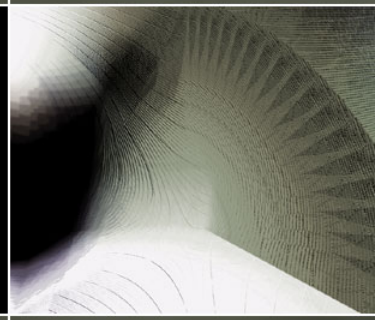
JPEG 2000

Specifications for The National Library of the Czech Republic



Lecture: Library of Congress, MAY 13, 2011
Lecturer: Bedřich Vychodil
Web: www.nkp.cz, www.ndk.cz
Contact: bedrich@gmail.com
bedrich.vychodil@nkp.cz

Klementinum 1653–1726

Overview

1992	Take-off	Pilot project under UNESCO
2005	Award	UNESCO/Jikji Memory of the World Prize
2011	Current state	~10,000,000 pages
2011-14	Our goal	~26,000,000 pages
2011-16	Google	~20,000,000 pages

(200,000 books)

Compression Ratio TEST

Scan **MC** **UC** **MC/UC**

TIFF JPEG DjVu JPEG2000

Format		BMP	TIFF	TIFF LZW	PNG	JPEG (12)	JPEG (11)	DJV photo MAX	DJV photo preset	DJV manuscript	JP2 (0)	JP2 (1:1)	JP2 (1:10)	JP2 (1:25)	JPM photo	JPM standard/good	JPM standard/low
Comparison %	A - 8bit, Gray	100%	100%	4,30%	2,83%	1,81%	1,20%	1,05%	0,25%	0,06%	2,45%	2,28%	1,15%	0,46%	0,41%	0,13%	0,09%
	A - 24bit, RGB	100%	100%	0,27%	0,21%	0,96%	0,76%	0,85%	0,38%	0,01%	0,71%	1,03%	0,38%	0,15%	0,14%	0,05%	0,05%
	B - 8bit, Gray	100%	100%	0,42%	0,19%	1,12%	0,90%	0,85%	0,38%	0,01%	0,70%	1,05%	1,05%	0,46%	0,41%	0,08%	0,08%
	B - 24bit, RGB	100%	100%	0,88%	0,60%	0,76%	0,55%	0,55%	0,20%	0,02%	0,71%	0,86%	0,37%	0,15%	0,14%	0,05%	0,04%

File size compare to TIFF	100%	100%	22,97%	15,70%	0,66%	14,36%	5,17%	**0,54%**	18,47%	0,78%		0,14%	
Storage gain	**0,0%**	**0,0%**	**77,0%**	**84,3%**	**91,2%**	**85,6%**	**94,8%**	**99,5%**	**81,5%**	**93,0%**		**98,0%**	
Number of layers	1 layer	1 layer		1 layer	1 layer	1 layer	1 layer	3 layer	1 layer			3 layers	

BMP TIFF (LZW) PNG

JPEG2000 Parameters Chart

	Master Copy	Production Master Copy	Production Master Copy
Used for	Books, periodicals, maps, manuscripts	Books, periodicals	Maps, manuscripts
Conversion software used	Kakadu	Kakadu	Kakadu
File format	Part 1 (.jp2)	Part 1 (.jp2)	Part 1 (.jp2)
Lossy or lossless	**Lossless**	**Lossy**	**Lossy**
Typical compression	1:2 to 1:3	1:20 to 1:30	1:8 to 1:10
Tiling	4096x4096	1024x1024	1024x1024
Progression order	RPCL	RPCL	RPCL
Number of decomposition levels	**5** or **6** /6 layers for over-sized material/	**5**	**5** or **6** /6 layers for over-sized material/
Number of quality layers	**1**	**12** /logarithmic/	**12** /logarithmic/
Code block size (xcb = ycb)	6	6	6
Transformation	5-3 reversible	9-7 irreversible	9-7 irreversible
Precinct size	256x256 for first two decomp. levels, 128 by 128 for lower levels	256x256 for first two decomp. levels, 128 by 128 for lower levels	256x256 for first two decomp. levels, 128 by 128 for lower levels
Regions of Interest	No	No	No
Code block size	64x64	64x64	64x64
TLM markers	Yes “R”	Yes “R”	Yes “R”
Bypass	**YES**	**YES**	**YES**
ICC profiles	YES	?	YES
Metadata	Embedded as XMP metadata in JP2 XML box	Embedded as XMP metadata in JP2 XML box	Embedded as XMP metadata in JP2 XML box
Greatly limits the impact on bit flipping, as it limits the damage to a single block in the JPEG 2000 file	**Cuse_sop=yes Cuse_eph=yes**	?	?

