



Podpořeno grantem z Norska
prostřednictvím
Norského finančního mechanismu

Projekt záchranné digitalizace Ad fontes v Archivu hlavního města Prahy

Projekt komplexní digitální ochrany, záchrany a obecného zpřístupnění nejvzácnějších historických dokumentů hlavního města Prahy

Název: Záchranná digitalizace a zpřístupnění nejvzácnějších historických archiválií hlavního města Prahy
CZ0041 Preservation digitizing and retrieval of the most valued historical archiving documents of Prague

Cíle projektu:

- **Záchrana a trvalé uchování historicky nejvýznamnějších, památkově nejceněnějších a fyzicky nejvíce ohrožených archiválií ze sbírek a fondů Archivu hlavního města Prahy, významných pro národní i evropské dějiny, cestou vytvoření všestranně využitelných digitálních kopií z originálů těchto archiválií**
- **Zpřístupnění archiválií v digitální podobě pro badatele a širokou veřejnost**

Zaměření / koncepce:

- **Uchování evropského kulturního dědictví**
- **Zlepšení péče a ochrana movitého kulturního dědictví**
- **Ochrana vzácných písemných dokumentů Prahy**
- **ePraha**

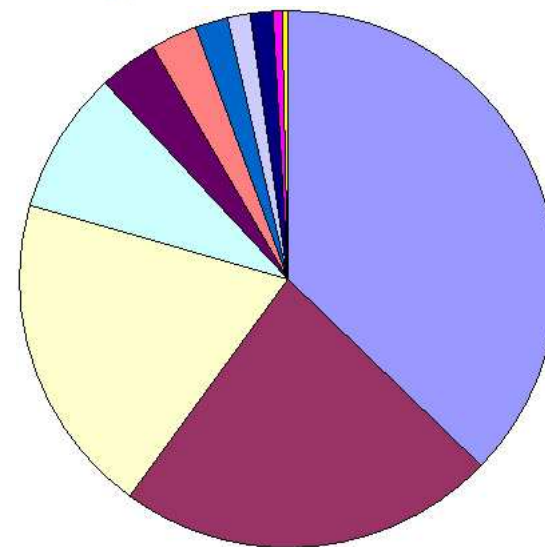
Motivace:

- **riziko zničení nebo ztráty archiválie** - archivace její digitální kopie na jiném bezpečném místě
- **riziko poškození archiválie jejím častým půjčováním do badatelný** - zpřístupnění její digitální kopie v Informačním systému v badatelně
- **nutnost prezenčního studia archiválie v badatelně archivu** - zpřístupnění na Internetu



Výběr archiválií a jeho kritéria

- historická dokumentární hodnota archiválie
- míra reálného nebo možného fyzického poškození
- frekvence jejího využívání badatelskou veřejností
- záchranná digitalizace více fondů a sbírek



- Sbírka matrik
- Magistrát hlavního města Prahy I. - konskripce Soupis pražských domovských příslušníků (1830-1910), Protokoly schůzí městské rady (1784-1814)
- Sbírka rukopisů
- Spolek výtvarných umělců Mánes Praha
- Heraldická (Rentzova) sbírka
- Sbírka papírových a sbírka pegamenových listin
- Hlavní škola židovská Praha V
- Sbírka fotografií
- Společnost pro povznesení německé vědy, umění a literatury v Čechách
- Všeobecná židovská nemocnice Praha V
- Sbírka map

stav k 1.6.2010:



Digitalizované fondy a sbírky

1. Prameny vhodné pro genalogické bádání

- Církevní a civilní matriky
- Magistrát hl. m. Prahy, ref. IV.: Soupis pražského obyvatelstva
- Sbíрка listin - erbovní a nobilitační privilegia, erbovní vývody
- Heraldická sbírka

2. Fondy a sbírky vzešlé z činnosti městských úřadů

- Magistrát hl. m. Prahy - protokoly sborů městské správy
- Magistrát hl. m. Prahy, ref. IV.: Soupis pražského obyvatelstva
- Sbíрка rukopisů - knihy městské správy (např. pamětní knihy, radní manuály, židovské knihy bílé)

3. Fondy a sbírky vzešlé z činnosti kulturních institucí a spolků

- Spolek výtvarných umělců Mánes
- Společnost pro podporu německé vědy, umění a literatury v Čechách - Německá akademie věd

Celkem má být pořizeno 230 000 skenů, z toho 150 000 skenů zpřístupněno na Internetu.

K 1.6.2010 bylo již pořizeno 230 937 skenů, z toho je 196 846 skenů zpřístupněno na Internetu.

