Long-term preservation of JPEG* images in the "Valley of the Clueless" - need for Standards

*JPEG-1 standard
(ITU-T T.81 | ISO/IEC IS 10918-1)

JPEG = JOINT PHOTOGRAPHIC EXPERT GROUP

DR. ISTVÁN SEBESTYÉN
(EX-CCITT SGVIII SPECIAL RAPPORTEUR ON NIC “NEW IMAGE COMMUNICATION” AND FOLLOW-ITU QUESTIONS BETWEEN 1986 – 2000)

NOW, IN 2019
SECRETARY GENERAL OF ECMA INTERNATIONAL IN GENEVA, CH
Ecma International is a small standardization organization in ICT. Located in Geneva, Switzerland, since 1961.

Our advantage:
- high speed
- high efficiency

Using flexible and unique procedures particularly suitable for Internet / Web standards

Our current “best-seller” standard:
- ECMAScript (=JavaScript)

ICT Storage Systems Programming Languages...

Good co-operation with other SDOs (ITU, ISO, IEC, IETF, W3C,...)

For more information just contact us:
www.ecma-international.org
Why this talk?

- To warn the public at large (not you...😊) that regarding JPEG photos they live in the «Valley of the Clueless» (in German: «Tal der Ahnungslosen»)

- Why?: The in talk identified issues have not been solved yet, and the need especially in the non-professional area to take action has became pressing

- It is evident that if no action is taken the „human heritage“ of the ever growing digital media objects may significantly disappear in a few decades

- Important to note that **ALL media types**, including text, graphics, photos, speech, music, videos etc. are also effected and **concerned**

- The public at large is in general unaware of this danger....Actions and help are needed!
“Valley of the Clueless” (“Tal der Ahnungslosen”)

Then in the 1960s-1980s…. “West-TV” reception in the GDR

“Valley of the Clueless”
(“Tal der Ahnungslosen”)

but today regarding the JPEG photos (and other digital docs….)
Challenges facing smaller archives: The case of the Photographic Record and Survey of Sussex

Dr Jaime Kaminski
University of Brighton
<table>
<thead>
<tr>
<th>Date</th>
<th>Name</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>29/4/12</td>
<td>Charles Job, R.P.S.</td>
<td>Neasden, 3 Chestnut Mansions</td>
</tr>
<tr>
<td>1905</td>
<td>A.H. Corder, 77 Buckingham Rd, Brighton</td>
<td>77 Buckingham Rd, Brighton</td>
</tr>
<tr>
<td>26/6/12</td>
<td>J. C. Stannage, Bevly, Eastbourne</td>
<td>77 Buckingham Rd, Brighton</td>
</tr>
<tr>
<td>1906</td>
<td>E. D. Solomon, Shoeburyness</td>
<td>77 Buckingham Rd, Brighton</td>
</tr>
<tr>
<td>1907</td>
<td>A.H. Corder, Brighton</td>
<td>Miss Clewlow, Hove</td>
</tr>
<tr>
<td>April 1907</td>
<td>R.G. Pearson, Hove</td>
<td>A.H. Corder, 77 Buckingham Rd, Brighton</td>
</tr>
<tr>
<td>May 1909</td>
<td>F. W. Solomon, Shoeburyness</td>
<td>T. W. Solomon, Shoeburyness</td>
</tr>
<tr>
<td>1906</td>
<td>E. D. Solomon, Shoeburyness</td>
<td>Miss Clewlow, Hove</td>
</tr>
<tr>
<td>June 1908</td>
<td>R.G. Pearson, Hove</td>
<td></td>
</tr>
</tbody>
</table>
DESCRIPTION
Ebernoe Horn Fair at which a horned sheep is roasted annually. The sheep is eaten afterwards and the horns carried away on a pole by the winners at cricket.

NAME OF PHOTOGRAPHER
Mrs. Packard

MEMBERS OF
Horsham Camera Club

ADDRESS
Horsham House

Use one form for each print. Write clearly. Make description brief. Leave blank.
Basic digitisation setup to digitalize the Photographic Record and Survey of Sussex:

- To make it accessible to wider audience
- Normal Lifespan of photos has been passed
Photo-storage media Lifespan

Source: Code42 Software Inc.
Observations and Questions

- 120-150 years old photos have survived, important “human heritage” data have been saved

- **The functionality of the old archive is exactly the same what is needed for today’s digital photo collections of the ordinary people.**

- But how many photos taken at that time have not survived? How much “human heritage” information got lost? We do not know that. But would be nice to have the potential to view them again, when needed.

