

# Současné využití formátu JPEG2000 pro obrazové dokumenty v zahraničí

Natalie Ostráková, Národní knihovna ČR



# JPEG2000

- JPEG Committee -příprava od 1996, zveřejněn v roce 2000
- Motivace
  - nárůst dat → výkonnější kompresní metoda-kompresse další generace
  - ztrátová a bezztrátová komprese
  - vyšší odolnost vůči přenosovým chybám
  - JPEG někdy nestačí (mobilní aplikace, lékařské diagnostické aplikace)
  - původně byl ale JPEG2000 plánován jako formát pro digitální fotoaparáty



# Národní digitální knihovna

- JPEG2000 Part 1 použit jako archivační i zpřístupňující formát
- Archivační kopie je bezztrátová, zpřístupňující ztrátová
- Zpřístupnění přes Image server který generuje obrazy JPEG
- Pro konverzi z TIFF do JPEG2000 se používají Kakadu a OpenJpeg

# Doporučení

- Kongresová knihovna- „[Recommended Formats Statement](#)“
- FADGI – „[Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Files](#)“

| Preferred |  |
|-----------|--|
| 1.        | <a href="#">TIFF (*.tif)</a>                 |
| 2.        | <a href="#">JPEG2000 (*.jp2)</a>             |
| 3.        | <a href="#">PNG (*.png)</a>                  |
| 4.        | <a href="#">JPEG/JFIF (*.jpg)</a>            |
| 5.        | <a href="#">Digital Negative DNG (*.dng)</a> |
| 6.        | <a href="#">BMP (*.bmp)</a>                  |
| 7.        | <a href="#">GIF (*.gif)</a>                  |

## Bound Volumes: General Collections

Performance Level:

|                     | 1 Star                 | 2 Star                               | 3 Star                               | 4 Star                         |
|---------------------|------------------------|--------------------------------------|--------------------------------------|--------------------------------|
| Master File Format  | TIFF, JPEG 2000, PDF/A | TIFF, JPEG 2000, PDF/A               | TIFF, JPEG 2000, PDF/A               | TIFF, JPEG 2000, PDF/A         |
| Access File Formats | All                    | All                                  | All                                  | All                            |
| Resolution          | 150 ppi                | 300 ppi                              | 300 ppi                              | 400 ppi                        |
| Bit Depth           | 8                      | 8                                    | 8 or 16                              | 8 or 16                        |
| Color Space         | Grey Gamma 2.2<br>SRGB | Grey Gamma 2.2<br>SRGB<br>Adobe 1998 | Grey Gamma 2.2<br>SRGB<br>Adobe 1998 | SRGB<br>Adobe 1998<br>ProPhoto |

- [KOST-CECO](#) (Švýcarsko) - [Katalog archivischer Dateiformate](#) a [tabulka](#)

| Kriterien   | Gewicht | Formate     |             |             |             |             |             |
|---|---------|-------------|-------------|-------------|-------------|-------------|-------------|
|   |         | Bild        |             |             |             |             |             |
|   |         | TIFF        | JPEG        | JPEG2000    | PNG         | DNG         | PDF/A-2     |
| <b>Otevřenost</b>   | 1       | 4           | 4           | 4           | 4           | 4           | 4           |
| <b>Patent.situace</b>   | 1       | 3           | 4           | 3           | 4           | 3           | 4           |
| <b>Rozšíření</b>  | 1       | 4           | 4           | 2           | 2           | 3           | 3           |
| <b>Funkcionalita</b>  | 1       | 4           | 1           | 4           | 3           | 4           | 3           |
| <b>Implementace</b>   | 0.5     | 3           | 4           | 3           | 4           | 3           | 3           |
| Speicherdichte  | 0.5     | 2           | 4           | 3           | 3           | 2           | 3           |
| <b>Validovatelnost</b>  | 0.5     | 4           | 4           | 4           | 3           | 3           | 4           |
| Komplexitát   | 0.5     |             |             |             |             |             |             |
| Selbstdokumentation   | 0.5     |             |             |             |             |             |             |
|   |         |             |             |             |             |             |             |
| Best Practice   | 0.5     | 3           | 2           | 3           | 1           | 3           | 3           |
| Perspektive   | 0.5     | 2           | 1           | 4           | 2           | 2           | 3           |
|   |         |             |             |             |             |             |             |
| <b>Formatklasse</b>   |         |             |             |             |             |             |             |
| A: Formát je známý a rozšířený, existuje velké množství souborů v tomto formátu |         | A           | A           | C           | C           | C           | C           |
| C: Formát s nadějnou budoucností, je široce používán                            |         | TIFF        | JPEG        | JPEG2000    | PNG         | DNG         | PDF/A-2     |
|   |         | 0.85        | 0.79        | 0.83        | 0.75        | 0.79        | 0.85        |
|   |         | <b>1.30</b> | <b>0.89</b> | <b>1.13</b> | <b>0.73</b> | <b>0.89</b> | <b>1.30</b> |