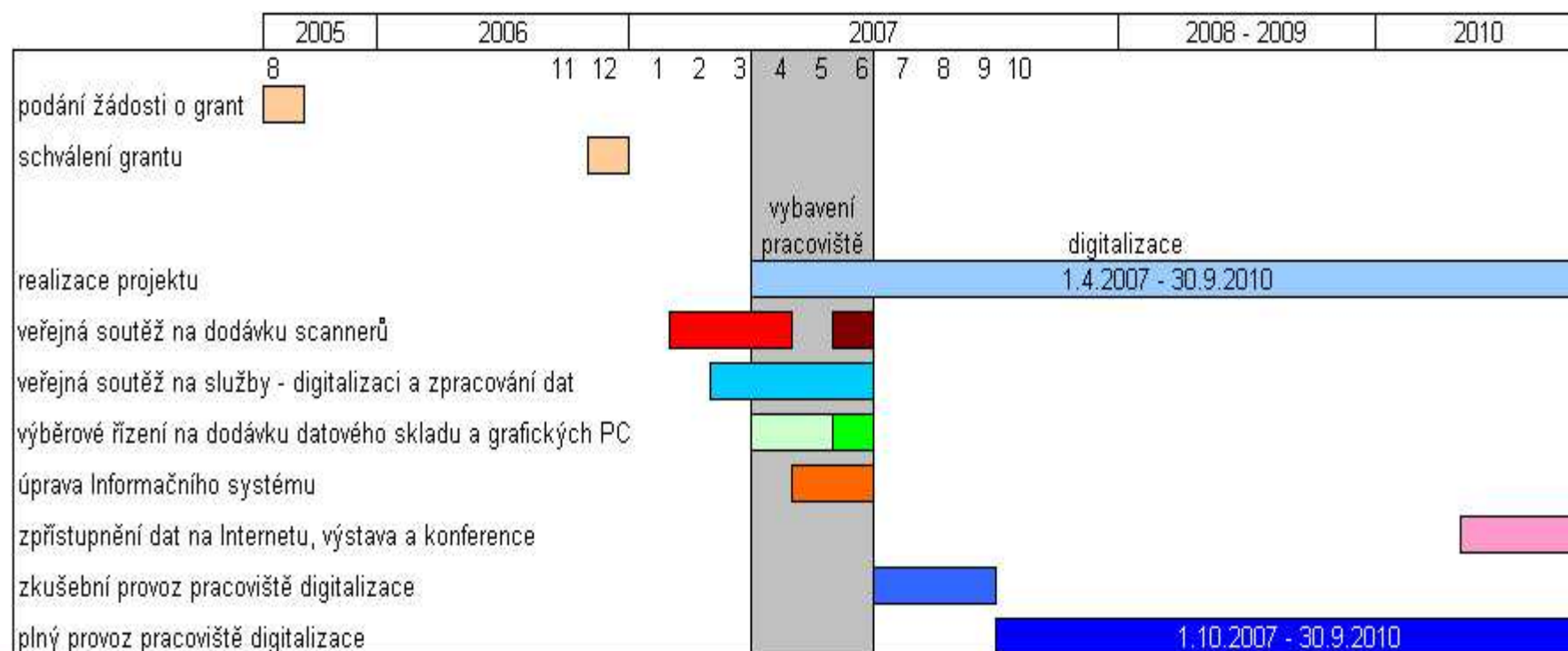


Časový harmonogram projektu a financování projektu

Doba trvání projektu: 1.4.2007 – 30.9.2010

Celkové náklady projektu: 778 855 EUR

Celková výše grantu: 483 591 EUR



Typy předloh k digitalizaci podle tvaru a velikosti:

- Knižní předlohy s vazbou (např. matriky, rukopisy)
- Volné listy (např. listiny)
- Velkoplošné předlohy (např. mapy a plány)
- Transparentní předlohy (např. skleněné negativy)

Technické vybavení – zařízení pro digitalizaci:

- 2 portálové skenery A1 600 DPI ProServ ScannTech 601i
 - + Vysoká kvalita výstupů, vysoké rozlišení
 - + Zajištění rovnoměrné geometrie a nasvícení předlohy
 - + Univerzální použití – knižní předlohy s vazbou (kolébka), plošné materiály
 - + Možnost nastavení přtlaku – šetrná manipulace
 - + Spolehlivost, rychlost a konektivita
 - Nutnost použití skla – riziko poškrábání
 - Malá hloubka ostrosti
 - skládání obrazu (3 CCD)
 - kalibrace CCD
- 1 portálový skener A2 600 DPI ProServ ScannTech 602i
- 1 plošný skener s nástavcem pro skenování transparentních předloh A3+ 2400 DPI Epson Expression 10000 XL



Technické vybavení - datové úložiště a zařízení pro archivaci

- Datový a aplikační server Hewlett Packard DL380 G5
- Zálohovací zařízení Hewlett Packard StorageWorks MSL4048 Ultrium 960
- Diskové pole RAID 6 Hewlett Packard 15TB
- Zařízení pro archivaci Plasmon Archive Appliance AA80-W2NDN2E

Záchranné digitální kopie archiválií ve formátu TIFF jsou spolu

s technickými a popisnými metadaty pro identifikaci archivována

na optická média UDO II typu WORM ve 2 kopiích.