Kakadu Command-lines

Master Copy

kdu_compress -i example.tif -o example.jp2 "Cblk={64,64}" Corder=RPCL "Stiles={4096,4096}" "Cprecincts={256,256},{128,128}" ORGtparts=R Creversible=yes Clayers=1 Clevels=5 "Cmodes={BYPASS}" -double_buffering **Cuse_sop=yes Cuse_eph=yes**

Production Master Copy

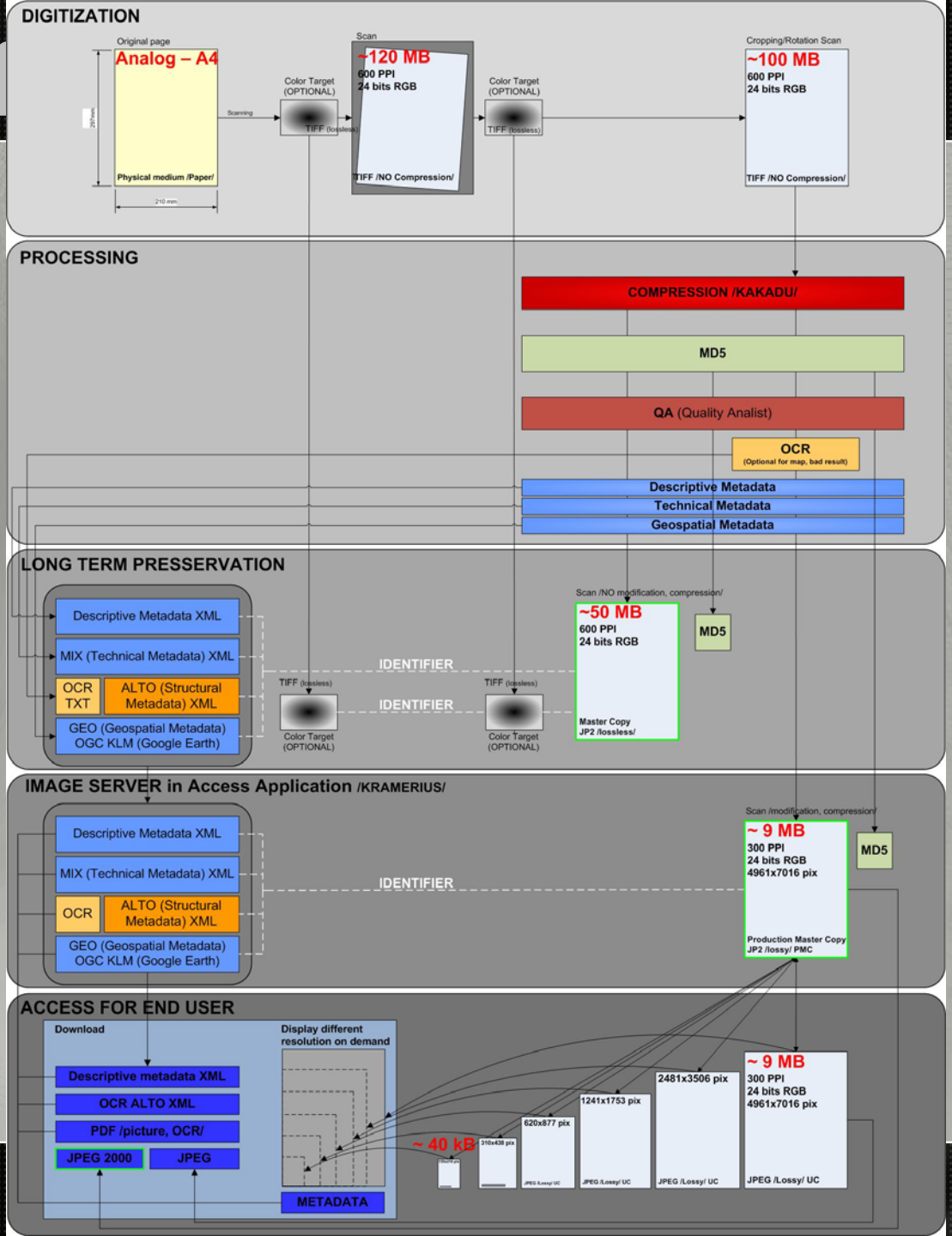
Compress Ratio 1:8

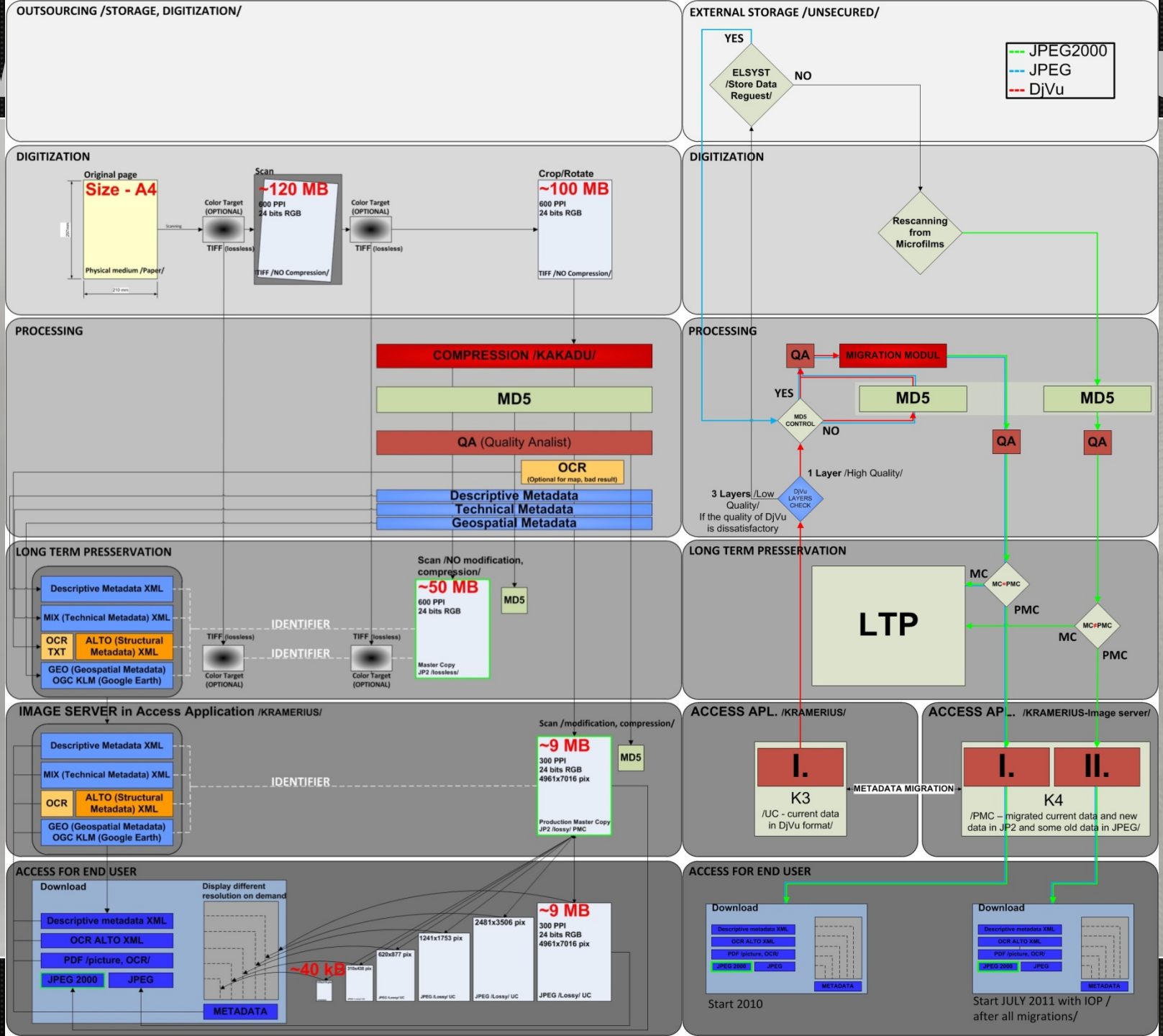
kdu_compress -i example.tif -o example.jp2 "Cblk={64,64}" Corder=RPCL "Stiles={1024,1024}" "Cprecincts={256,256},{128,128}" ORGtparts=R -rate 3 Clayers=12 Clevels=5 "Cmodes={BYPASS}"

Compress Ratio 1:20

kdu_compress -i example.tif -o example.jp2 "Cblk={64,64}" Corder=RPCL "Stiles={1024,1024}" "Cprecincts={256,256},{128,128}" ORGtparts=R -rate 1.2 Clayers=12 Clevels=5 "Cmodes={BYPASS}"