# JPEG2000 ve světě- předcházející průzkumy



# UConn Survey

- 2008
- odpovědi od 175 institucí (knihovny, státní agentury, galerie, školy, muzea, vládní archiv..)
- 59,5% (103 inst) formát JPEG2000 používalo
- Jako archivační jej používalo 19,7 % (28 inst)
- Cca 50% jej používalo jako formát pro online přístup
- Důvody proč nepoužívají: malá podpora v sw, nedostatek zkušeností u personálu, patentové nejasnosti .....+ špatné PR



# Průzkum formátových politik knihoven ARL

- Sběr dat 2012-2013 od 118 členů ARL (Association of Research Libraries; severoamerické a kanadské knihovny)
- Nejčastěji používané formáty: TIFF (115), WAV (80), PDF, JPEG(70), TXT



- 15 nejpoužívanějších formátů:

| Format Name                   | Extension       | Format Type          | Medium Confidence | High Confidence | Total occurrence in policies |
|-------------------------------|-----------------|----------------------|-------------------|-----------------|------------------------------|
| Tagged Image File Format      | TIFF, TIF       | image                | 27                | 88              | 115                          |
| Waveform Audio File Format    | WAV             | audio                | 33                | 47              | 80                           |
| Portable Document Format      | PDF             | document             | 25                | 49              | 74                           |
| JPEG                          | JPG, JPEG       | image                | 28                | 42              | 70                           |
| Plain text document           | TXT, ASC        | document             | 17                | 52              | 69                           |
| JPEG 2000                     | JP2             | image                | 23                | 45              | 68                           |
| Quicktime                     | MOV, QT         | video                | 47                | 20              | 67                           |
| Extensible Markup Language    | XML             | document             | 16                | 47              | 63                           |
| Audio Interchange File Format | AIF, AIFC, AIFF | audio                | 14                | 45              | 59                           |
| Portable Network Graphics     | PNG             | image                | 14                | 39              | 53                           |
| Graphics Interchange Format   | GIF             | image                | 22                | 31              | 53                           |
| HyperText Markup Language     | HTML, HTM       | document             | 18                | 33              | 51                           |
| Microsoft Word                | DOC             | document             | 38                | 8               | 46                           |
| Microsoft Powerpoint          | PPT             | presentation         | 38                | 6               | 44                           |
| Microsoft Excel               | XLS             | spreadsheet/database | 39                | 5               | 44                           |

- Autoři na základě údajů spočítali důvěru ve formáty

| File Format Name            | Extension | Relative Confidence |
|-----------------------------|-----------|---------------------|
| Tagged Image File Format    | TIFF, TIF | 53%                 |
| Portable Network Graphics   | PNG       | 47%                 |
| JPEG 2000                   | JP2       | 32%                 |
| JPEG                        | JPG, JPEG | 20%                 |
| Graphics Interchange Format | GIF       | 17%                 |
| Bitmap image file           | BMP       | -88%                |
| Photo CD                    | CD        | -100%               |

# Succeed Survey

- 2014
- Průzkum existujících doporučení (17)
- Data od New York State Archives, BnF, NARA, California Digital Library, NISO Framework of Guidance for Building Digital Collections....
- JPEG2000 je celkem populární-vyskytuje se v 53% doporučení
- Ale jen 14% jej užívá pro archivaci
- Limity formátu: složitost, málá podpora v sw-> alternativní formát k TIFF