Kapacita UDO II média je 60 GB a deklarovaná životnost média typu WORM je 180 let.



Výstupy digitalizace:

- Záchranné digitální kopie nejvzácnějších archiválií pořízené v maximální možné technické kvalitě se zachováním obsahu i formy předlohy
- Skeny v nahlížecí kvalitě pro účely dalšího využití a zpřístupnění
- Popisná metadata pro identifikaci archiválie
- Technická metadata

Formát výstupů:

- Záchranné digitální kopie
ve formátu TIFF group 6 bez komprese, 600 DPI (24 bitů/pixel, RGB)
fotografie ze skleněných negativů 1200 DPI
- Skeny v nahlížecí kvalitě
ve formátu JPEG 300 DPI
Optimální nastavení míry komprese pro dosažení
nejlepšího poměru kvality obrázku a jeho velikosti v Bytech je 83% (10)
- Technická metadata
pro jednotlivé obrázky se ukládají v metadatovém formátu **MIX v.2.0**.
Specifikace se odkazuje na standard popisu technických metadat
digitálních statických obrazů Z39.87.
- Popisná metadata pro identifikaci
ve formátu XML podle vlastního schématu IDA AMP v.2.0
(Identifikace archiválie AMP)
Definice vazby obrázek – metadatový záznam v katalogu



Srovnání grafických formátů (pro rastrovou 2D grafiku) TIFF a JPEG

TIFF group 6 bez komprese

- + Není nutné pro účely archivace řešit migraci formátu
- + Vysoká podpora napříč všemi počítačovými platformami a operačními systémy
- + Případná ztráta dat má zpravidla malý vliv na změnu původního obrazu
- Vysoký objem dat geometricky úměrný rozlišení resp. fyzické velikosti předlohy
- Z výše uvedeného důvodu malá podpora v prostředí Internetu
- Časově a výkonově náročné počítačové zpracování

JPEG

- + Malý objem dat
- + Podpora na Internetu
- Ztrátová komprese – každou další úpravou obrázku dochází k jeho degradaci
- Není vhodný pro reprodukci textu
- Složitý kompresní algoritmus DCT
- Případná ztráta dat zásadně ovlivní výsledný obraz



Co se archivuje:

- Digitální objekty (skeny)
- Technická metadata pro každý sken
- Popisná metadata pro základní identifikaci archiválie

Příklad uložení dokumentů v souborovém systému na archivační médium:

```
AMP/      00153/   ida.xml   { popisná metadata sady 00153 pro základní identifikaci
                                archiválie nebo archiválií }
                                00001.tif { naskenovaný obrázek ve formátu TIFF }
                                00001.xml { technická metadata obrázku 00001.tif ze sady 00153 }
                                00002.tif
                                00002.xml
                                00003.tif
                                00003.xml
                                00004.tif
                                00004.xml
                                00005.tif
                                00005.xml

          00154/   ida.xml   { popisná metadata sady 00154 pro základní identifikaci
                                archiválie nebo archiválií }
                                00001.tif { naskenovaný obrázek ve formátu TIFF }
                                00001.xml { technická metadata obrázku 00001.tif ze sady 00154 }
                                00002.tif
```



Archivace

Archivace dat: dlouhodobé uložení dat se zajištěním:

- Udržitelnosti
 - Formátu – TIFF group 6 bez komprese
 - Datového archivačního média
životnost UDO WORM média až 180 let
- Neměnnosti (autentičnosti) dokumentu
média typu WORM, hashovací řetězec
- Důvěryhodnosti – historická sbírka, garant instituce

