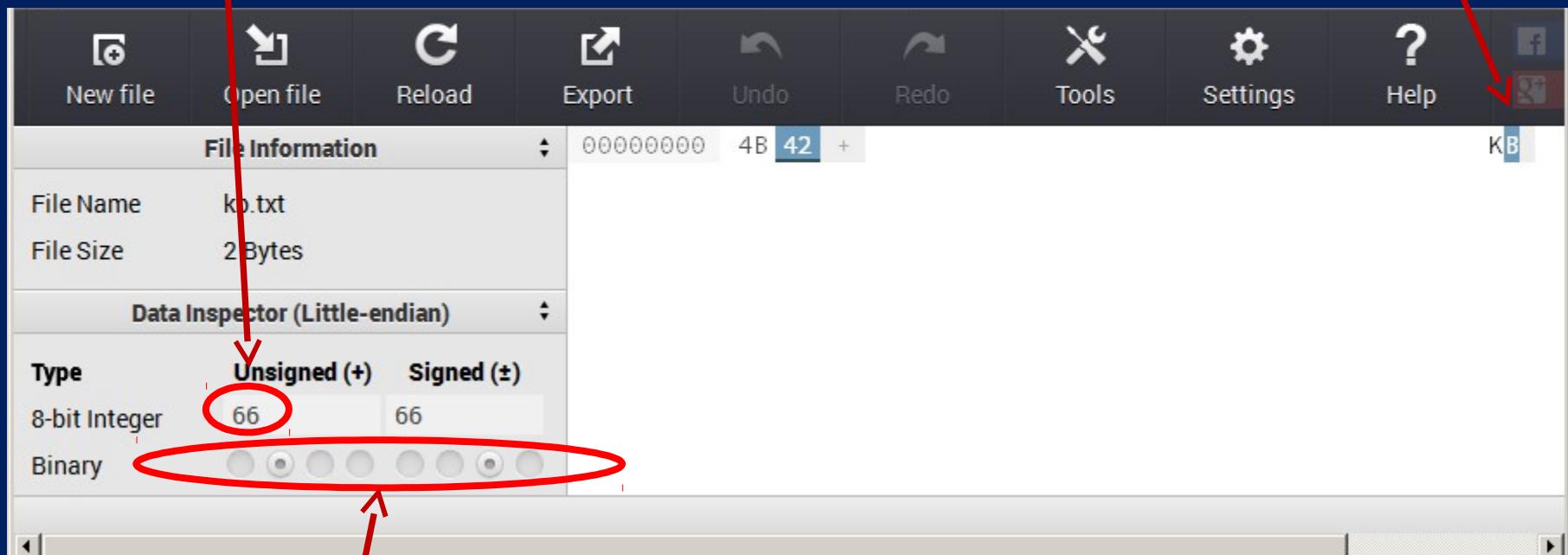
01001011

ASCII Table

Bin	Dec	Char
::	::	::
01000001	65	A
01000010	66	B
01000011	67	C
01000100	68	D
01000101	69	E
01000110	70	F
01000111	71	G
01001000	72	H
01001001	73	I
01001010	74	J
01001011	75	K
01001100	76	L
::	::	::

66

B



01000010

ASCII Table

Bin	Dec	Char
::	::	::
01000001	65	A
01000010	66	B
01000011	67	C
01000100	68	D
01000101	69	E
01000110	70	F
01000111	71	G
01001000	72	H
01001001	73	I
01001010	74	J
01001011	75	K
01001100	76	L
::	::	::

Text-based formats

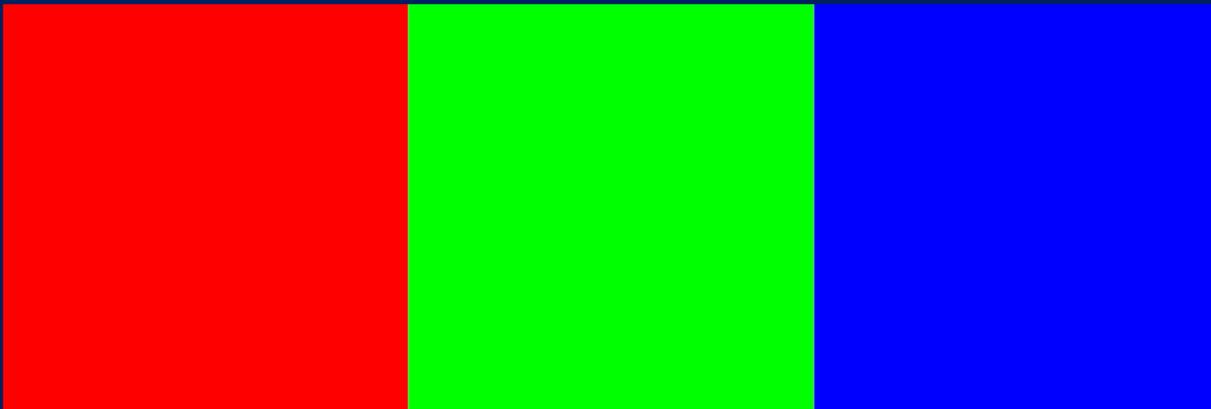
- * Simple text
- * Each character represented by unique (sequence of) byte(s)
- * Viewing / editing possible with any text editor (e.g: Notepad)
- * Examples: ASCII, UTF-8

MINGW64:/f/Johan/DACH2016

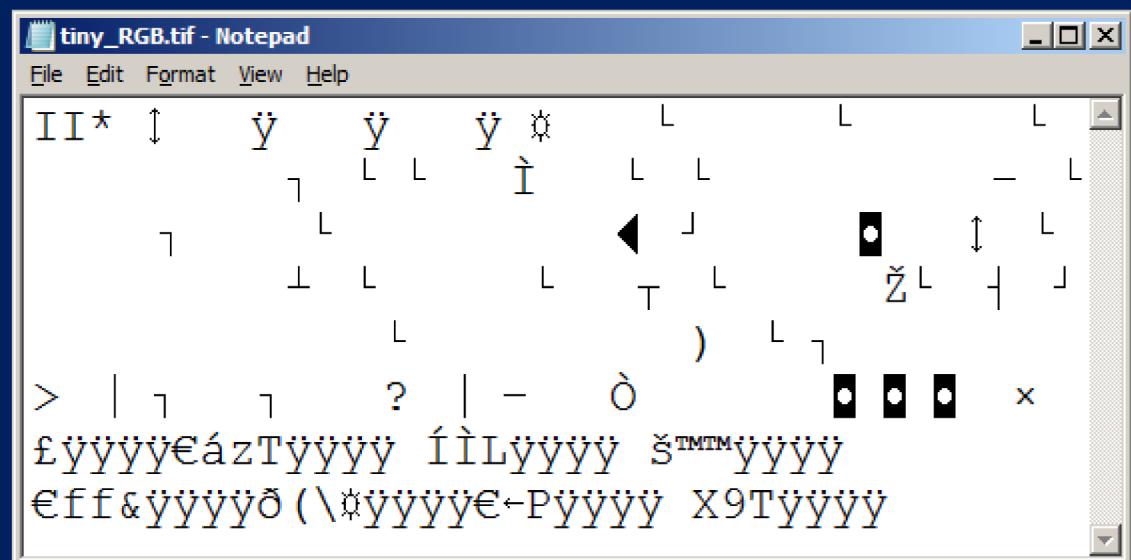
Applications

- * System files (e.g. Windows)
- * Simple data files (e.g.
comma-delimited values)
- * Structured data / text (e.g. XML)
- * Marked up text: (e.g. HTML)

