

Zpráva ze služební cesty

Jméno a příjmení účastníka cesty	PhDr. Bedřich Vychodil
Pracoviště – dle organizační struktury	Odbor digitálních fondů
Pracoviště – zařazení	Útvar 2.4.2
Důvod cesty	Pozvání na konferenci EWASS 2012 Rome
Místo - město	Řím
Místo – země	Vatikán
Datum (od – do)	4. -7. 7. 2012
Podrobný časový harmonogram	4. 7. přelet, 5. 7. konference, 6. 7. odlet
Spolucestující z NK	Ne
Finanční zajištění	Vatikán (cestovné, konferenční poplatek, část ubytování), NK (část ubytování, stravné pojištění)
Vztah k projektu	Prezentace aplikace DIFFER na základě oficiálního pozvání z Biblioteca Apostolica Vaticana
Cíle cesty	Výměna zkušeností užitých grafických formátů a metadat digitalizace, prezentace projektu DIFFER (VaV)
Plnění cílů cesty	Prezentace aplikace DIFFER, diskuze o dlouhodobém ukládání digitálních obrazových reprezentací.
Další podrobnější informace	DIFFER je určen v další fázi projektu pro validace formátů
Podpora publicity projektu	V rámci diskuse se zainteresovanými účastníky

Související materiály	
Materiál	Místo uložení
Sborník 2012	EWASS 2012
Prezentované materiály:	http://www.vaticanlibrary.va/home.php?pag=in_evidenza_art_00168&BC=11

Datum předložení zprávy	25. Září 2012
Podpis předkladatele zprávy	Bedřich Vychodil

Datum	Podpis
--------------	---------------

Podpis nadřízeného

Vloženo na intranet

Přijato v mezinárodním oddělení

Tato služební cesta na konferenci *EWAS 2012* byla podniknuta na základě oficiálního pozvání vatikánské knihovny (Biblioteca Apostolica Vaticana), viz příloha.

V odborné sekci *Special Session 12 Program* byla prezentována a následně probírána problematika dlouhodobého uchování digitálních (dvoj i více dimenzionálních) reprezentací. Přičemž se nejednalo o klasické 2D a 3D v pravém slova smyslu, ale o tzv. *datasets*, které slouží pro uchování vícedimenzionálních dat z prostředí astrofyziky a astronomie.

V rámci konference byl představen a do podrobnosti rozebrán (obrazový) formát FITS (Flexible Image Transport System) sloužící pro uchování a šíření takovýchto dat.

Vatikánská knihovna přistoupila k realizaci využití formátu FITS pro své digitální (2D) reprezentace vzniknuvší v rámci digitalizace. A to na základě dlouhodobé spolupráce s vědci z oblasti astrofyziky a astronomie. Již v současné době disponují velkou bází tzv. *datasets* uchovávaných ve formátu FITS. Bedřichem Vychodilem byl prezentován projekt DIFFER vyvíjený v rámci VaV (oblast 2a č. 133). Velký zájem způsobilo i nasazení formátu JPEG 2000, který bude využit v rámci projektu NDK, NK ČR.

V diskusním bloku byly projednávány rozdíly, výhody a nevýhody nasazení formátů JPEG 2000 a FITS.

Nevýhody formátu FITS:

- OFAF limitace
- ASCII (68 znaků)

Výhody formátu FITS:

- Formát byl již od svého počátku navrhován pro účely dlouhodobého uchování s důrazem na zpětnou kompatibilitu.
- Možnost ukládání vícerozměrných polí v rozsahu od 1D spekter, 2D obrázků nebo 3D datových prostorů až po 999D rozměrná pole)
- Možnost ukládání dvourozměrných tabulek
- Verze 3.0 z července roku 2008
- Standardizace specifikace se odhaduje do konce roku 2012
- Dokumentace na bitové úrovni (velmi důležitá vlastnost pro dlouhodobé uchování)
- Kompletní dokumentace: http://fits.gsfc.nasa.gov/fits_documentation.html

Výsledek diskuze byl takový, že v případě vatikánské knihovny je formát FITS adekvátním rozhodnutím z důvodů diverzity zpracovaných dat a především z důvodu přístupu k datům. Využívání formátu JPEG 2000 též nevylučují jako alternativu pro dlouhodobé uložení.