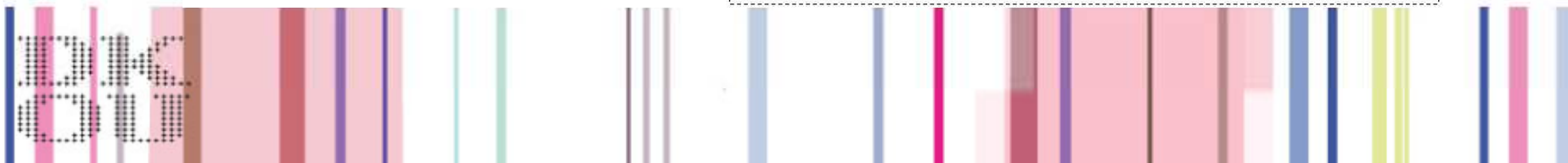
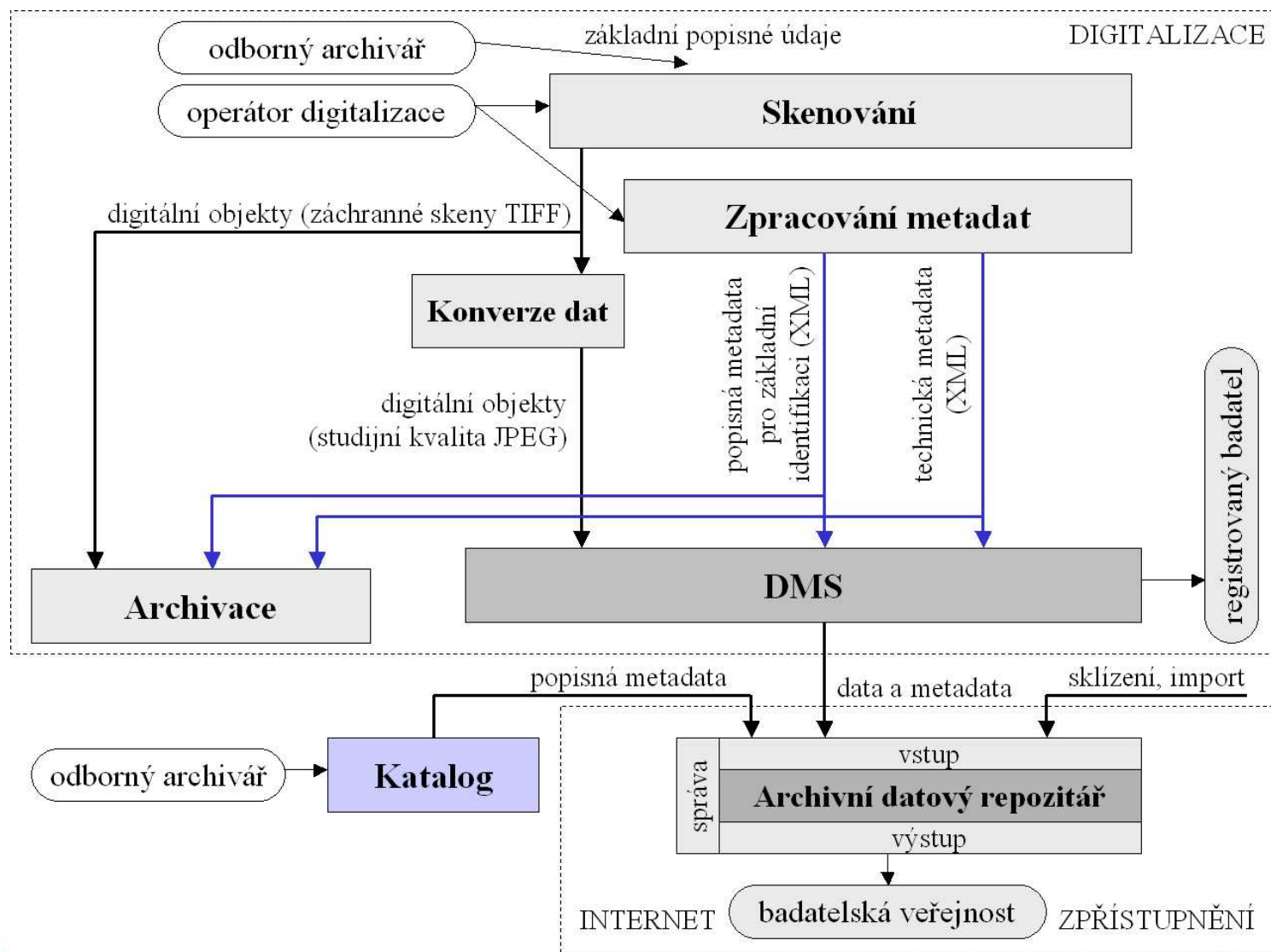


Technologie:

- Archivace na optická média UDO II typu WORM ve 2 kopiích (primární a sekundární řada). Kapacita UDO II média je 60 GB a udávaná životnost 180 let.
- Archivace souborů ve formátu TIFF spolu s technickými a popisnými metadaty pro identifikaci
- Použití kontrolního řetězce (64 znaků) generovaného na základě hashovací funkce sha-256 (jednocestná funkce) používající pro výpočet hashovacího řetězce 256 bitový klíč – součást technických metadat
- Oddělení dat (obrázků) a metadat (popisu) – obsah některých elementů se mění



Model digitalizace v Archivu hlavního města Prahy



1. Webové stránky projektu Ad fontes

www.ahmp.cz/adfontes

Archiv hlavního města Prahy - Mozilla Firefox

Soubor Úpravy Zobrazení Historie Záložky Nástroje nápověda

http://www.ahmp.cz/

ARCHIV HLAVNÍHO MĚSTA PRAHY

O nás Návštěva archivu Fondy a pomůcky **Projekty a výzkum** Předarchivní péče Nabídka publikací

hledat

mapa webu

Projekty a výzkum

- **Projekty**
- Konference
 - 2009
- Mezinárodní spolupráce
- Ediční činnost
 - Pražský sborník historický
 - Redakce a redakční rada
 - Pokyny pro autory
 - Obsah vydaných čísel
 - Documenta Pragensia
 - Redakce a redakční rada
 - Pokyny pro autory
 - Obsah vydaných čísel
- Dějiny Prahy
 - Vývoj pražské samosprávy
 - Správní a územněsprávní členění Prahy
 - Přehled primasů a primátorů Prahy
- Kroužek pro dějiny raného novověku
 - archiv přednášek

AD FONTES

Projekt hromadné digitalizace

Projekt komplexní digitální ochrany, záchrany a obecného zpřístupnění nejvýznamnějších historických dokumentů hlavního města Prahy.