# Výsledky průzkumu

Formáty, které s vyskytují v doporučeních pro archivaci obrazových dat

- TIFF „de facto“ standard
- JP2 alternativa k TIFF
- PNG není nejlepší volbou

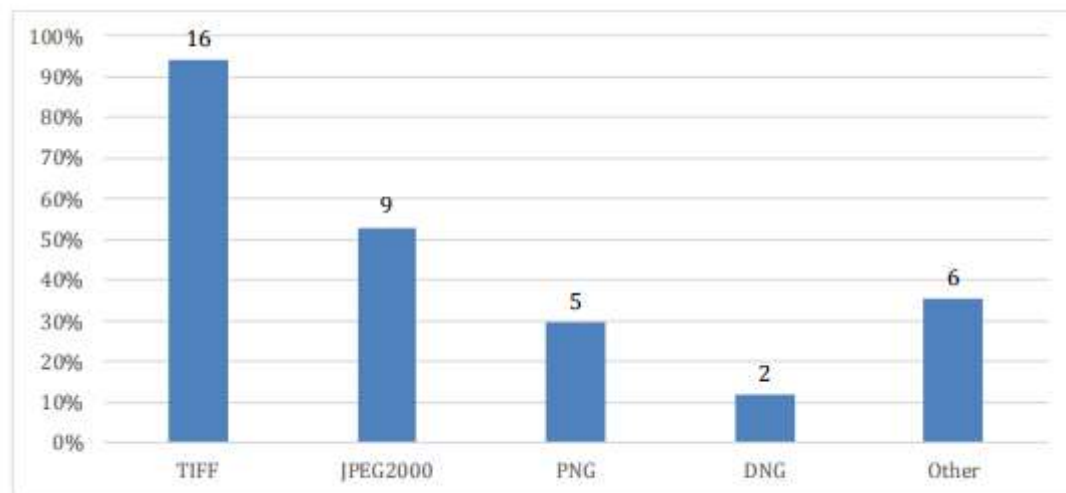


Figure 1 Summary for master file formats

Formáty, které s vyskytují v doporučeních pro zpřístupnění obrazových dat

- PDF pro textový obsah
- TIFF není nejlepší volbou

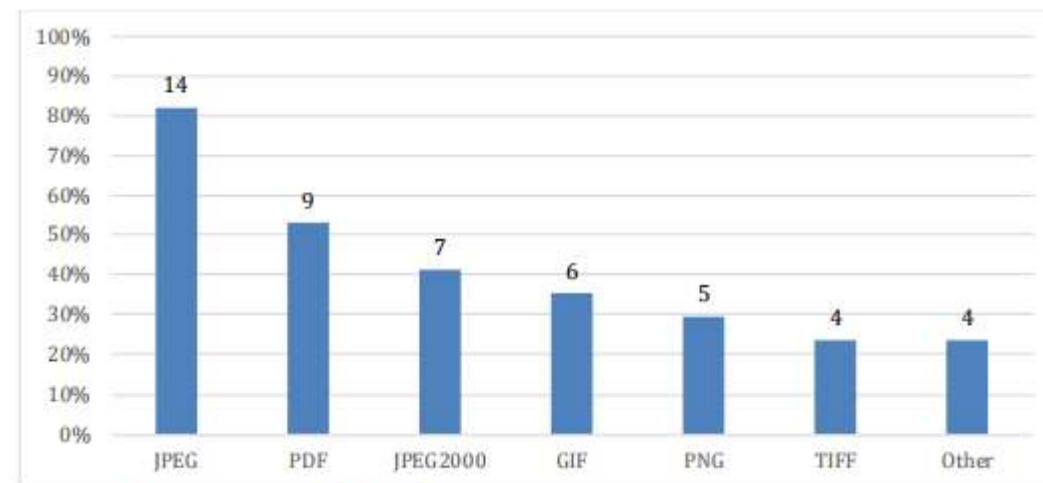


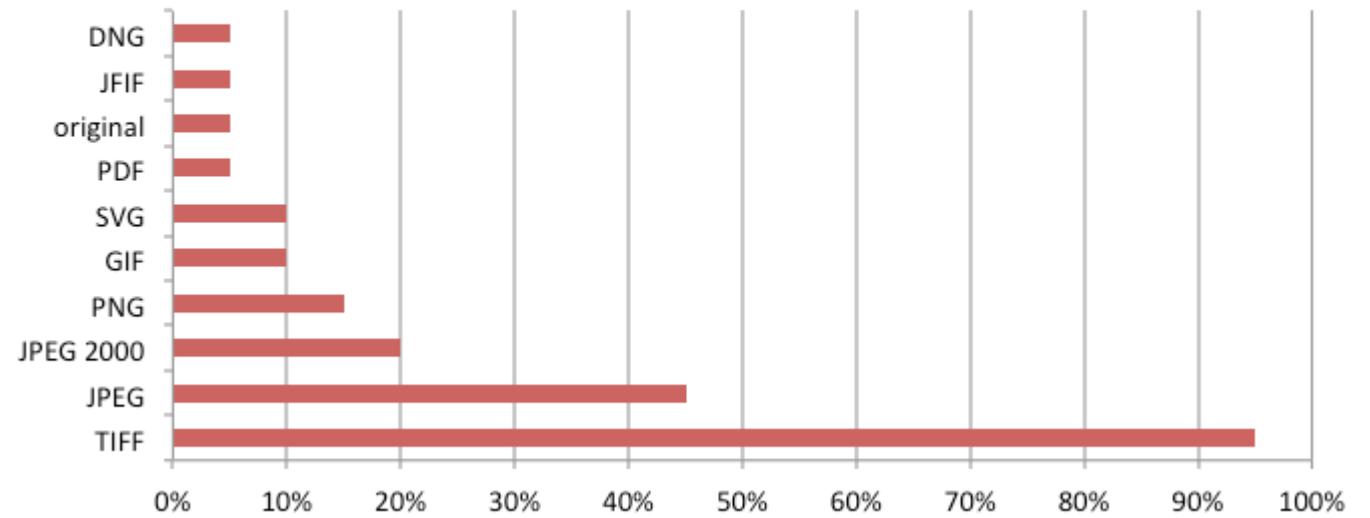
Figure 2 Summary for delivery file formats

# University of Alabama Libraries-průzkum 2016

- Průzkum se týkal celkového zacházení s digitálními dokumenty
  - Co instituce přijímají, workflow, metadata, formáty
- Formáty
- 62 institucí (50% akademické knihovny, dále archivy, muzea, vládní organizace)
  - Jaké přijímají a jaké používají pro archivaci
  - Vítěz TIFF



## Target Formats for Still Images



TIFF -používá 19 institucí

JPEG -9 institucí

JPEG2000 -4 instituce

# Významné instituce používající JPEG2000

- **Archivační formát**

- Bibliothèque nationale de France od r. 2014 (nahradil TIFF)
- Kungliga Biblioteket pro digitalizované noviny (knihy, mapy, plakáty archivují ve formátu TIFF)
- National Archives in Britain od r. 2013 pro archivaci i zpřístupnění
- Library and Archives Canada
- Digitální knihovna/repozitář HathiTrust
- Internet Archive (i pro zpřístupnění)