Workf


DIGITIZATION
Original page
Analog – A4
Physical medium /Paper/
Scanning
Color Target (OPTIONAL)
TIFF (lossless)
Scan
~120 MB
600 PPI
24 bits RGB
TIFF /NO Compression/
Color Target (OPTIONAL)
TIFF (lossless)
Cropping/Rotation Scan
~100 MB
600 PPI
24 bits RGB
TIFF /NO Compression/
PROCESSING
COMPRESSION /KAKADU/
MD5
QA (Quality Analist)
OCR
(Optional for map, bad result)
Descriptive Metadata
Technical Metadata
Geospatial Metadata
LONG TERM PRESSERVATION
Descriptive Metadata XML
MIX (Technical Metadata) XML
OCR TXT
ALTO (Structural Metadata) XML
GEO (Geospatial Metadata) OGC KLM (Google Earth)
IDENTIFIER
IDENTIFIER
TIFF (lossless)
Color Target (OPTIONAL)
TIFF (lossless)
Color Target (OPTIONAL)
Scan /NO modification, compression/
~50 MB
600 PPI
24 bits RGB
Master Copy
JP2 /lossless/
MD5
IMAGE SERVER in Access Application /KRAMERIUS/
Descriptive Metadata XML
MIX (Technical Metadata) XML
OCR
ALTO (Structural Metadata) XML
GEO (Geospatial Metadata) OGC KLM (Google Earth)
IDENTIFIER
Scan /modification, compression/
~ 9 MB
300 PPI
24 bits RGB
4961x7016 pix
Production Master Copy
JP2 /lossy/ PMC
MD5
ACCESS FOR END USER
Download
Descriptive metadata XML
OCR ALTO XML
PDF /picture, OCR/
JPEG 2000
JPEG
Display different resolution on demand
METADATA
~ 40 kB
310x438 pix
620x877 pix
1241x1753 pix
2481x3506 pix
~ 9 MB
300 PPI
24 bits RGB
4961x7016 pix
JPEG /Lossy/ UC


OUTSOURCING /STORAGE, DIGITIZATION/
EXTERNAL STORAGE /UNSECURED/
YES
ELSYST /Store Data Reguest/
NO
JPEG2000
JPEG
DjVu
DIGITIZATION
Original page
Size - A4
Physical medium /Paper/
Color Target (OPTIONAL)
TIFF (lossless)
Scan
~120 MB
600 PPI
24 bits RGB
TIFF /NO Compression/
Color Target (OPTIONAL)
TIFF (lossless)
Crop/Rotate
~100 MB
600 PPI
24 bits RGB
TIFF /NO Compression/
DIGITIZATION
Rescanning from Microfilms
PROCESSING
COMPRESSION /KAKADU/
MD5
QA (Quality Analist)
OCR
(Optional for map, bad result)
Descriptive Metadata
Technical Metadata
Geospatial Metadata
PROCESSING
QA
MIGRATION MODUL
YES
MD5
MD5
MD5 CONTROL
NO
QA
QA
1 Layer /High Quality/
3 Layers /Low Quality/
If the quality of DjVu is dissatisfactory
DjVu LAYERS CHECK
LONG TERM PRESSERVATION
Descriptive Metadata XML
MIX (Technical Metadata) XML
OCR TXT
ALTO (Structural Metadata) XML
GEO (Geospatial Metadata) OGC KLM (Google Earth)
IDENTIFIER
IDENTIFIER
TIFF (lossless)
Color Target (OPTIONAL)
TIFF (lossless)
Color Target (OPTIONAL)
Scan /NO modification, compression/
~50 MB
600 PPI
24 bits RGB
Master Copy JP2 /lossless/
MD5
LONG TERM PRESSERVATION
LTP
MC
MC=PMC
PMC
MC=PMC
MC
PMC
IMAGE SERVER in Access Application /KRAMERIUS/
Descriptive Metadata XML
MIX (Technical Metadata) XML
OCR
ALTO (Structural Metadata) XML
GEO (Geospatial Metadata) OGC KLM (Google Earth)
IDENTIFIER
Scan /modification, compression/
~9 MB
300 PPI
24 bits RGB
4961x7016 pix
Production Master Copy JP2 /lossy/ PMC
MD5
ACCESS APL. /KRAMERIUS/
I.
K3
/UC - current data in DjVu format/
METADATA MIGRATION
ACCESS APL. /KRAMERIUS-Image server/
I.
II.
K4
/PMC – migrated current data and new data in JP2 and some old data in JPEG/
ACCESS FOR END USER
Download
Descriptive metadata XML
OCR ALTO XML
PDF /picture, OCR/
JPEG 2000
JPEG
METADATA
Display different resolution on demand
~40 kB
310x438 pix
620x877 pix
1241x1753 pix
2481x3506 pix
~9 MB
300 PPI
24 bits RGB
4961x7016 pix
JPEG /Lossy/ UC
ACCESS FOR END USER
Download
METADATA
Start 2010
Download
METADATA
Start JULY 2011 with IOP / after all migrations/