Hlavní cíle projektu:

- Záchrana a trvalé uchování historicky nejvýznamnějších, památkově nejcennějších a fyzicky nejvíce ohrožených archiválií ze sbírek a fondů Archivu hlavního města Prahy, významných pro národní i evropské dějiny, cestou vytvoření všestranně využitelných digitálních kopií z originálů těchto archiválií
- Zpřístupnění archiválií v digitální podobě pro badatele a širokou veřejnost

Zaměření/koncepce:

- Uchování evropského kulturního dědictví
- Zlepšení péče a ochrana movitého kulturního dědictví
- Ochrana vzácných písemných dokumentů Prahy
- ePraha

Výběr archiválií a jeho kritéria:

- historická dokumentární hodnota archiválie
- míra reálného nebo možného fyzického poškození
- frekvence jejího využívání badatelskou veřejností

Sbírka matrik

Sbírka rukopisů

Stará konškrípce

Spolek výtvarných umělců Mánes Praha

Protokoly samosprávných sborů

Heraldická (Rentzova) sbírka

Divadlo Labyrint

Společnost pro povznesení německé vědy, umění a literatury v Čechách

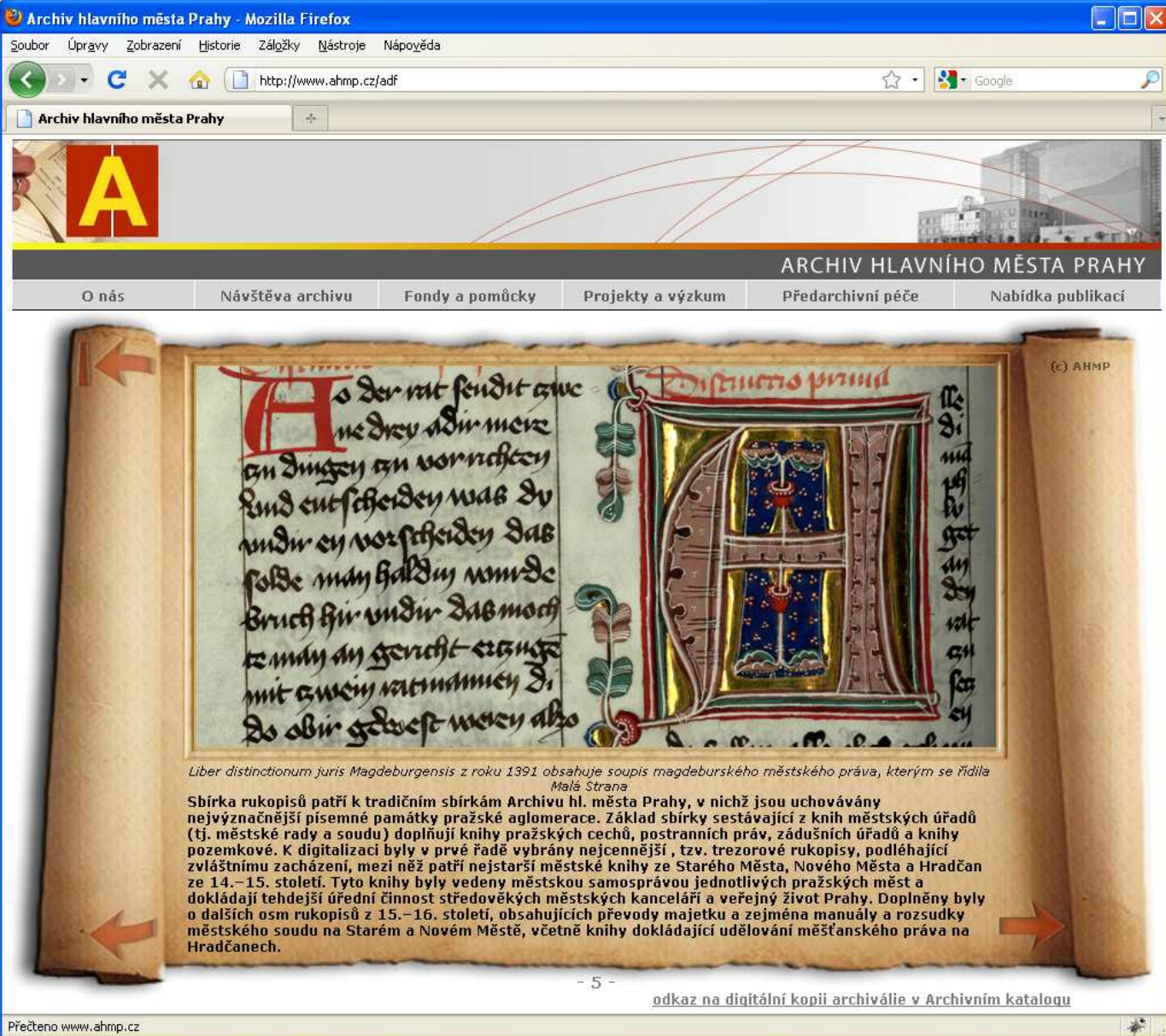
Archiv hlavního města Prahy
Archivní 6, 149 00 Praha 4