- Google Books
- British Library
- Koninklijke Bibliotheek
- New York Public Library
- Oxford University's Bodleian Library (jeden soubor pro archivaci i zpřístupnění)
- FBI pro otisky prstů
- Preferovaný archivační formát (někdy spolu s TIFF aj) pro: Finská národní digitální knihovna, Missouri University of Science and Technology, Cornell University Library, Ecole Polytechnique Bibliotheque, Smithsonian Institution Archives, University of Maryland University Libraries, CSU Digital Repository—Colorado state university libraries, University of Washington University Libraries



# Zpřístupňující kopie

- Kongresová knihovna (od 2006)- projekt NDNP a American Memory (pro mapy nahradil MrSID)
- U.S National Library of Medicine (pro archivaci používají TIFF)





# Kdo nemá rád JPEG2000

- Repozitáře které jej nepodporují:
  - UCD Library, Virginia tech, University of Sydney, The University of Western Australia
- University of Illinois at Urbana-Champaign
  - Jpeg2000 od r 2007 jako archivační formát pro obrazová data
  - Testy 2015-2016 → Photoshop většinu JP2 neotevře= problém pro zaměstnance i uživatele → vrátili se k TIFFu jako archivačnímu formátu (JP2 ale používají stále jako zpřístupňující)
- Některé instituce jej akceptují ale není jejich preferovaným formátem (National Archives and Records Administration, UK Data Archive apod.)