Podstatný význam této služební cesty je v navázání osobních kontaktů a rozhovorech mimo hlavní přednášky, ve kterých účastníci konference diskutují konkrétní otázky vlastních projektů a domlouvají další spolupráci, neméně významný je vliv na reprezentaci Národní knihovny i celé ČR a posílení její vědecké a výzkumné pozice v globálním měřítku.

Odkazy na prezentované příspěvky:

http://www.vaticanlibrary.va/home.php?pag=in_evidenza_art_00168&BC=11

Program konference: EWASS 2012 Special Session 12 Program

Opening – chair Lucio Chiappetti

11:00–11:15 Rosa Caffo – Direttore dell'Istituto Centrale per il Catalogo unico delle Biblioteche italiane

11.15-11.45: Opening – chair

Lucio Chiappetti The Flexible Image Transport System (FITS) Format

11.45-12.15 Luciano Ammenti The FITS (Flexible Image Transport System) Format: 50 Years of Experience in Long-Term Digital Conservation

12.15-12.45 Bedrich Vychodil DIFFER m(Determinator of Image File Format propERTies)

12.45-13.00 Discussion

13.00-14.30 Lunch

14.30-15.00 Stefano Allegrezza Flexible Image Transport System: a new standard file format for long-term preservation projects??

15.00-15.30 Michael Seadle The Elements of Authenticity in Digital Preservation

15.30-16.00 Giovanni Michetti Digital Preservation: No One-Size-Fits-All Format

16.00-16.30 Coffee Break

16.30-16.45 Discussion

16.45-17.15 Thomas mLedoux Long-term preservation at the National Library of France: the Scalable Preservation and Archiving Repository (SPAR)

17.15-17.30 Discussion and Conclusion

Prezentace:

Autor/Přednášející: PhDr. Bedrich Vychodil

Téma: DIFFER (Determinator of Image File Format propERTies)

Odkaz na prezentaci: http://www.vaticanlibrary.va/moduli/Vychodil_EWASS2012.pdf

DIFFER

Determinator of Image File Format propERTies

Lecture: 2012 EWASS, 5 JUL, 2012
 Lecturer: Bedrich Vychodil
 Web: www.nkp.cz, www.ndk.cz
 Contact: bedrich.vychodil@nkp.cz
 bedrich@gmail.com

Klementinum - built (1653–1726)



Overview

1992	Take-off	Pilot project under UNESCO
2005	Award	UNESCO/Jikji Memory of the World Prize
2012	Current state	~10,000,000 pages
2012-15	Our goal	~26,000,000 pages
2012-16	Google	~20,000,000 pages

Compression Ratio TEST

Format	Scan		MC		UC		MC/UC	
	TIFF	JPEG	JPEG	DjVu	JPEG	DjVu	JPEG	JPEG2000
Storage gain	0.0%	0.0%	77.0%	84.3%	81.3%	85.6%	84.8%	81.3%
Storage loss	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Migration from JPEG to JP2

Difference between @yans
 DEVIATION
 Black - Min
 White - Max

JPEG2000 Reference Chart

	Master Copy	Production Master Copy	Production Master Copy
Used for:	Books, periodicals, maps, microfilm	Books, periodicals	Maps, microfilm
Characteristics software used:	Kodak	Kodak	Kodak
File format:	Part 1: j2k	Part 1: j2k	Part 1: j2k
Color or grayscale:	Lossless	Lossy	Lossy
Typical compression:	12:1 to 3	10:1 to 16	1:1 to 1:16
Typical resolution:	4096 x 4096	4096 x 4096	1024 x 1024
Interpolation used:	4:1 to 8	4:1 to 8	4:1 to 8
Number of reconstruction levels:	1 (used for non-archival)	9	17 (used for archival material)
Number of quality layers:	1	10 (dependent)	10 (dependent)
Code block size (width x height):	64 x 64	64 x 64	64 x 64
Transformation:	0 (vertical)	0 (vertical)	0 (vertical)
Residual size:	minimum 64 x 64 (vertical), 32 x 32 (horizontal)	minimum 64 x 64 (vertical), 32 x 32 (horizontal)	minimum 64 x 64 (vertical), 32 x 32 (horizontal)
Segment of interest:	no	no	no
Code block size:	64x64	64x64	64x64
TIFF metadata:	Yes (Y)	Yes (Y)	Yes (Y)
Keywords:	YES	YES	YES
ICC profiles:	YES	YES	YES
Metadata:	Embedded as J2K metadata in J2K file	Embedded as J2K metadata in J2K file	Embedded as J2K metadata in J2K file
Quality limits the impact on bit mapping, avoid limits for damage or to avoid loss in the J2K file	Color: 4096x4096 128, 400-100	0	0