Hotovo

2. Prezentace výstupů projektu Ad fontes

www.ahmp.cz/adf

Flash prezentace popisující fondy a sbírky vybrané k digitalizaci v rámci projektu



Archiv hlavního města Prahy - Mozilla Firefox

Soubor Úpravy Zobrazení Historie Záložky Nástroje nápověda

http://www.ahmp.cz/adf

ARCHIV HLAVNÍHO MĚSTA PRAHY

O nás Návštěva archivu Fondy a pomůcky Projekty a výzkum Předarchivní péče Nabídka publikací

Ao der rat sendit que
ne drey adir meiz
an dungen an vorracheen
und entscheyden was dy
undir en vor scheyden das
solde man halduy nomde
bruch hir undir das mocht
te man an gericht eruge
mit einem ratmannen di
do obir gelvest wachen also

Distinctiones primae

(c) AHMP

Liber distinctionum juris Magdeburgensis z roku 1391 obsahuje soupis magdeburského městského práva, kterým se řídila Malá Strana

Sbírka rukopisů patří k tradičním sbírkám Archivu hl. města Prahy, v nichž jsou uchovávány nejvýznačnější písemné památky pražské aglomerace. Základ sbírky sestávající z knih městských úřadů (tj. městské rady a soudu) doplňují knihy pražských cechů, postranních práv, zádušních úřadů a knihy pozemkové. K digitalizaci byly v první řadě vybrány nejcenější, tzv. trezorové rukopisy, podléhající zvláštnímu zacházení, mezi něž patří nejstarší městské knihy ze Starého Města, Nového Města a Hradčan ze 14.–15. století. Tyto knihy byly vedeny městskou samosprávou jednotlivých pražských měst a dokládají tehdejší úřední činnost středověkých městských kancelářů a veřejný život Prahy. Doplněny byly o dalších osm rukopisů z 15.–16. století, obsahujících převody majetku a zejména manuály a rozsudky městského soudu na Starém a Novém Městě, včetně knihy dokládající udělování měšťanského práva na Hradčanech.

- 5 -

[odkaz na digitální kopii archiválie v Archivním katalogu](#)

Přečteno www.ahmp.cz

3. doprovodná výstava v konf. sále Archivu hl.m. Prahy, Praha 4 Chodovec

Ukázky originálů vybraných archiválií digitalizovaných fondů a sbírek
Srovnání s vytištěnými digitalními kopiemi archiválií



4. Zpřístupnění digitálních kopií archiválií v Archivním katalogu Praga Publica Historica www.ahmp.cz/katalog

podzim 2010:
5. DVD
6. kniha

The screenshot shows a Mozilla Firefox browser window displaying the website 'Archiv hlavního města Prahy - Mozila Firefox'. The address bar shows 'http://www.ahmp.cz/katalog'. The website header features a large 'A' logo and the text 'ARCHIVNÍ KATALOG' and 'ARCHIV HLAVNÍHO MĚSTA PRAHY'. Below the header is a navigation menu with links: 'O nás', 'Návštěva archivu', 'Fondy a pomůcky', 'Projekty a výzkum', 'Předarchivní péče', and 'Nabídka publikací'. The main content area is titled 'Digitalizované' and 'Procházet a hledat'. It displays a grid of 12 digitalized archival collections, each with a small image and a title:

- Sbírka matrik
- Sbírka pergamenových listin
- Sbírka fotografií
- Sbírka papírových listin
- Sbírka map a plánů
- Soupis pražských domovských příslušníků (1830-1910)
- Sbírka rukopisů
- Protokoly schůzí městské rady (1784-1814)
- Heraldická sbírka
- Spolek výtvarných umělců Mánes
- Společnost pro povznesení německé vědy, umění a literatury v Čechách
- Všeobecná židovská nemocnice

At the bottom left of the browser window, the text 'Hotovo' is visible.