Differences in rendering

/24bits, RGB, 300 PPI/

Photoshop CS5 v.12,0 x64

IrfanView 4.27

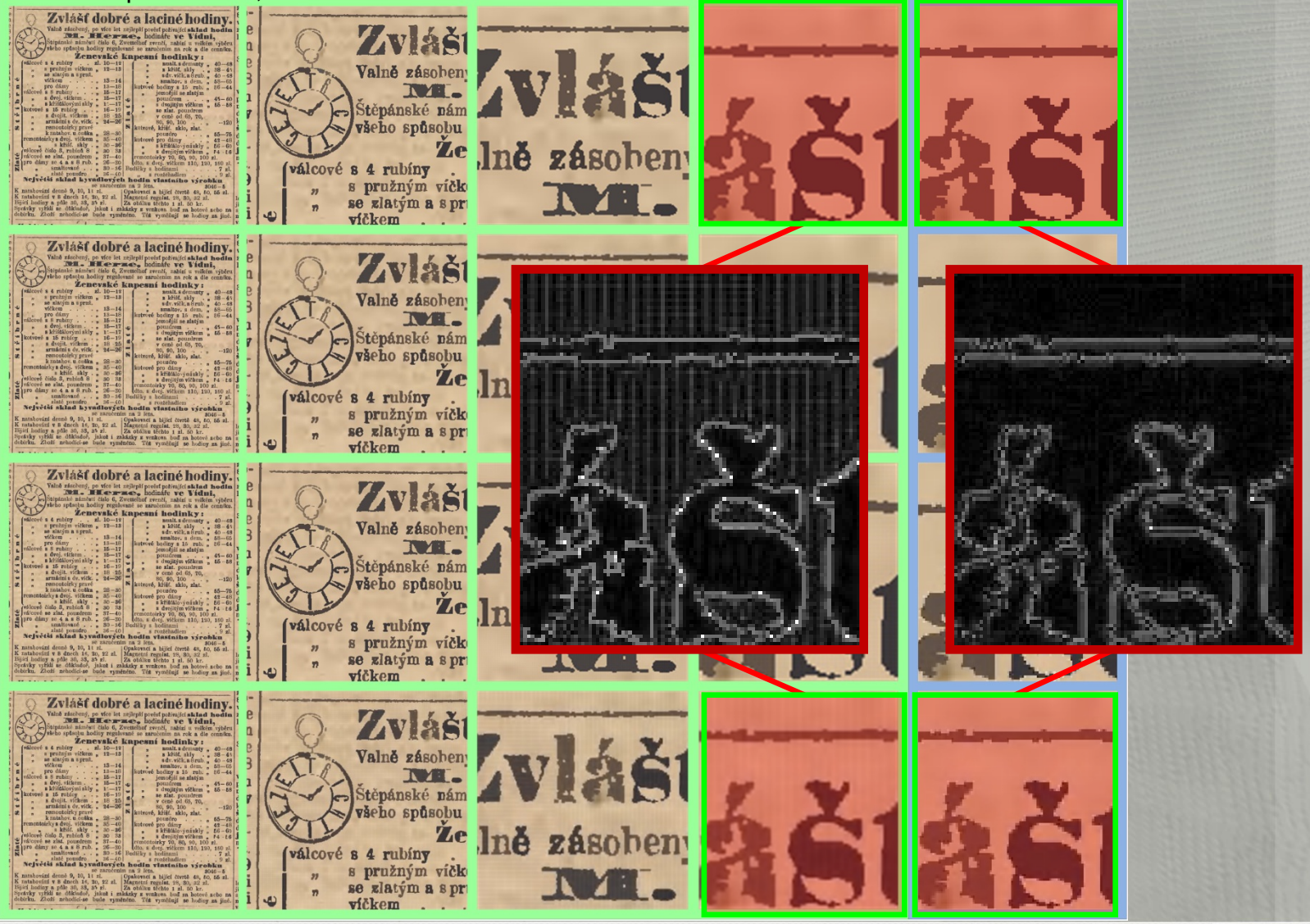
TIFF No compression
123 MB

JP2 lossless
21,5 MB

JP2 1:8
11,5 MB

JP2 1:20
4,6 MB

JP2 1:30