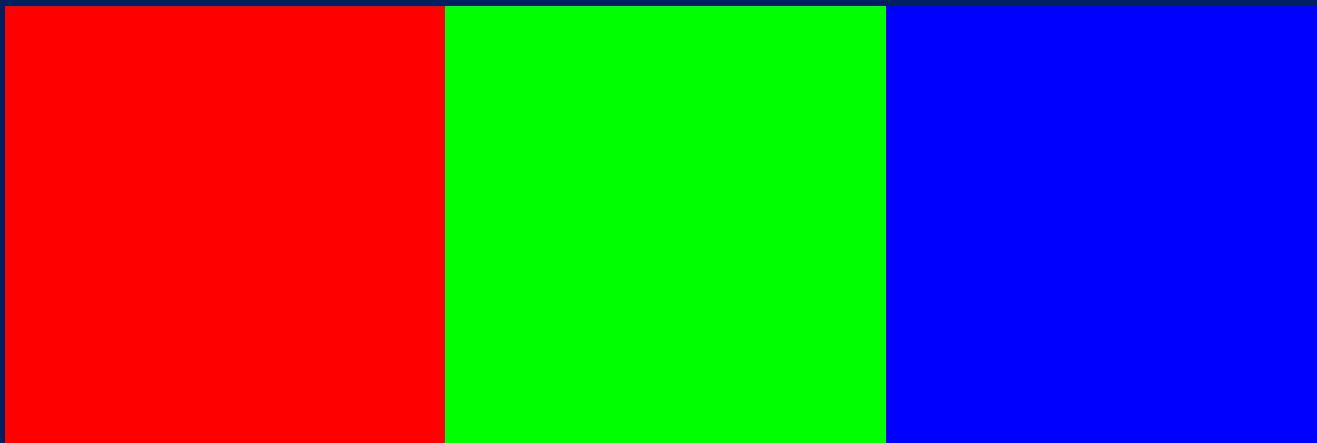
Binary formats



TIFF image



- Pixel: Red, Green, Blue component
- Each component: 8 bits

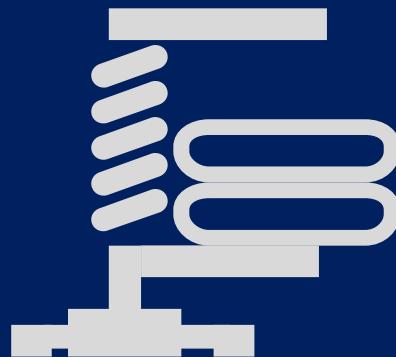
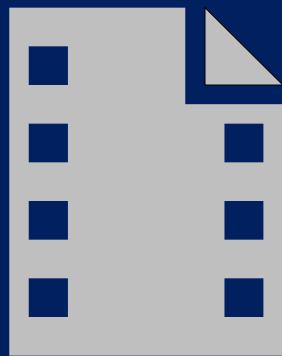
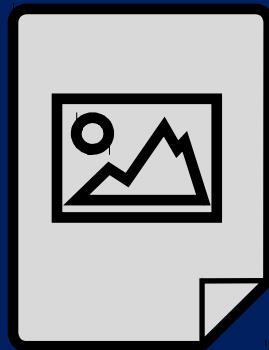


10000000 00000000 00000000 00000000 10000000 00000000 00000000 00000000 10000000

R = 255 G = 0 B = 0 R = 0 G = 255 B = 0 R = 0 G = 0 B = 255

Adapted from Ange Albertini: <https://github.com/corkami>

Applications



Mixed binary / text content

C H A P T E R

7

Detection of Microbial and Viral Contaminants in Cell Lines

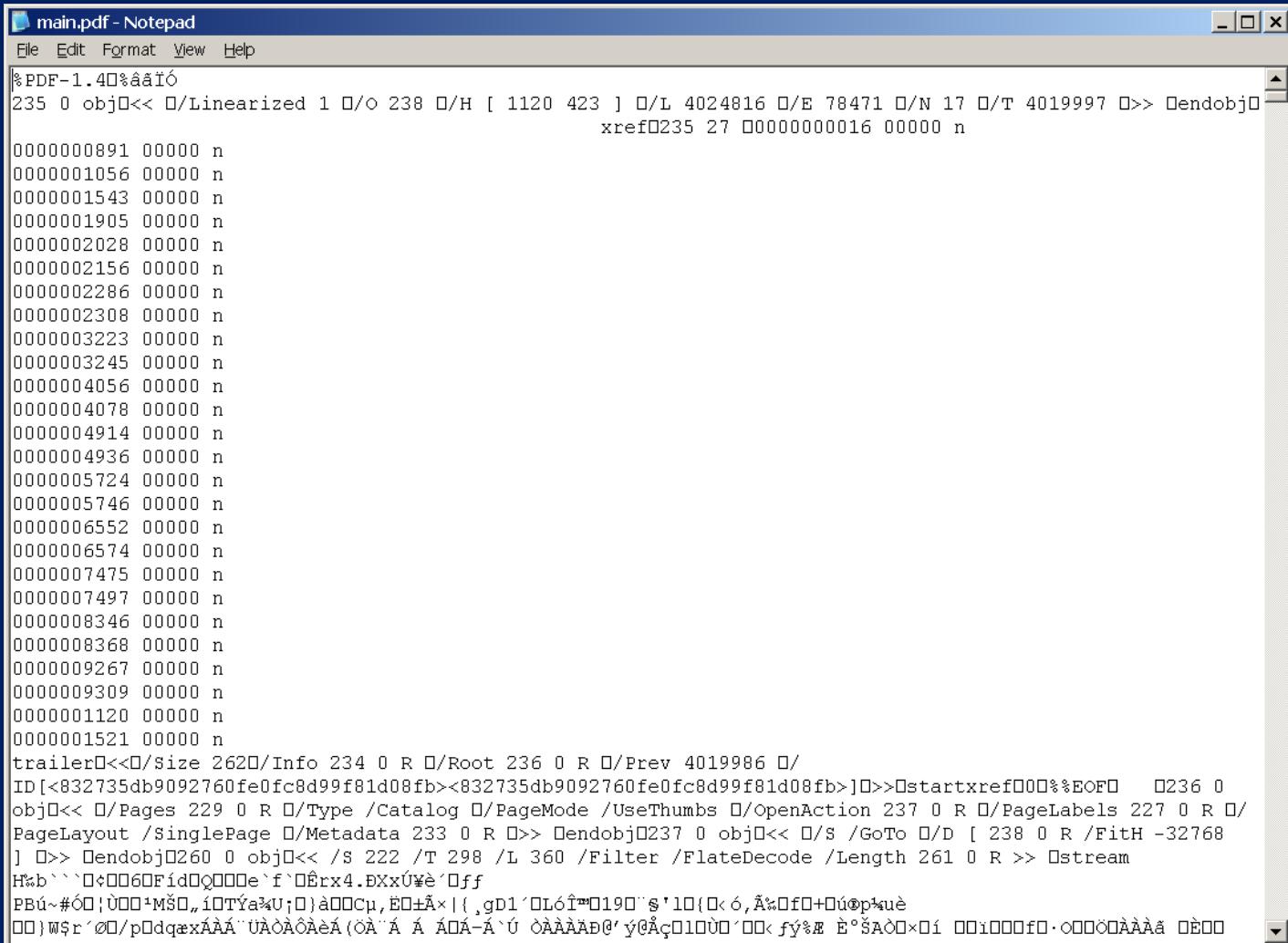
Robert J. Hay and Pranvera Ikonomi

I. INTRODUCTION

The presence of microbial contaminants—bacteria, fungi, mycoplasma, or protozoa—in cell cultures seriously compromises virtually all research or production work involving culture technology. Although many contamination events are overt and readily apparent,

Systems (formerly BBL). Fresh defibrinated rabbit blood (Cat. No. 82-8614) and sheep blood are from Editek, and North American Biologicals provided Diamond's TP-S-1 broth base powder (Cat. No. 73-9502) and Diamond's TP-S-1 vitamin solution (40X, Cat. No. 72-2315). American Hoechst is the source of bisbenzamide fluorochrome stain (Cat. No. 33217), and oil-free, dry annealed aluminum foil is from Reynolds

Same PDF in Notepad



MS Word (docx)

L 1 · 2 · 1 · 1 · 3 · 4 · 5 · 6 · 7 · 8 · 9 · 10 · 11 · 12 · 13 · 14 · 15 · 16 · 17 · 18 ·

Test docx document

This is a section
Here's some text

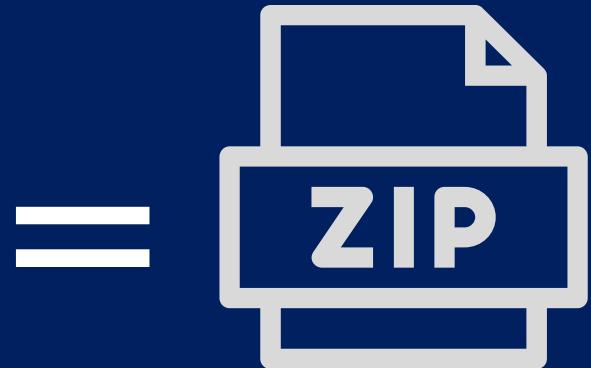
Another section
Some more text. And an image as well:

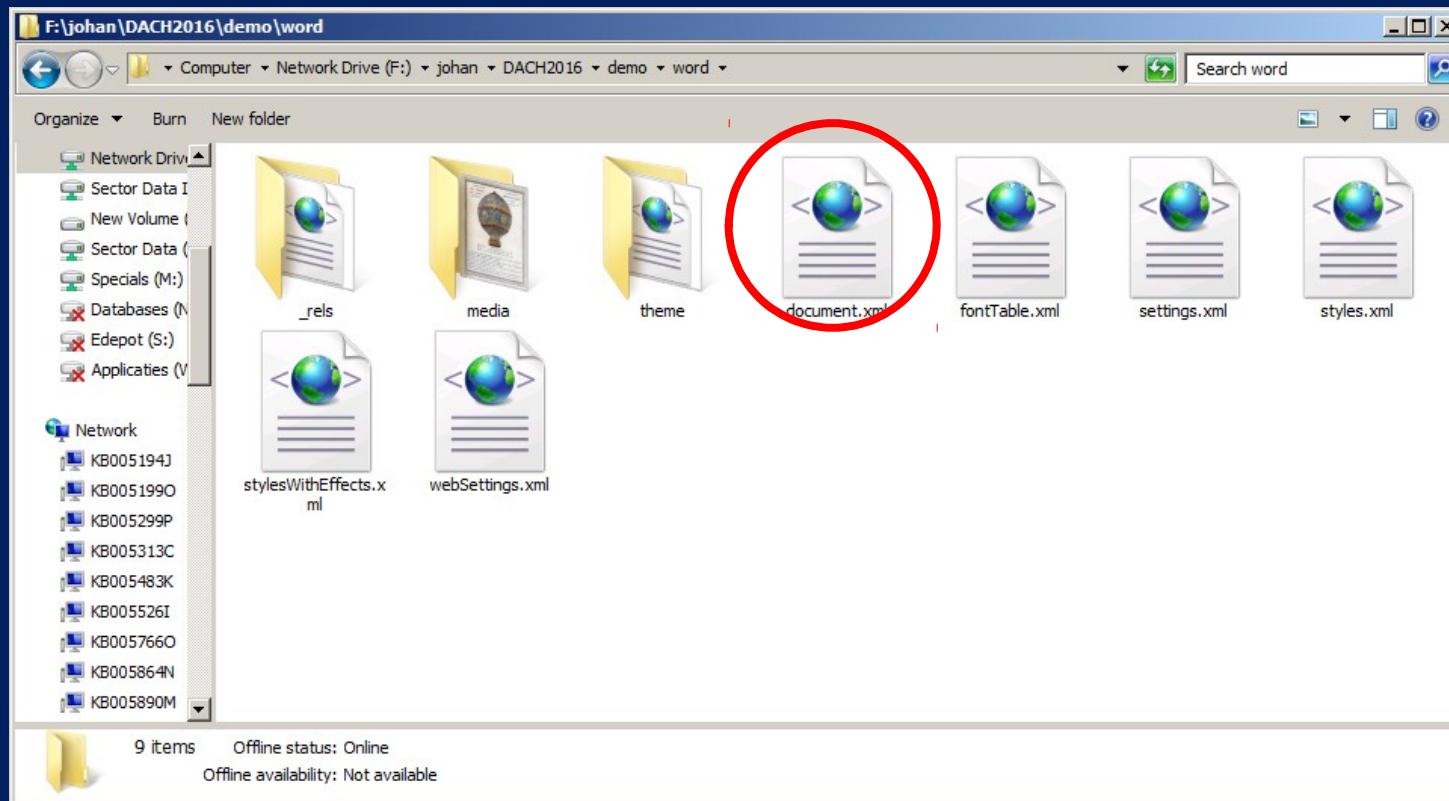
FIGURE EXACTE
DU GLOBE
AÉROSTATIQUE,
Qui, le premier,
des Hommes
a enlevé
dans les Airs.

ET PROPORTIONS.

Le diamètre du Globe..... 70 pieds | Poids du Globe..... 1600 Livres.
Diamètre..... 46 pieds | Poids qu'il a levé de 4000 Livres.
Capacité..... 6000 pieds cubes | La ballon avec 3 pieds de hauteur.

La partie supérieure est entourée de fleurs-de-lis au-dessous des 12 signes du Zodiaque.
Au milieu des chiffres du Roi, entourés de Soleils.
Le bas, écrit grec de Mars, carre et de Céphéus, plusieurs figures à tête égale qui paraissent se référer en tour cette machine.
Tour ces symboles devant de couleur d'or sur un beau fond bleu, entouré que ce corps globe par une bordure d'or et d'azur.
La ballon circulaire, dans laquelle on voit M. le Marquis d'ARLANDES et M. PILATRE de ROZIER, sont peints en Drapierie et armoiries à franges d'or.





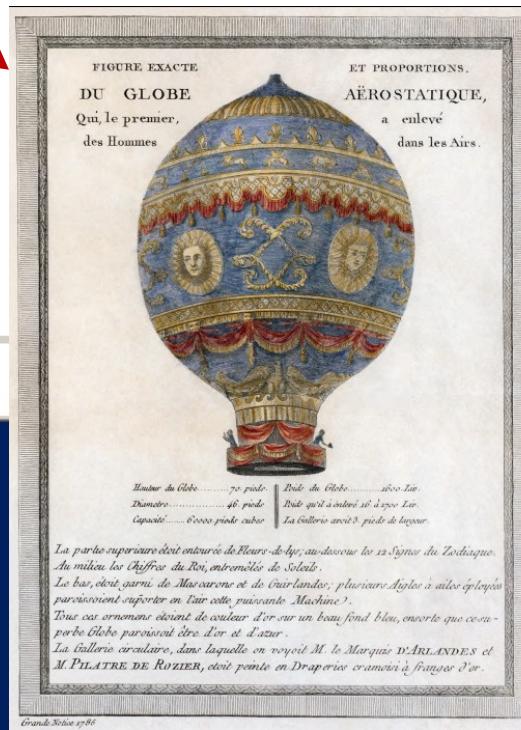
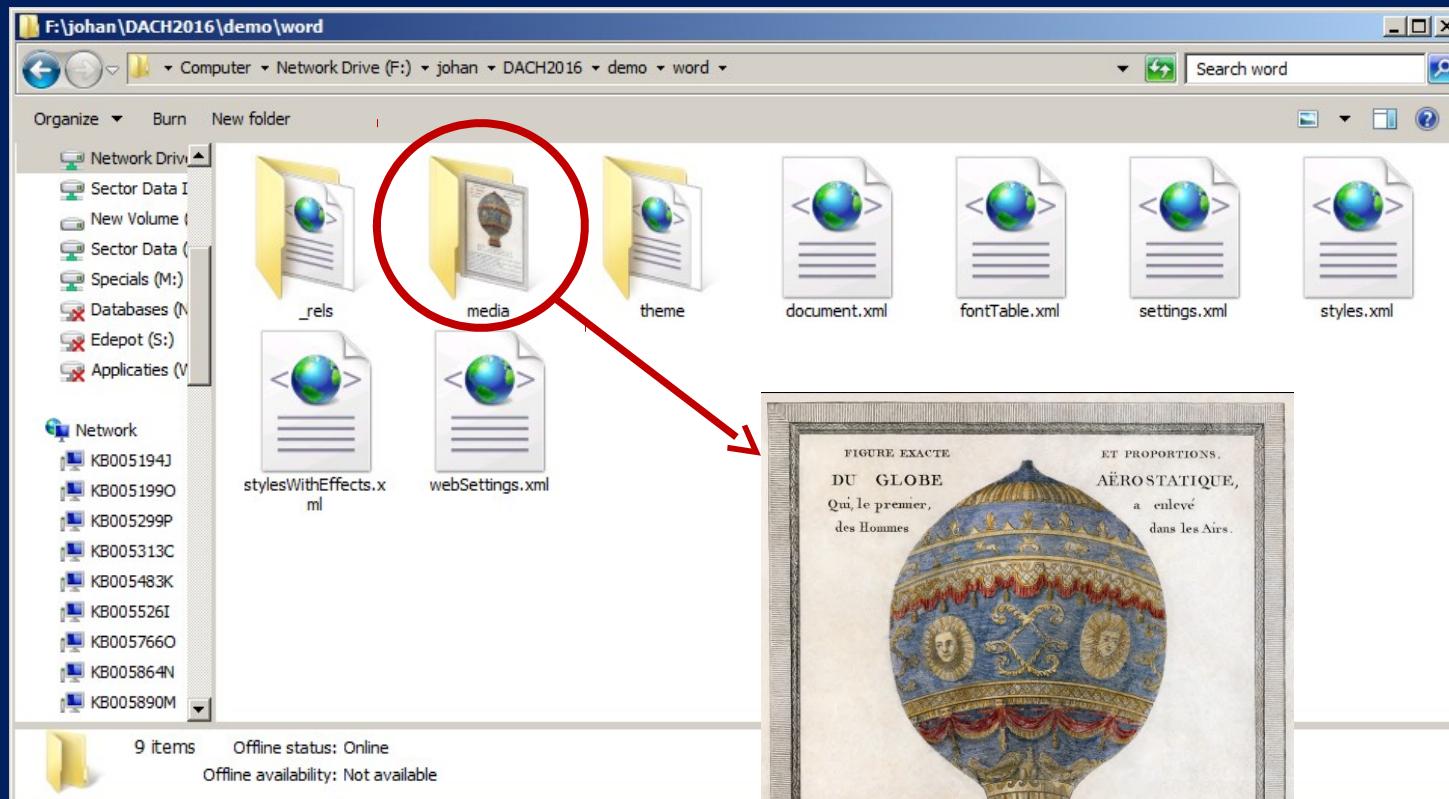
F:\johan\DACh2016\demo\word\document.xml - Notepad++