- The same is true for the JPEG photos of the past 25 years and for future days.

- Difference: The old photos were from the beginning “valuable” and rare taken by a few Photographers (cameras). Today there is “inflation” of “no cost” JPEG images (and “photographers”).
Further Outline

- Some market figures on the dominance of JPEG-1 images in today’s photography (The digital photos example)
- Is there a future for the JPEG-1 Standard?
- Need for long term archival of huge volume of JPEG-1 images taken by all (incl. public at large)
- Lifetime span of storage media, equipment and relevant system and application software
- “OSI-like” model suits also for long term archival of digital media (incl. images). Some standardization and perhaps regulation is needed.
Share of yearly produced photo-camera-types between 1933-2015 (top of smartphones cut off)
Photo camera types (cont.)

- The data on cameras are from CIPA (Japan) and from TomiAhonen Consulting for the smartphones with cameras.
- Kodak experiment in 1976 came to the conclusion that analog photography is not in danger for a long time.
- The first JPEG compact digital cameras came out around 2000 and have smashed the analogue camera market.
- The smartphones with digital cameras smashed the compact digital camera market around 2010.
Share of photo-camera-types between 1933-2015 (smartphones – yellow included)

Global installed base of cameras still operational is 5.8 Billion units.

Out of those only 4.0 B are in use (as cameras, most that 'are not used' are on mobile phones/smartphones which are used in other ways but not for their camera).

Of the 4 B cameras in use, 440 million (11%) are stand-alone 'traditional' digital cameras and 3.56 B (89%) of all cameras in use on the planet today are on mobile phones/smartphones as cameraphones.

Source: TomiAhonen Consulting
Yearly Smartphone Camera production growth saturated in 2018 (1.56 billions were sold)

Sources: TomiAhonen Consulting; Statista 2019
Photocameras picture statistics estimations

- The average stand-alone digital camera user takes 375 pictures in a year while the average cameraphone user snaps 259 pictures a year.

- When multiplied across the total user bases, that produces 1 Trillion (1,000 Billion) photographs taken per year by digital camera owners.

- That brings humankind's cumulative picture production total to 5.7 Trillion photographs taken since the first camera was invented.”

(All stats are from TomiAhonen Phone Book 2014).
Future expectations

- What is not told in the statistics, but it is obvious that close to 100% of the pictures taken by digital cameras have been using JPEG-1. The rest is marginal (including the analogue pictures).

- As a result the many trillions of JPEG images created and yet to be created stay here (except those that will not survive long-term storage – another open issue) – , and from the practical point of view it is hard to imagine that those JPEG-1 images could be mass-converted to a new “post-JPEG” format even if that new method was superior.

- Consequently JPEG-1 – which apparently today fully satisfies average user demand – is expected to stay here also for picture capturing at least for a few more decades.

- This also means that a potential successor of JPEG-1 in the consumer area must be either backwards compatible to JPEG-1 or it has to implement two parallel compression and coding methods including JPEG-1.
Future expectations and conclusions

- To **decode and present JPEG-1 pictures** will *always be* a requirement in the future.

- JPEG-1 is a standard that satisfies an eternal basic human need, namely **the preservation of human heritage**, like writing, drawing etc.

- Its difference is that JPEG images are **digital** and not **analog**, which imposes additional challenges. These must be handled if mankind does not want to lose its human heritage in the future.

- To ensure that must probably be a governmental (?) task and not just leave it e.g. to market forces or just to neglect it. E.g. by the creation of “National Digital Archives” (?) who not only archive long-term selected picture materials but also provide regulations for standards and ensure equipment and software stability to decode and present the images at any time in the future.
What are the standardization challenges for a better solution?

1. We need a standardized storage medium that lasts “forever” (well for 500-1000 years…) – some proprietary solutions exist (e.g. Milleniata optical storage)

2. We need a standardized model of an image retrieval- and presentation system that lasts “forever” and which is resilient to technological changes (hardware, systems software) over time

3. Implement that medium and model in 1. and 2. and provide long-term solution with them.

Note: In what is described later is just an idea of the author, not an Ecma committee work…
1. Current Computer-Media Lifespan – not enough

Source: Code42 Software Inc.
Some Disadvantages:

• What if power failure, network failure, cyberattack, etc.
• Resource intensive storage (energy intensive)
• In practice limited: how many media items one will put on the cloud?
• It is not “forever”. One has to take action by “data migration” periodically.
Additional considerations: Environmental Comparison per Media type