# Vybrané profily JP2 pro archivační kopie

|                                  | NDK                         | <u>National Archives UK</u> | <u>Japan National Diet Library</u> | <u>BnF</u>                  | Švédská národní knihovna               | <u>Oxford Digital Library</u> |
|----------------------------------|-----------------------------|-----------------------------|------------------------------------|-----------------------------|--|-------------------------------|
| <b>Pro</b>                       | Knihy, noviny, mapy         | Knihy apod.                 |                                    |                             | Noviny                                 |                               |
| <b>UC v MC</b>                   | MC                          | MC                          | MC                                 | MC                          | MC                                     | MC                            |
| <b>Dlaždice</b>                  | 4096x4096                   | 1 dlaždice na obraz         | 1024x1024                          | 1024x1024                   | 1024x1024                              | 1 dlaždice (default Kakadu)   |
| <b>Transformace</b>              | 5-3 ( <u>lossless</u> )     | 5-3                         | 5-3                                | 9-7 ( <u>lossy</u> )        | 9-7                                    | 5-3                           |
| <b>Kompresní poměr</b>           | -                           | -                           | -                                  | 1:4-1:6                     | 1:6                                    | -                             |
| <b>Průběh zobrazení</b>          | RPCL                        | RPCL                        | RLCP                               | RPCL                        | RPCL                                   | RPCL                          |
| <b>Dekompoziční úroveň</b>       | 5-6                         | 7                           | 6                                  | 10                          | 6                                      | 6                             |
| <b>Vrstvy kvality</b>            | 1                           | 1                           | 1                                  | 10                          | 14                                     | 1                             |
| <b>Velikost bloků</b>            | 64x64                       | 64x64                       | 64x64                              | 64x64                       | 64x64                                  | 64x64                         |
| <b><u>Precinct (regiony)</u></b> | {256,256}{256,256}{128,128} | NE                          | NE                                 | {256,256}{256,256}{128,128} | NE                                     | {256,256}{256,256}{128,128}   |
| <b>SOP</b>                       | ANO                         | NE                          | NE                                 | ANO                         | ANO                                    | ANO                           |
| <b>EPH</b>                       | ANO                         | NE                          | NE                                 | ANO                         | ANO                                    | ANO                           |
| <b>Režim BYPASS</b>              | ANO                         | Výběrový                    | NE                                 | NE                          | NE                                     | NE                            |
| <b><u>Vložená metadata</u></b>   | NE                          | Neuvádí                     | Neuvádí                            | ANO ( <u>uuid</u> )         | ANO (nastavení použita při tvorbě JP2) | NE                            |

# Vybrané profily JP2 pro zpřístupňující kopie

|                                       | NDK                             | <u>National Archives</u><br>UK | <u>Japan National</u><br><u>Diet Library</u> | Kongresová<br>knihovna<br>program NDNP  | IGN (Francouzský<br>geografický<br>institut) |
|---------------------------------------|---------------------------------|--------------------------------|--|---|--|
| Pro                                   | Knihy, noviny, mapy             | Knihy apod.                    |  | noviny  | mapy   |
| UC v MC                               | UC                              | UC                             | UC   | UC  | UC   |
| Dlaždice                              | 1024x1024                       | 1024x1024                      | 1024x1024                                    | 1024x1024   | 1024x1024                                    |
| Transformace                          | 9-7 ( <u>lossy</u> )            | 9-7                            | 9-7  | 9-7 ( <u>lossy</u> )  | 9-7  |
| Kompresní<br>poměr                    | 1:8-1:30                        | 1:6, 1:4-1:10 dle<br>materiálu | 1:50   | 1:8   | 1:20   |
| Průběh<br>zobrazení                   | RPCL                            | RPCL                           | RLCP   | RLCP nebo RLPC  | RPCL   |
| Dekompoziční<br>úroveň                | 5-6                             | 7                              | 6  | 6   | 8  |
| Vrstvy kvality                        | 12                              | 1                              | 1  | 25  | 12   |
| Velikost<br>bloků                     | 64x64                           | 64x64                          | 64x64  | 64x64   | 64x64  |
| <u>Precinct</u><br>( <u>regiony</u> ) | {256,256}{256,256}{12<br>8,128} | NE                             | NE   | NE  | {256,256}{256,256}{1<br>28,128}              |
| SOP                                   | Volitelné                       | NE                             | NE   | NE  | Volitelné                                    |
| EPH                                   | Volitelné                       | NE                             | NE   | NE  | Volitelné                                    |
| Režim<br>BYPASS                       | ANO                             | Výběrový                       | NE   | ANO   | NE   |
| <u>Vložená</u><br><u>metadata</u>     | NE                              | Neuvádí                        | Neuvádí                                      | ANO (informace<br>o použitém<br><u>kodeku</u> , základní<br><u>bibliog. údaje</u> ) | NE   |

# JPEG2000 is not dead

(yet)