File Edit Search View Encoding Language Settings Macro Run TextFX Plugins Window ? X

markup.txt ascii.txt tiff.py changedbit.txt balloon.jpg imagecredits.txt offsets.txt

```
"005316A9" w:rsidRDefault="007D2D82" w:rsidP="007D2D82"><w:pPr><w:pStyle w:val="Heading1"/><w:rPr><w:lang w:val="en-GB"/></w:rPr></w:pPr><w:r w:rsidRPr="005316A9"><w:rPr><w:lang w:val="en-GB"/></w:rPr><w:t xml:space="preserve">Test </w:t></w:r><w:r w:rsidR="00FC039C"><w:rPr><w:lang w:val="en-GB"/></w:rPr><w:t xml:space="preserve">docx</w:t></w:r><w:bookmarkStart w:id="0" w:name="_GoBack"/><w:bookmarkEnd w:id="0"/><w:r w:rsidRPr="005316A9"><w:rPr><w:lang w:val="en-GB"/></w:rPr><w:t>document</w:t></w:r></w:p><w:p w:rsidR="007D2D82" w:rsidRPr="005316A9" w:rsidRDefault="007D2D82" w:rsidP="007D2D82"><w:pPr><w:pStyle w:val="Heading2"/><w:rPr><w:lang w:val="en-GB"/></w:rPr></w:pPr><w:r w:rsidRPr="005316A9"><w:rPr><w:lang w:val="en-GB"/></w:rPr><w:t>This is a section</w:t></w:r></w:p><w:p w:rsidR="007D2D82" w:rsidRPr="005316A9" w:rsidRDefault="007D2D82" w:rsidP="007D2D82"><w:pPr><w:rPr><w:lang w:val="en-GB"/></w:rPr></w:pPr><w:r w:rsidRPr="005316A9"><w:rPr><w:lang w:val="en-GB"/></w:rPr><w:t>Here's some text</w:t></w:r></w:p><w:p w:rsidR="007D2D82" w:rsidRPr="005316A9" w:rsidRDefault="007D2D82" w:rsidP="007D2D82"><w:pPr><w:pStyle w:val="Heading2"/><w:rPr><w:lang w:val="en-GB"/></w:rPr></w:pPr><w:r w:rsidRPr="005316A9"><w:rPr><w:lang w:val="en-GB"/></w:rPr><w:t>Another section</w:t></w:r></w:p><w:p w:rsidR="005316A9" w:rsidRDefault="007D2D82" w:rsidP="005316A9"><w:pPr><w:rPr><w:lang w:val="en-GB"/></w:rPr></w:pPr><w:r w:rsidRPr="005316A9"><w:rPr><w:lang w:val="en-GB"/></w:rPr><w:t>Some more text.</w:t></w:r><w:r w:rsidR="005316A9"><w:rPr><w:lang w:val="en-GB"/></w:rPr><w:t xml:space="preserve"> And an image as well:</w:t></w:r></w:p><w:p w:rsidR="005316A9" w:rsidRDefault="005316A9" w:rsidP="005316A9"><w:pPr><w:rPr><w:lang w:val="en-GB"/></w:rPr></w:pPr><w:r w:rsidRPr="007D2D82" w:rsidRPr="005316A9" w:rsidRDefault="005316A9" w:rsidP="005316A9"><w:pPr><w:rPr><w:lang w:val="en-GB"/></w:rPr></w:pPr><w:r w:rsidRPr="007D2D82" w:rsidRPr="005316A9" w:rsidRDefault="005316A9" w:rsidP="005316A9"><w:pPr><w:rPr><w:lang w:val="en-US"/></w:rPr></w:pPr><w:rPr><w:lang w:val="en-US"/></w:rPr><w:drawing><wp:inline distT="0" distB="0" distL="0" distR="0" wp14:anchorId="14444E51" wp14:editId="04278A27"><wp:extent cx="3722400" cy="5068800"/><wp:effectExtent l="0" t="0" r="0" b="0"/><wp:docPr id="1"
```

length : 4561 lines : 2 Ln : 1 Col : 1 Sel : 0 | 0 Dos\Windows ANSI as UTF-8 INS //





W3C®

Portable Network Graphics (PNG) Specification
(Second Edition)

Information technology — Computer graphics and image processing — Portable Network Graphics (PNG): Functional specification.
ISO/IEC 15948:2003 (E)

W3C Recommendation 10 November 2003

This version:
<http://www.w3.org/TR/2003/REC-PNG-20031110>

Latest version:
<http://www.w3.org/TR/PNG>

Previous version:

ITU INTERNATIONAL TELECOMMUNICATION UNION

CCITT
THE INTERNATIONAL
TELEGRAPH AND TELEPHONE
CONSULTATIVE COMMITTEE

T.81
(09/92)

TERMINAL EQUIPMENT AND PROTOCOLS
FOR TELEMATIC SERVICES

INFORMATION TECHNOLOGY –
DIGITAL COMPRESSION AND CODING
OF CONTINUOUS-TONE STILL IMAGES –
REQUIREMENTS AND GUIDELINES

Recommendation T.81

Open vs closed formats



Open formats

- Specifications published and accessible (not always free!)
- Often published by standards body (ISO, ITU, IDPF)
- Specs contain all info needed to build reader