50TB Solution – Power Comparison

[kW/h vs. tons]

CO₂ Emissions (tons) for 1PB in 1year

- HDD
- Tape
- Blu-ray Optical

• Source: Report of U.S. Environmental Protection Agency
• Source: Japanese Gov. MIC report ‘08
2. We need a model for the archival of images of the „public at large“

- The OAIS reference model is for „professional use“ (e.g. National archives) and not for the archival of digital images for „non-professional use“ of the „public at large“

- However, the needed functionalities are similar

- Task: To define a reference model for the images of the public at large:
  - It has to be much simpler, suitable for daily use by people and small service shops
  - It has to be much more robust, error resilient, error corrective and image recovery supporting and take into account the sometimes lower knowledge level and non-expert behaviour of people
Model of a paper archive
(all layers always function – over the total lifetime of the archive)

<table>
<thead>
<tr>
<th>Layers</th>
<th>Macro-layers</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Application Layer”</td>
<td>Human brain</td>
</tr>
<tr>
<td>(Image semantic processing)</td>
<td>(Image semantic</td>
</tr>
<tr>
<td></td>
<td>processing)</td>
</tr>
<tr>
<td>“Image Restoring Layer”</td>
<td>Human brain</td>
</tr>
<tr>
<td></td>
<td>(Image Syntax</td>
</tr>
<tr>
<td></td>
<td>processing)</td>
</tr>
<tr>
<td>“Visual Reception Layer”</td>
<td>Human eye</td>
</tr>
<tr>
<td>“Store &amp; Retrieve Layer”</td>
<td>Image Boxes &amp;</td>
</tr>
<tr>
<td></td>
<td>Image “Metadata”</td>
</tr>
<tr>
<td>“Presentation Layer”</td>
<td>Photo material</td>
</tr>
<tr>
<td></td>
<td>chemical components</td>
</tr>
<tr>
<td></td>
<td>(dots)</td>
</tr>
<tr>
<td>“Physical Layer”</td>
<td>Photo material</td>
</tr>
<tr>
<td></td>
<td>paper</td>
</tr>
</tbody>
</table>
Model of a digital archive
(only few layers function all the time – the digital archive is not really time resistant – this needs remedy)

<table>
<thead>
<tr>
<th>Layers</th>
<th>Macro-layers</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Application Layer” (Image semantic processing)</td>
<td>Human brain (Image semantic processing)</td>
</tr>
<tr>
<td>“Image Restoring Layer”</td>
<td>Human brain (Image Syntax processing)</td>
</tr>
<tr>
<td>“Visual Reception Layer”</td>
<td>Human eye</td>
</tr>
<tr>
<td>“Visual Presentation Layer”</td>
<td>Image Display &amp; Print (Variable)</td>
</tr>
<tr>
<td>“Application Layer”</td>
<td>Image Presentation Program (Variable)</td>
</tr>
<tr>
<td>“Presentation Layer”</td>
<td>JPEG-1 Standard Code/decode (is FIX)</td>
</tr>
<tr>
<td>“Image Store &amp; Retrieve Layer”</td>
<td>Image Files &amp; Directories (Variable)</td>
</tr>
<tr>
<td>“Media Access Layer”</td>
<td>HW/SW Device Drivers (Variable)</td>
</tr>
<tr>
<td>“Physical Layer” (write, read operation)</td>
<td>Digital Media (floppy, CD, etc.) (must be FIX-ed)</td>
</tr>
</tbody>
</table>
Standardization tasks:

- Ensure that on the **Presentation Layer (JPEG-1 coded)** all data **remains intact** over the lifetime of archival on **Physical Layer** by
  - E.g. New media types (like Milleniata-like optical storage media, Holographic storage media...)
  - Data Migration method - periodical rewriting of the archive

- Ensure that appropriate Application Layer and Lower Layers **are replaceable** and **appropriate implementation are available** according to the requirements of archival and media retrieval e.g. By standardized hardware / software profiles. This may change over the years with the advancement of ICT technologies. To be standardized are the **interfaces between layers and communication protocols between them** (similar to the OSI model)
Regulatory and service tasks(?):  

- This is to be discussed...
- These might require also some sort of Governmental Regulation and perhaps special Services(?) e.g. Selection and implementation of archival profiles, maybe providing service centers for image conversions. Not clear if such service functions can just be left just to market forces.
Conclusion: Solutions and standards for long-term media archival is needed now to save “human heritage data”.... Let us do it.... THANKS....

Source: Original test images of JPEG for the tested 4 progressive stages