Kakadu Command-lines

Master Copy

```
kdj_compress -i example.tif -o example.jp2 "CbL=64,64" "Corder=RPCL" "Stiles=4096,4096" "Cprecincts=[256,256],[128,128]" "OROparts=R" "Oreversible=yes" "Clayers=1" "Clevels=5" "Cmodes=[BYPASS]" "Csize_topbytes Csize_bottombytes"
```

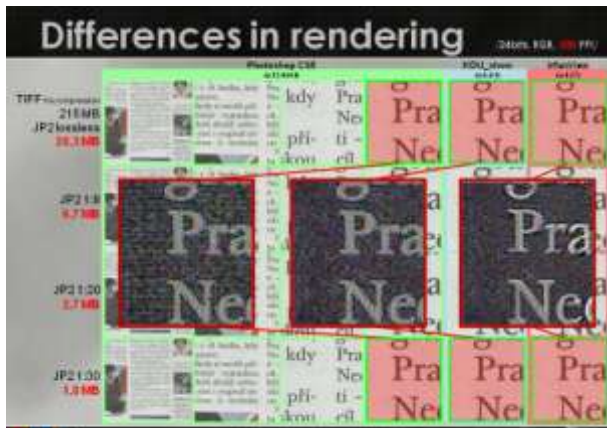
Production Master Copy

Compress Ratio 1:8
 kdj_compress -i example.tif -o example.jp2 "CbL=64,64" "Corder=RPCL" "Stiles=1024,1024" "Cprecincts=[256,256],[128,128]" "OROparts=R" -rate 3 "Clayers=12" "Clevels=5" "Cmodes=[BYPASS]"

Compress Ratio 1:20
 kdj_compress -i example.tif -o example.jp2 "CbL=64,64" "Corder=RPCL" "Stiles=1024,1024" "Cprecincts=[256,256],[128,128]" "OROparts=R" -rate 1.2 "Clayers=12" "Clevels=5" "Cmodes=[BYPASS]"

Differences in rendering

TIFF (uncompressed) 123 MB
 JP2 lossless 25.8 MB
 JP2 1:8 11.5 MB
 JP2 1:20 6.0 MB
 JP2 1:30 3.0 MB



PROJECT - Tool Wrapper

DIFFER

(Determinator of Image File Format propERTies)

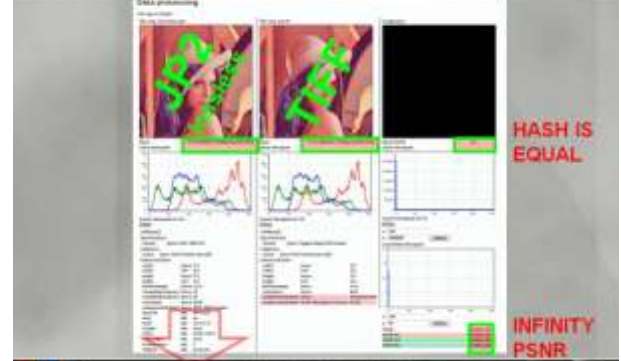
WHAT IT DOES

- TIFF, JPEG, JP2, DjVu, (PNG, PDF)
- Identification
- Characterization
- Validation
- Visual comparison
- Similarity metrics
- Detection of glitches
- JP2 profile validator

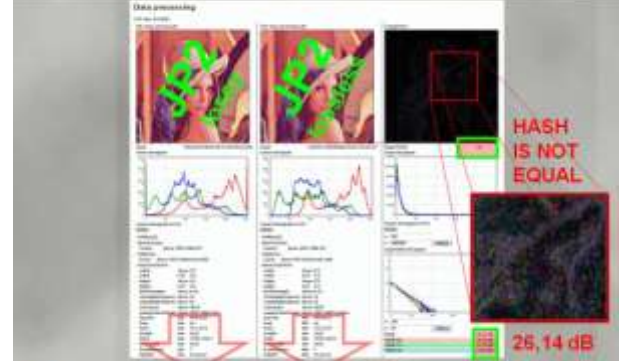
WHAT IS IN IT

- JHOVE (JSTORHarvard Object Validation Environment)
- ExifTool (Read, Write and Edit Meta-information)
- KDU_expand (library of Kakadu)
- DJVUDUMP
- DROID (Digital Record Object Identification)
- FFIdent (tool wrapper)
- FITS (File Information Tool Set)
- NLNZ MTD Extraction Tool (tool wrapper)
- PRONOM (The technical registry PRONOM)
- Jpylyzer (by van der Knijff)