3,0 MB



Differences in rendering

/24bits, RGB, 600 PPI/

Photoshop CS5 v.12,0 x64

IrfanView 4.27

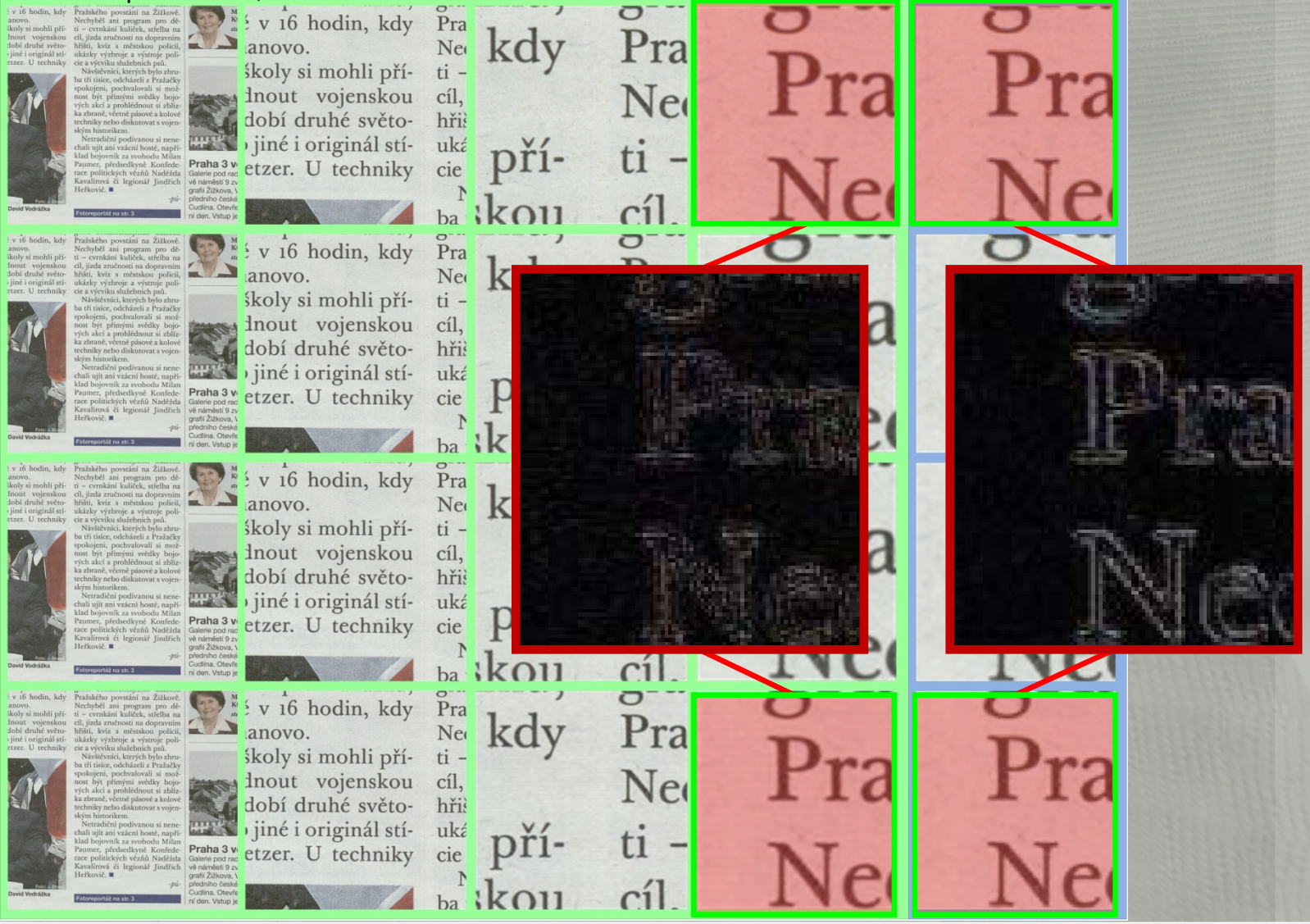
TIFF No compression
215 MB

JP2 lossless
28,3 MB

JP2 1:8
6,7 MB

JP2 1:20
2,7 MB