HTML



ePUB

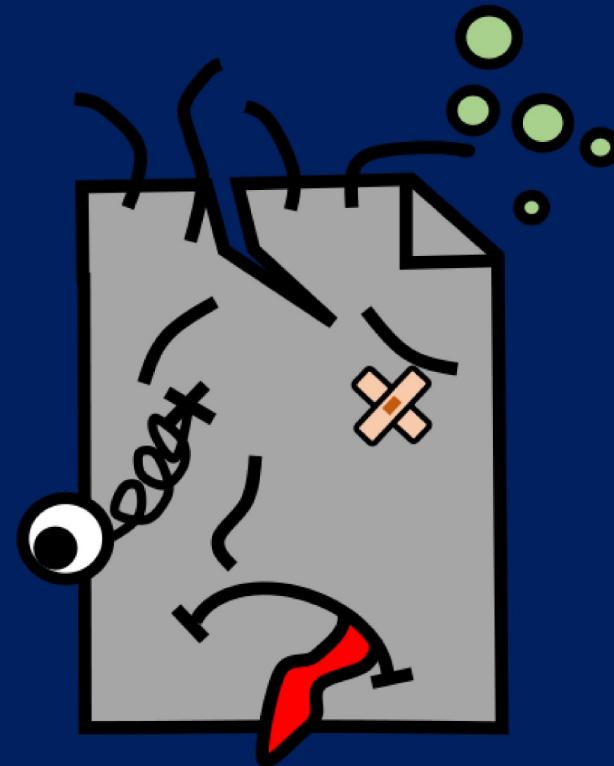
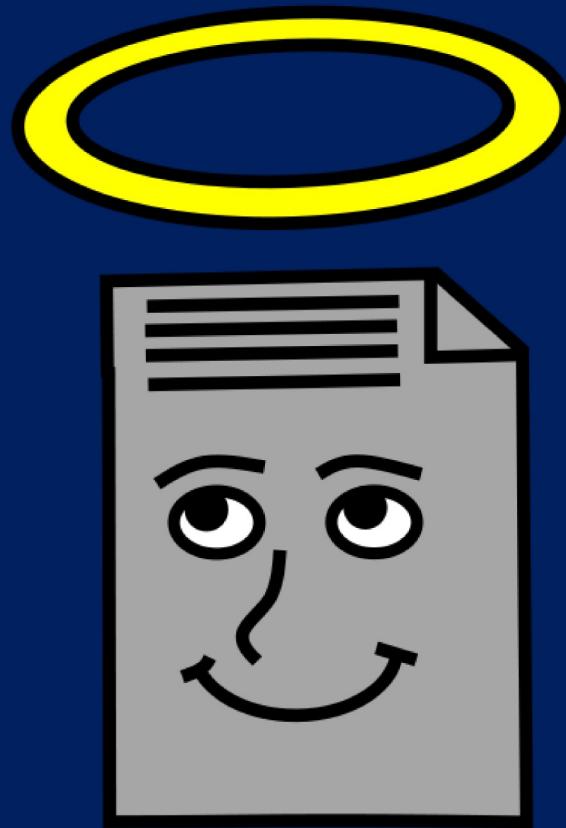


- Dependency on owner format
- Owner may stop offering reader software
- Long-term accessibility uncertain



- WMA
- DOC, XLS, PPT
- DWG (AutoCAD)

Effects of damage and corruption



JPEG

FIGURE EXACTE
DU GLOBE
Qui, le premier,
des Hommes

ET PROPORTIONS.
AÉROSTATIQUE,
a enlevé
dans les Airs.



Hautur du Globe.....	70 pieds.	Poids du Globe.....	1600 Liv.
Diametres.....	46 pieds	Poids qu'il a élevé de 16 à 1700 Liv.	
Capacité.....	60000 pieds cubes	La Gallerie avoit 3 pieds de largeur.	

La partie supérieure éroit entourée de Fleurs-de-lis; au-dessous les 12 Signes du Zodiaque.
Au milieu les Chiffres du Roi, entremêlés de Soleils.

Le bas, éroit garni de Mas carons et de Guirlandes; plusieurs Aigles à ailes éployées
paroisoient suporter en l'air cette puissante Machine².

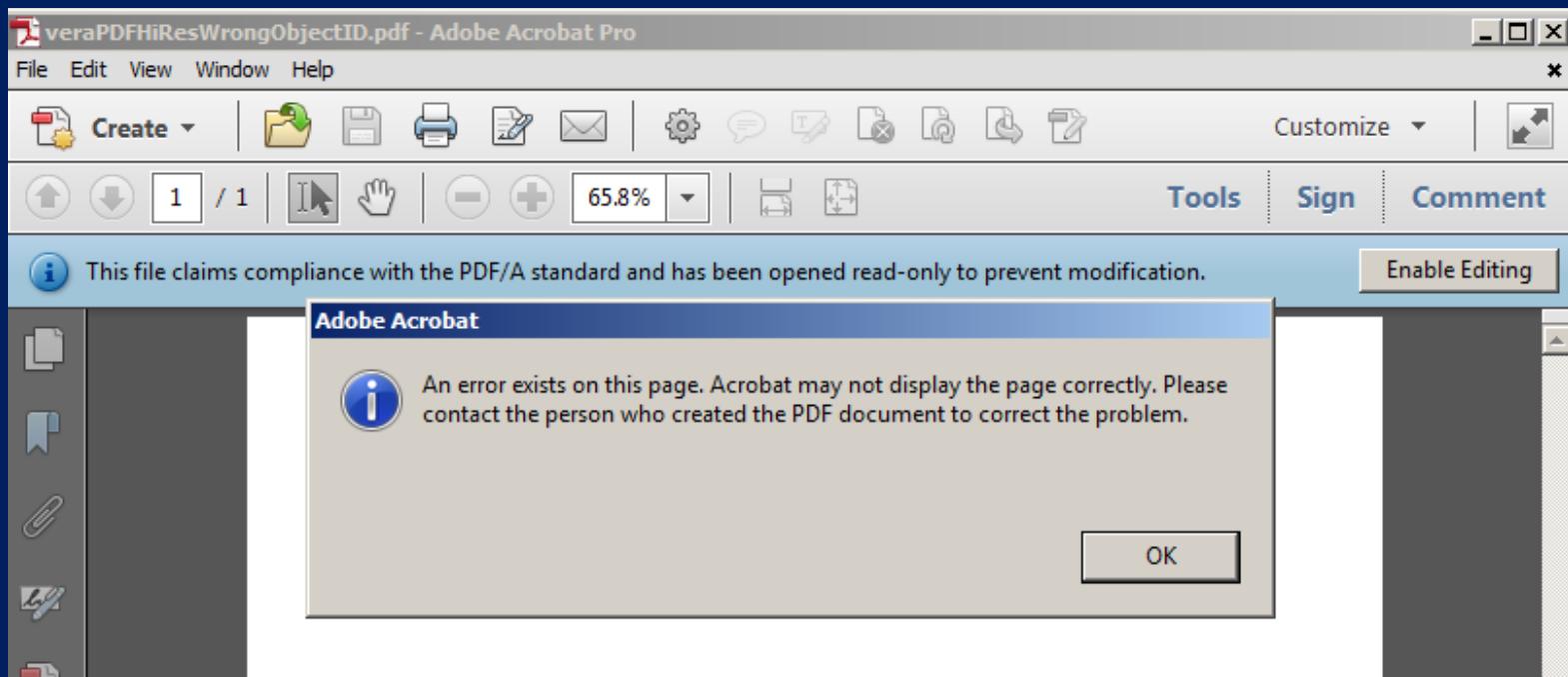
Tous ces ornementz étoient de couleur d'or sur un beau fond bleu, en sorte que ce superbe
Globe paroisoit être d'or et d'azur.

La Gallerie circulaire, dans laquelle on voyoit M. le Marquis D'ARLANDES et
M. PILATRE DE ROZIER, éoit peinte en Draperies cramoisi à franges d'or.



Same JPEG,
1 bit
changed!

PDF



hoewel verre van rendeerdend, ten minste nog op eigen beenen staat, kunnen de producten van den landbouw, deriet- en grondcultuur het lage loon van f 12,— van den grondarbeider nog niet eens opbrengen. Zoo wijst alles erop, dat de aanpassing, die noodig is, vooral moet komen door verlaging van de loonen in de arbeid-intensieve bedrijven met lage arbeidsprestatie.

7. De vraag naar kapitaal.

Ik zou nu nog terugkomen op de moeilijkheden van het kapitaal.

Wij zagen, dat er overvloed was van kapitaal. Dit zou volgens het bestaande inzicht der economie (zie b.v. prof. BORDEWIJK, Grondbeginselen, blz. 256) een zeer gunstig verschijnsel moeten heeten. Er kan volgens dat inzicht nooit te veel kapitaal zijn. Het wordt dan immers — zegt men — op ruime schaal aangeboden en een toenemend aanbod van kapitaal betekent een toenemende vraag naar arbeid, beduidt daalende rente en stijgend loon.

De daling van den rentestandaard treedt inderdaad op; maar in geringere mate dan we zouden verwachten op grond van den geconstateerde overvloed van vast kapitaal.

