

DigiDepot: JPEG 2000 jako ukládací formát

Implementace JPEG 2000 jako ukládacího formátu
v datovém úložišti DigiDepot - zkušenosti a výsledky

Ing. Martin Hankovec, Státní oblastní archiv v Třeboni
martin.hankovec@ceskearchivy.cz

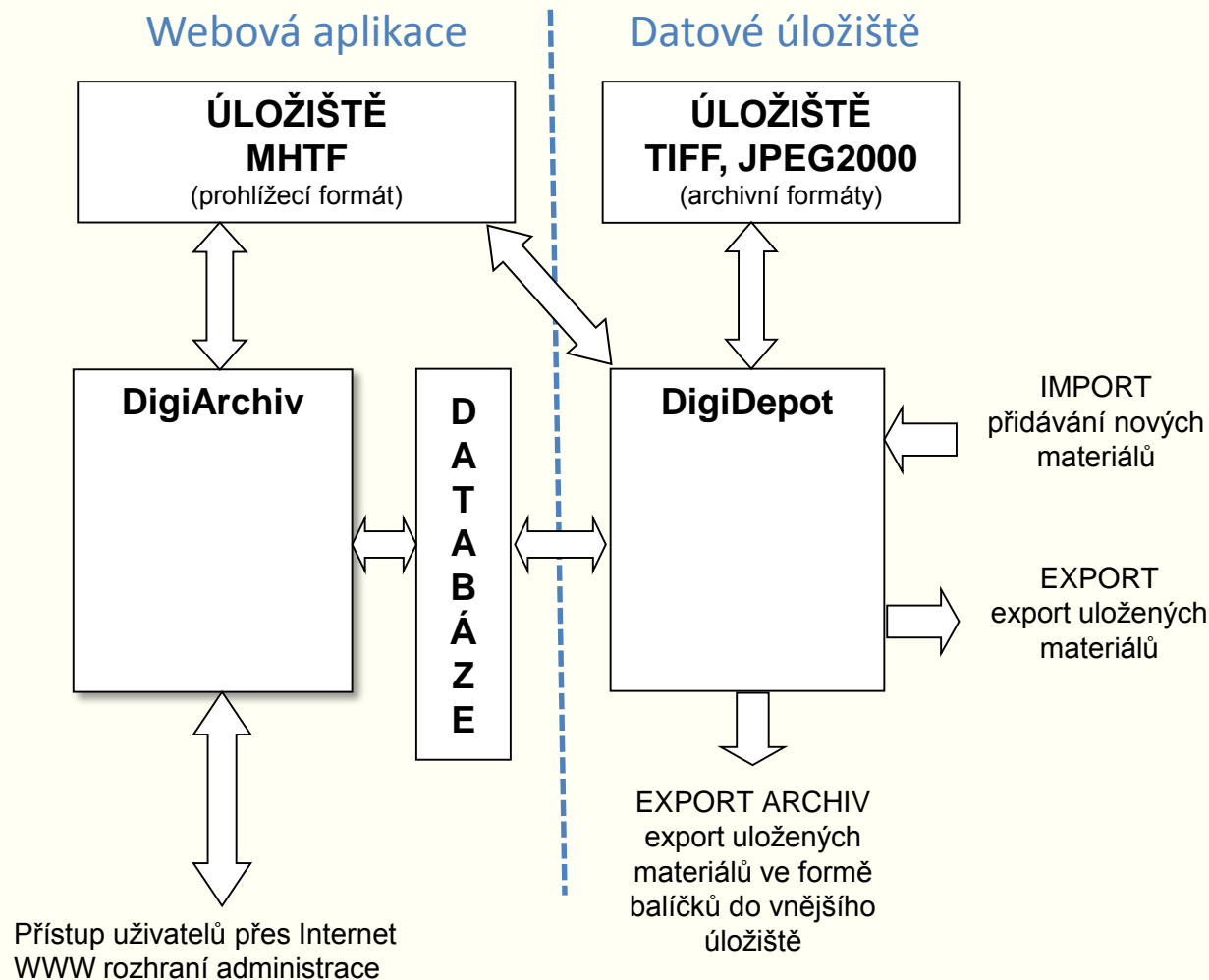
Obsah

- co je DigiDepot
- dosud používaný formát pro originální snímky
- formát JPEG 2000
- cíle zavedení JPEG 2000 v DigiDepotu
- předimplementační testy (+ výsledky)
- způsob implementace
- reálné výsledky a dopady implementace

Co je DigiDepot?

- součást systému digitálního archivu, který používá SOA v Třeboni
- samostatný paralelní vícevláknový server datového úložiště
 - import, zpracování, konverze, export originálních snímků z digitalizačních pracovišť
 - automatická archivace snímků na páskovou knihovnu včetně správy zálohovacích médií
 - správa metadat

Digitální archiv SOA v Třeboni



Dosud používaný formát

- originální snímky ukládány ve formátu TIFF s kompresí LZW
- metadata z originálních snímků ukládána ve snímcích
- metadata k materiálům ukládána zvlášť – soubory XML, CSV

Formát JPEG 2000

- komplexní formát vhodný pro ukládání originálních snímků i zpřístupňovacích kopií
- dle Metodiky pro vytváření bezpečnostních kopií archiválií v digitální podobě (Národní archiv, SOA v Praze) jeden z formátů vhodných k dlouhodobému ukládání originálů
- možnost bezztrátové komprese stejně jako u formátu TIFF
- lepší stupeň komprese než při použití formátu TIFF (LZW) i při bezztrátové kompresi