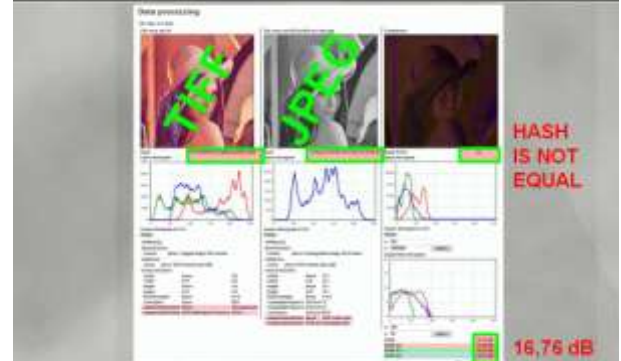
DIFFER – Finds Differences



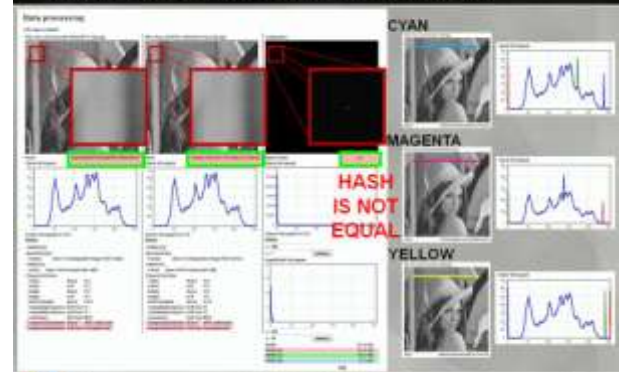
DIFFER – Finds Differences



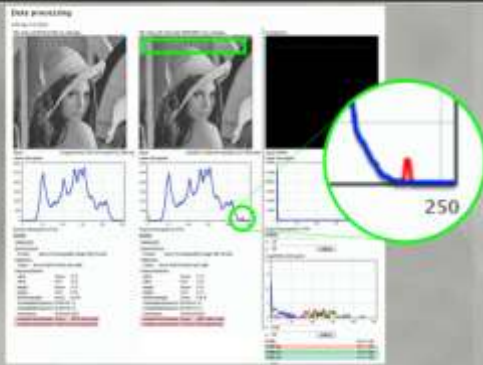
DIFFER – Finds Differences



DIFFER – Pixels Detection



DIFFER – Glitches Detection



DIFFER – Glitches Detection



DIFFER – Corrupted file Detection



DIFFER – Corrupted file Detection



DIFFER – JP2 profile validator

Follow-up Study

- Web Service – JAVA
- Google Summer of Code
<http://www.google.com/events/opensource/2012/programingoogle/2012/track>
- Open Source
<https://github.com/nkp/wendarydiff>
- MSSIM (Multi-Structural Similarity Index)
- UIQI (Universal Image Quality Index)
- Lossless vs. Lossy for Master Copy
- Digital Images Production and QC

Link to DIFFER

BETA VERSION:
<http://differ.nkp.cz>

Questions...?

Lecture: 2012 EWASS, 5 JUL, 2012
Lecturer: Bedrich Vychodil
Web: www.nkp.cz, www.ndk.cz
Contact: bedrich.vychodil@nkp.cz
bedrich@gmail.com

Fotodokumentace z konference EWASS2012 Special Session 12 Program



Obrázek 1: Diskusní blok, EWASS 2012 Special Session 12 Program, Pontificia Universita Lateranese, Řím



Obrázek 2: Diskusní blok, EWASS 2012 Special Session 12 Program, Pontificia Universita Lateranese, Řím



Obrázek 3: Bedřich Vychodil, prezentace projektu DIFFER, Pontificia Universita Lateranese, Řím



Obrázek 4: Budova Pontificia Universita Lateranese, Řím