De geldmarkt, welke de hoogte van den rentestandaard bepaalt, wordt blijkbaar volstrekt niet gehele en zelfs niet grootendeels beheerscht door het tevel of te weinig aan vast kapitaal. De besparingen van het vaste kapitaal — wij zagen het ook in ons voorbeeld — gingen voor een belangrijk gedeelte buiten de geldmarkt om: zij kwamen tot stand uit de winsten en reserveeringen van de bedrijven zelve. Dat geschiedt niet alleen in het baggerbedrijf, het was een kenmerkende eigenschap voor het gehele kapitaal-intensieve grootbedrijf. Omgekeerd wordt het tevel aan vast-kapitaal ook niet op zoondagine wijze op de geldmarkt aangeboden, dat het deze beheerscht. Men bedenke trouwens, hoe gering de waarde is van het overcomplete vaste kapitaal.

De geldmarkt wordt veel meer beheerscht en de rentestandaard op peil gehouden door de vraag naar vloottend kapitaal, handels- en bankredicten en schuldbriefen van openbare lichamen. Hierdoor wordt natuurlijk de productie niet gestimuleerd, de schulden betrekken meestal slechts een verplaatsing van kapitaal. Zij houden echter den rentestandaard op en zo is deze in ons land nog ruim 3½ % terwijl — als het inderdaad waar was, dat de vraag naar het aanbod van productief vast kapitaal de bepaalende aquatiek den rentestandaard bepaalden — deze laatst wel onder 1 % gedaald zou moeten zijn.

Belangrijker nog is echter de tweede reden, dat de vraag naar industrieel kapitaal nog niet betrekkelijk ruime schaal nieuw kapitaal wordt aangeboden en dat een nieuwe voortbrenging.

Volgens PIERSON zou er wel geen prijsverhoging optreden, zulk een prijkelend invloed heeft op de vraag na verlaging van de rente op de vraag naar kapitaal. De prijsverhoging uit ongeveer 1900 blijkt thans wel volledig verdwenen te zijn. De reeds verkregen niet onbelangrijke verlaging deed ten minste de vraag naar productief vast kapitaal nog in het geheel niet herleven. Blijkt dat de hoogte van de rente volstrekt niet meer de bepalende factor te zijn? Veel belangrijker is het risico geworden en daar is de mogelijkheid. Het risico voor nieuwe technisch bedrijf wordt te groot geacht. Het beleggend publiek leeft door pijnlijke ervaringen. Misschien heeft het bewust de wet ontdekt, die uitspreekt, wat de bedrijven moeten zijn van de vrije concurrentie in de kapitaal-intensieve bedrijven, maar op het onderbewustzinnige niveau. Beleggend publiek heeft deze wet toch reeds kennen geleerd.

Wil men dus komen tot nieuwe beleggingen, dan moet men dat zuur noch nodig zijn voor een industrieel bedrijfsvolwassenheid — dan zal men waarborgen moeten geven, dat de aanvankelijke monopoliepositie van het kapitaal in de industrie ook meer duurzaam zal worden beschermd.

Laten wij nog even terugkeeren tot ons voorbeeld. De zandprijs van f 0,15 per m³ met inbegrip van domeinrechten, laat niets over voor rente, afschrijving en winst.

Het is begrijpelijk, dat de eigenaren van de baggermolens niet konden berusten in dezen toestand. De aanvankelijke monopoliepositie van den eersten baggerondernemer kon natuurlijk niet terug worden verkregen, ook de veel kortstondiger monopoliepositie van den eersten grooten molen en van de intussen ingevoerde technische verbeteringen waren niet houdbaar gebleken, maar zou een kartelovereenkomst tusschen alle deelnemers aan het concessiezandbedrijf niet kunnen helpen?

Het heeft nu inderdaad niet onthouden aan pogingen door de eigenaren der baggermolens in één concerto te vereenigen. Tot nu toe duurde dat echter nooit lang. Van een verhogening van den zandprijs beleefde men namelijk niet veel plezier: het gevolg was het ten toonsele verschijnen van nieuwe concurrenten en het op slinksche wijze ontduiken der prijsafspraken door de baggeraars zelve. Daardoor viel de in 1932 opgerichte Zand-Unie weer spoedig uiteen. Thans bestaat er — naar ik meen — opnieuw een overeenkomst. Men heeft nu echter vooroerst den zandprijs nog niet willen verhogen en alleen enkele molens uit het bedrijf genomen om den belastingscoöfficiënt der overblijvende molens te verbeteren. We zullen moeten afwachten of deze nieuwe poging tot samenwerking zal slagen. Uit alles blijkt echter, dat er op den duur een kunstmatig monopolie zal moeten komen, daar alleen hierdoor de op den duur onhoudbare positie van het vaste kapitaal kan worden verbeterd. Soortgelijke ontwikkelingen ziet men in vele kapitaal-intensieve bedrijven. Men voelt de betekenis van de wet der vrije concurrentie in deze bedrijven aan den lijve en wordt wel gedwongen tot het zoeken van een oplossing.

In hoeverre de overheid daarbij moet ingrijpen, wil ik buiten het bestel van mijn beschouwingen laten. Alleen wil ik erop wijzen, dat het vraagstuk zoo moeilijk is, omdat eenerzijds — zoals ook in ons voorbeeld bleek — de industrie zelve veelal onvoldoende in staat blijkt om uit eigen kracht een voldoende samenwerking tot stand te brengen, terwijl anderzijds — om bekende redenen — het ingrijpen van de overheid ongetwijfeld zoo veel en zoo lang mogelijk dient te worden vermeden.

Glukkig gaat het natuurlijke voorsprong-monopolie echter niet overal zoo spoedig te niet als in het technisch bedrijf. Zoals we al gezien hebben, blijft die monopoliepositie in de industrieel kapitaal-intensieve fabrieksbedrijven en in de industrieel kapitaal-intensieve bedrijven.

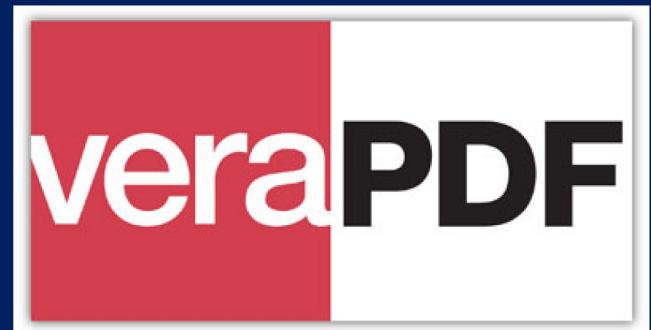
De industrieel kapitaal-intensieve bedrijven, aan de zelfstandigheid van de industrieel organisatie gebonden, blijven inderdaad bestaan. En dat is: wedijver tussen de verschillende industrieën en goedkoop arbeid. De industrieel kapitaal-intensieve bedrijven zijn ovaalgestanst door de vrije concurrentie in de arbeid-intensieve bedrijven. De industrieel kapitaal-intensieve bedrijven zijn ovaalgestanst door de vrije concurrentie in de arbeid-intensieve bedrijven.

De industrieel kapitaal-intensieve bedrijven blijken inderdaad bestaan te kunnen, maar de oudere en goedkoop arbeid blijft veelal een belangrijke factor in de scheepvaart, de industrieel kapitaal-intensieve bedrijven.

De industrieel kapitaal-intensieve bedrijven blijven bestaan, maar de oudere en goedkoop arbeid blijft veelal een belangrijke factor in de scheepvaart, de industrieel kapitaal-intensieve bedrijven. De industrieel kapitaal-intensieve bedrijven blijven bestaan, maar de oudere en goedkoop arbeid blijft veelal een belangrijke factor in de scheepvaart, de industrieel kapitaal-intensieve bedrijven. De industrieel kapitaal-intensieve bedrijven blijven bestaan, maar de oudere en goedkoop arbeid blijft veelal een belangrijke factor in de scheepvaart, de industrieel kapitaal-intensieve bedrijven.