JP2 1:30
1,8 MB



Migration from JPEG to JP2

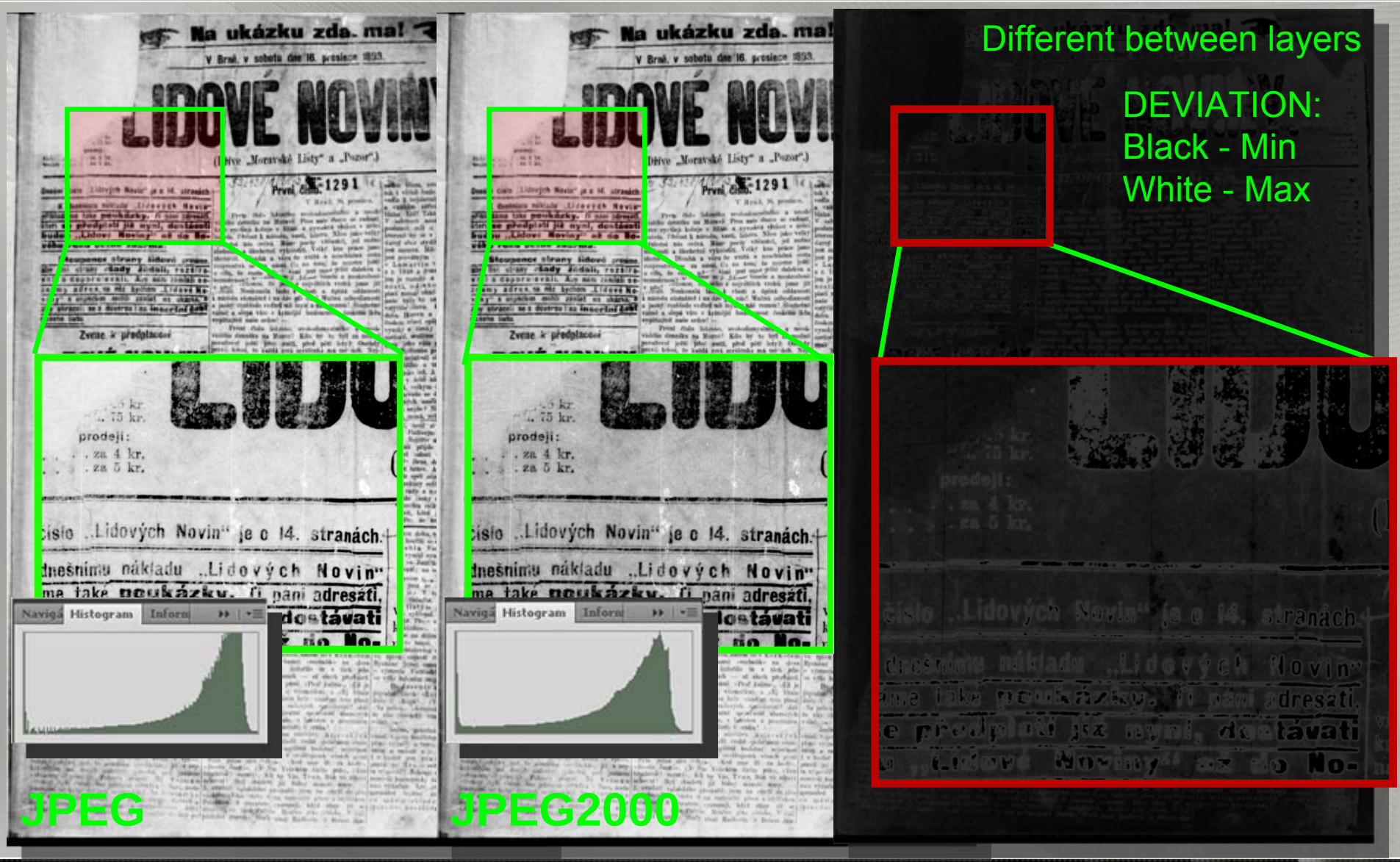
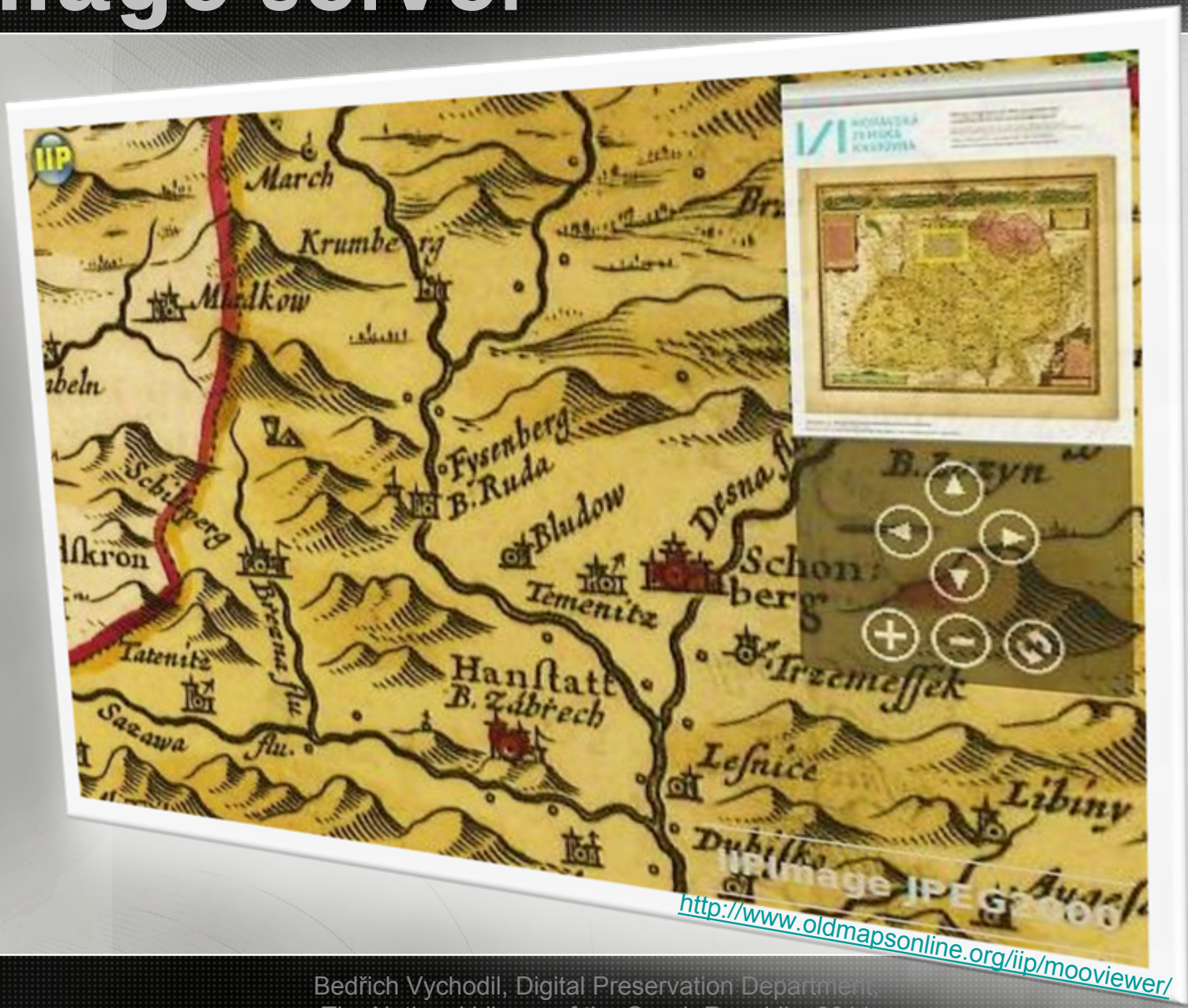


Image server

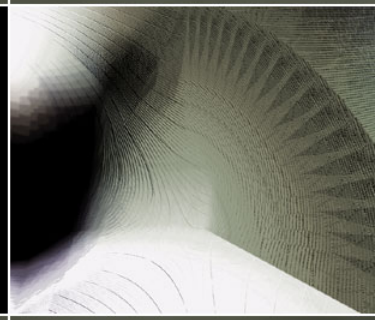


Follow-up Study

- Lossless vs. Lossy for Master Copy
- Quality Check
- Robustness

The End...

Questions...?



Lecture: Library of Congress, MAY 13, 2011
Lecturer: Bedřich Vychodil
Web: www.nkp.cz, www.ndk.cz
Contact: bedrich@gmail.com
bedrich.vychodil@nkp.cz