Cíle zavedení JPEG 2000 v DigiDepotu

- především snaha o snížení datové náročnosti ukládání originálních snímků (ve formátu TIFF zabírají nyní cca 60 TB)
- nebude používán současně jako prohlížečí formát

Předimplementační testy

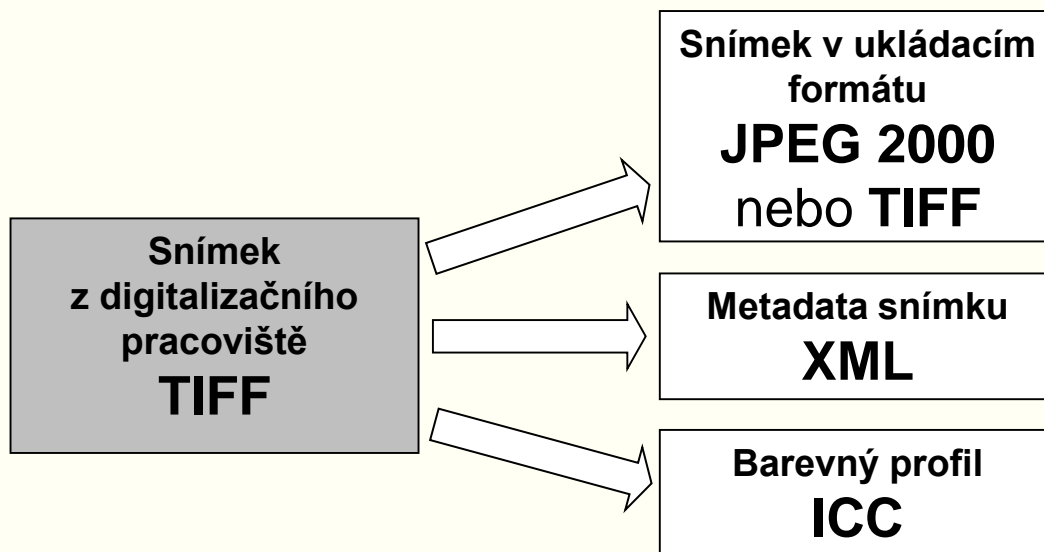
- testování několika volně dostupných implementací konvertoru do JPEG 2000
- OpenJPEG, NConvert (Jasper), ImageMagick (Jasper), Jasper
- analýza výsledného JPEG 2000 – jpylyzer
- analýza bezztrátovosti zpětným převodem do TIFF a porovnáním (ImageMagick Compare)
- analýza možnosti přenosu metadat do výsledného snímku (exiftool) a barevných profilů ICC

Předimplementační testy - výsledky

- převod do bezztrátově komprimovaného JP2 souboru zvládají všechny testované knihovny bez problému
- žádná knihovna neuchovává metadata (technická metadata, rozlišení snímku, ICC profil)

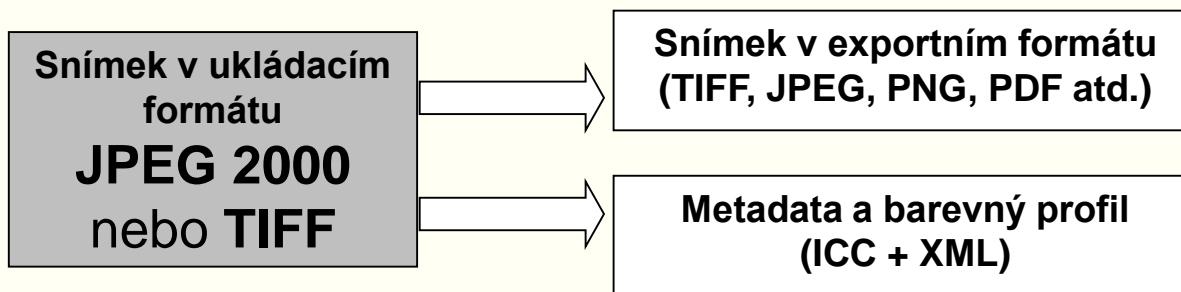
Způsob implementace - import

- vycházejí z předimplementačních testů, jsou metadata a barevné profily extrahovány ze zdrojového snímku a uloženy zvlášť



Způsob implementace - export

- metadata a barevné profily jsou exportovány zvlášť, nebo jsou vloženy zpět do exportovaného snímku



Způsob implementace - ovládání

- veškeré vnitřní operace jsou pro obsluhu transparentní
- při importu do DigiDepotu se pouze nastaví, v jakém formátu se mají snímky daného materiálu ukládat (JPEG 2000, TIFF)
- možnost kdykoli převést snímky z jednoho ukládacího formátu na druhý (JPEG 2000 <-> TIFF) pomocí příkazu DigiDepotu
- možnost kdykoli on-line zobrazit metadata ze snímku

Reálné výsledky a dopady implementace

- původní zaplněná kapacita úložiště 60 TB
- zaplněná kapacita po konverzi 44 TB
- ušetřeno 16 TB (cca 26.5%)
- ušetřeno velké množství zálohovacích pásek LTO (čtvrtina)
- rozdíl komprese výsledného snímku oproti TIFF závisí na jeho obsahu a velikosti
- rychlost převodu kolísá podle velikosti snímku (cca 1x – 2x oproti TIFF), není významná – DigiDepot zpracovává úlohy dávkově a paralelně bez nutnosti zásahů obsluhy

Závěr

Podářilo se implementovat podporu formátu JPEG2000, který přinesl nemalou úsporu kapacity úložiště. To přeneseně přináší úsporu finančních prostředků a to jak na straně investic (disková pole), tak na straně provozních výdajů (pásy do knihovny).

Děkuji za pozornost

Martin Hankovec

martin.hankovec@ceskearchivy.cz