JP2 (JPEG 2000)

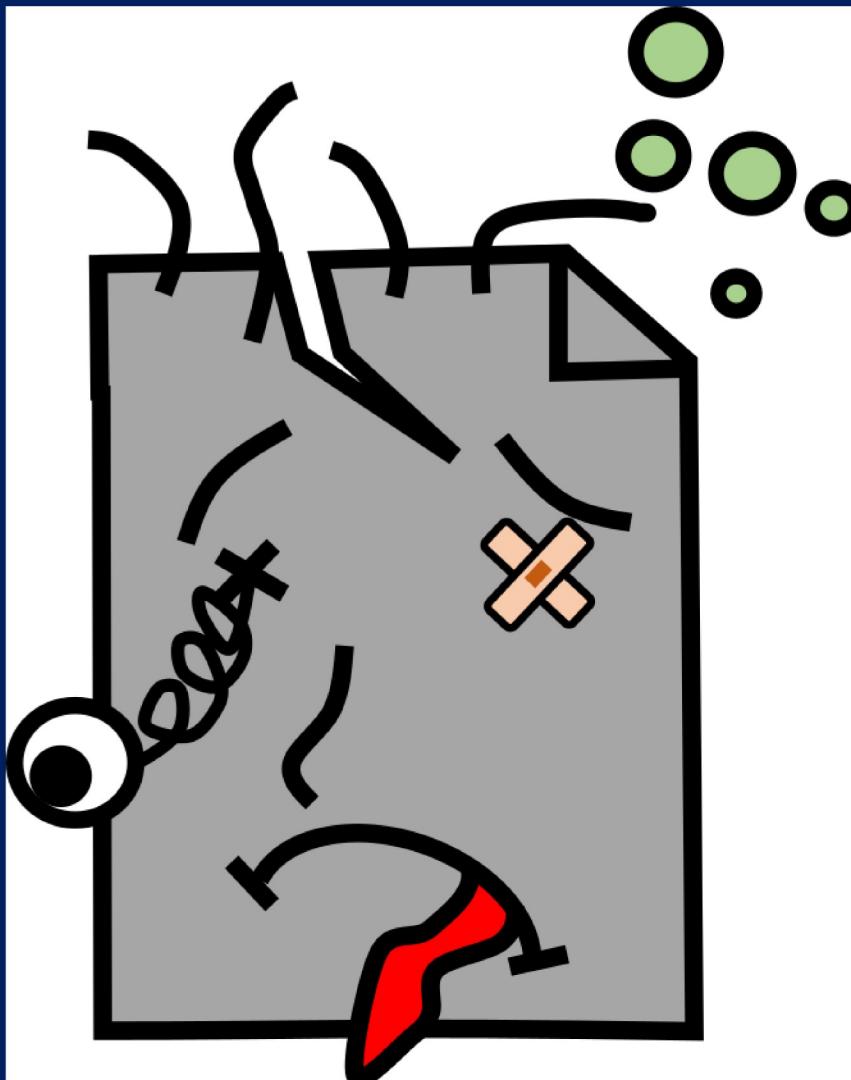
Format validation



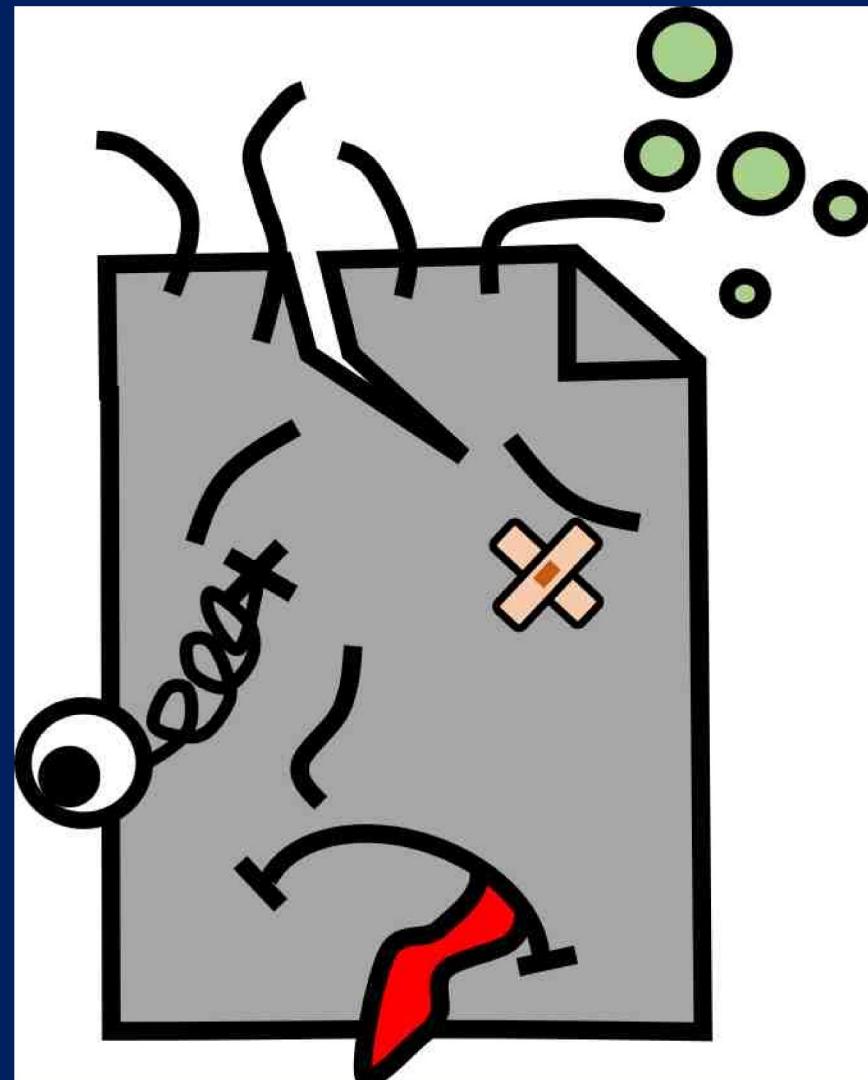
Master formats digitisation

- Quality
- Features (metadata, ICC profiles)
- Cost (file size!)
- Long-term accessibility
- Software support

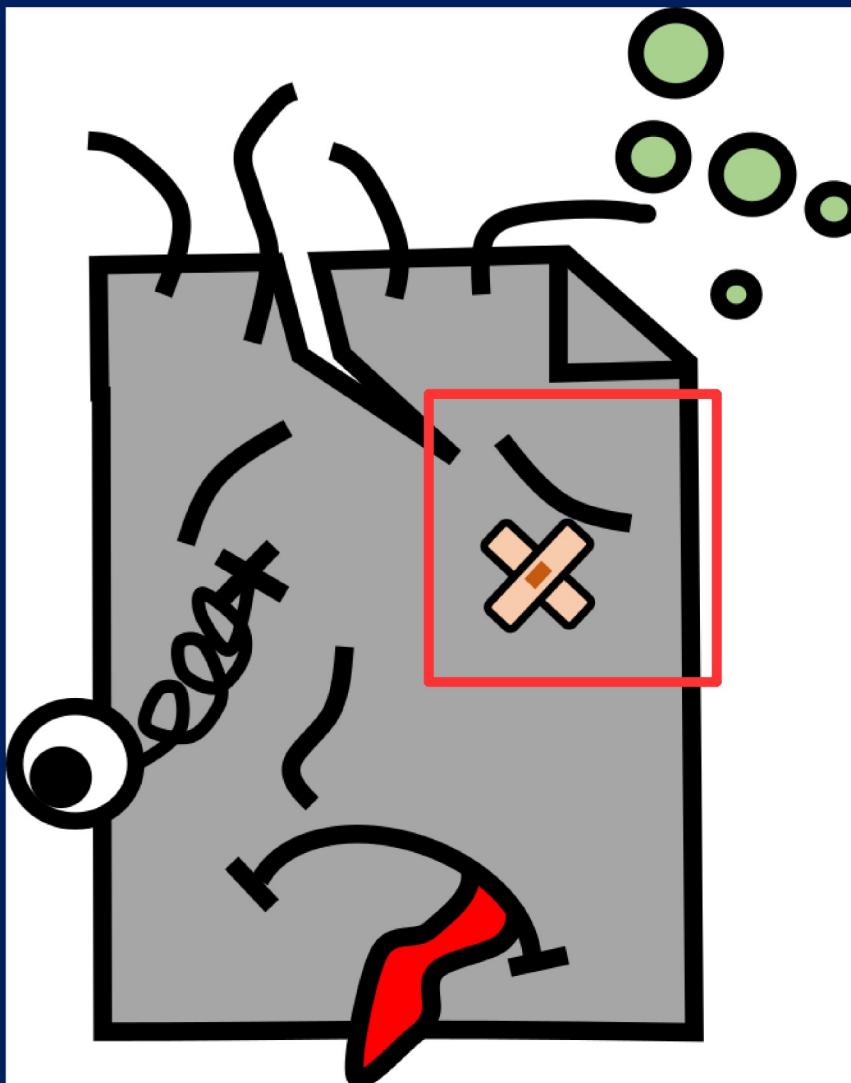
TIFF/ PNG



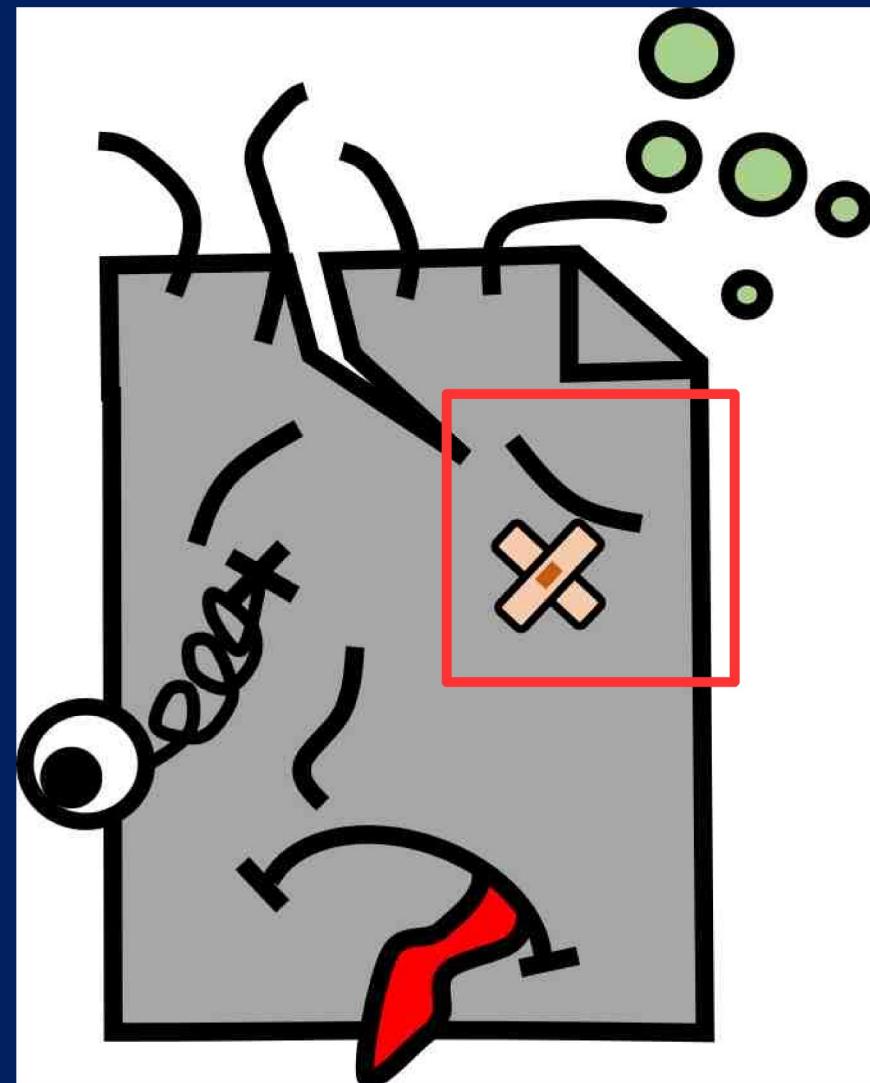
JPEG

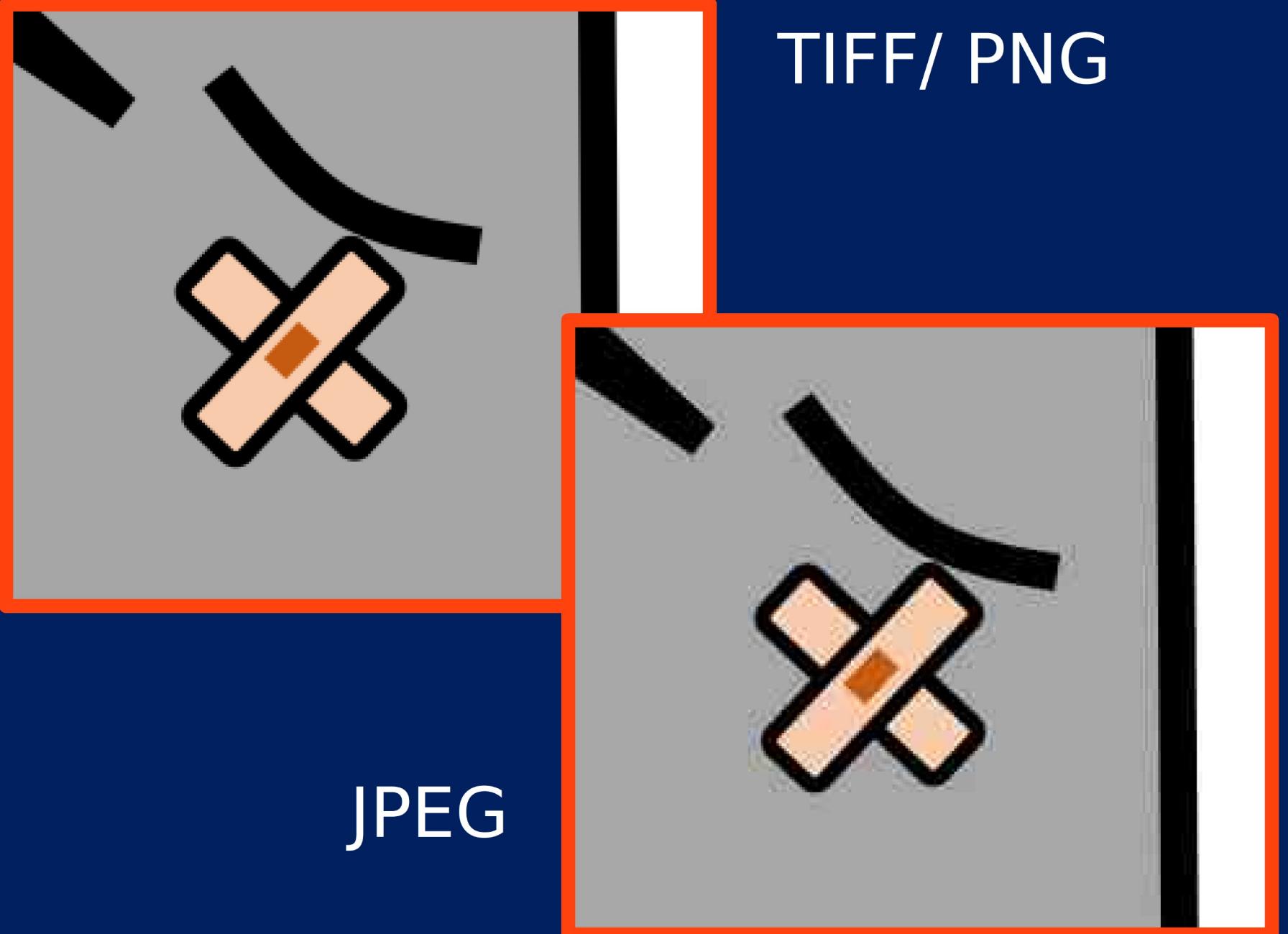


TIFF/ PNG



JPEG





TIFF/ PNG

JPEG

Master formats digitisation

- Uncompressed TIFF (huge files!)
- TIFF, LZW compression
- JPEG 2000 Part 1 / JP2 is on the rise
(but limited software support)

Information on formats

- <http://fileformats.archiveteam.org/>
- <http://www.digitalpreservation.gov/formats/index.shtml>

More links / resources:

- <https://github.com/bitsgalore/DACH2016>

Image credits

Music icon created by Piotrek Chuchla

Video icon created by Arthur Shlain

Image icon created by BenPixels

Compress icon created by Mister Pixel

Zip file icon created by Arthur Shlain

TIFF icon created by To Uyen

All from the Noun Project: <https://thenounproject.com/>

Top Secret by BP22Heber (Own work) [CC BY-SA 3.0],
via Wikimedia Commons