Organizace pořizovaných skenů a identifikace

Sady:

- soubor archivního materiálu, který lze identifikovat a popisovat
- knihy (např. rukopisy, matriky), soubor jednotlivin (např. stavební plány)
- zakládá se před skenováním
- 1 sada = 1 adresář s názvem sady
- název sady: 5 ciferné přírustkové číslo strojově generované IS
- Sada obsahuje n skenů, n technických metadat, 1 soubor s popisnými metadaty

Archiválie:

- ID může vycházet z ID sady pro referencování v rámci metadat
- Problematika perzistentních identifikátorů pro sdílení a zpřístupnění

Skeny (objekty):

- Název souboru: 5 ciferné přírustkové číslo 00001.tif – nnnnn.tif
- Názvy souborů neobsahují odkazy na signatury či inventární čísla
- Obsah objektů v popisných metadatech

Identifikace objektů (ID):

- jedinečný identifikátor
- 32 znaků
- generovaný algoritmicky
- není nutné centrální přidělení
- vazba metadata – data – záznamy v databázi
- možnost off-line zpracování dat



Popisné údaje sady

Obsah sady je vhodné popisovat po částech:

- **Popis sady:** informace o vytvoření celé sady

Typické položky:

označení sady, datace, jméno operátora digitalizace, zařízení pro digitalizaci

- **Popis archiválie:** informace identifikující digitalizovanou předlohu. Tyto údaje musí poskytnout příslušný archivář, do systému/tabulky je pak zanáší digitalizátor

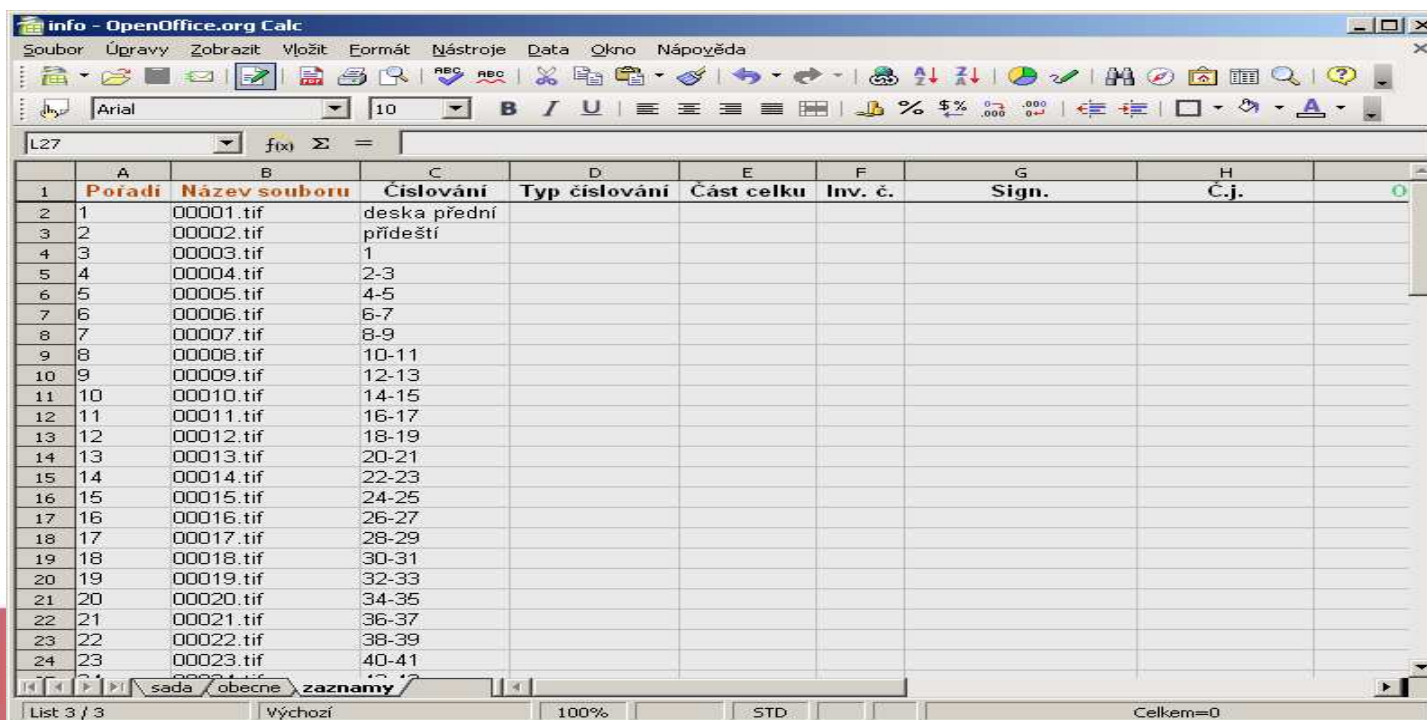
Typické položky:

číslo NAD a/nebo název fondu/sbírky, část fondu, manipulace, úložná jednotka, inv. č., signatura, č. jednací, typ, poznámka

- **Popis záznamů (skenů):** informace o jednotlivých snímcích (např. stranách knihy)

Typické položky:

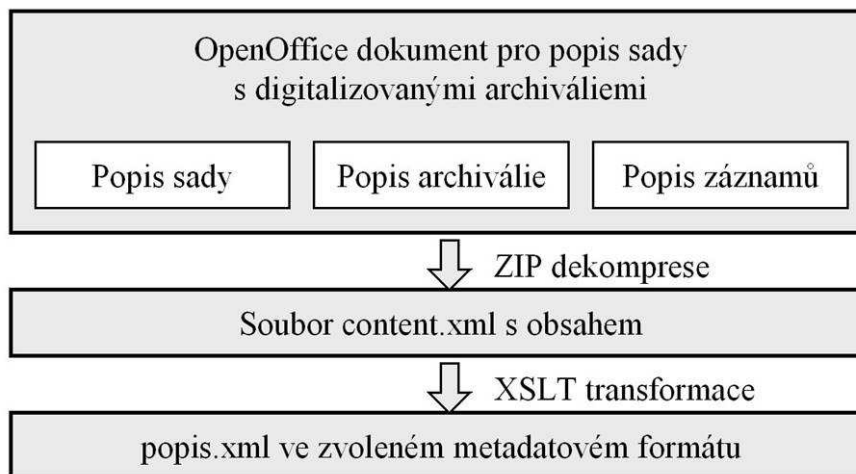
pořadí, název souboru, číslování, typ číslování, část celku, poznámka



	A	B	C	D	E	F	G	H
	Pořadí	Název souboru	Číslování	Typ číslování	Část celku	Inv. č.	Sign.	Č.j.
1	1	00001.tif	deska přední					
2	2	00002.tif	přídeščí					
3	3	00003.tif	1					
4	4	00004.tif	2-3					
5	5	00005.tif	4-5					
6	6	00006.tif	6-7					
7	7	00007.tif	8-9					
8	8	00008.tif	10-11					
9	9	00009.tif	12-13					
10	10	00010.tif	14-15					
11	11	00011.tif	16-17					
12	12	00012.tif	18-19					
13	13	00013.tif	20-21					
14	14	00014.tif	22-23					
15	15	00015.tif	24-25					
16	16	00016.tif	26-27					
17	17	00017.tif	28-29					
18	18	00018.tif	30-31					
19	19	00019.tif	32-33					
20	20	00020.tif	34-35					
21	21	00021.tif	36-37					
22	22	00022.tif	38-39					
23	23	00023.tif	40-41					
24	24	00024.tif	42-43					

Transformace popisu

(s použitím dokumentu OpenOffice)



Popisná metadata pro základní identifikaci

Varianty popisu sady z hlediska vztahu k archiválii:

1 sada = 1 archiválie
(např. rukopis, matrika):

```
<sada>
  <arch>
    <zaznamy>
      <zaznam1 />
      <zaznam2 />
      ...
    </zaznamy>
  </arch>
</sada>
```

1 sada obsahuje více archiválií
(např. sada fotografií):

```
<sada>
  <arch1>
    <zaznamy>
      <zaznam1 />
      <zaznam2 />
      ...
    </zaznamy>
  </arch1>
  <arch2>
    <zaznamy>
      <zaznam1 />
      <zaznam2 />
      ...
    </zaznamy>
  </arch2>
</sada>
```

1 archiválie obsahuje více sad
(např. doskenovaný rukopis):

```
<sada1>
  <arch>
    <zaznamy>
      <zaznam1 />
      <zaznam2 />
    </zaznamy>
  </arch>
</sada1>
<sada2>
  <arch>
    <zaznamy>
      <zaznam3 />
      <zaznam4 />
    </zaznamy>
  </arch>
</sada2>
```

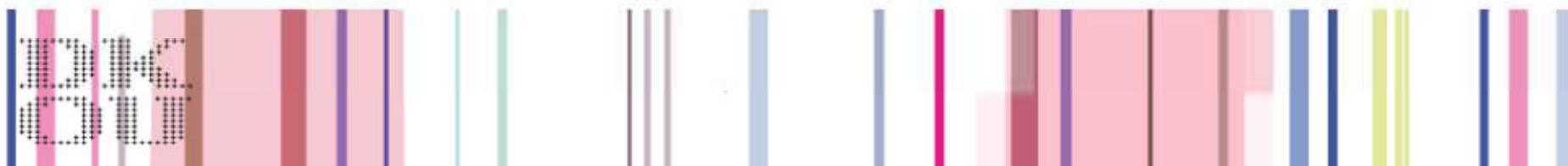


Technická metadata

Technická metadata pro jednotlivé skeny v kvalitě určené pro archivaci se ukládají v metadatovém formátu **MIX v.2.0**. Specifikace se odkazuje na standard popisu technických metadat digitálních statických obrazů Z39.87.

Technická metadata by měla obsahovat alespoň elementy obsahující tyto údaje:

- Identifikátor objektu (popisovaného obrázku) v rámci použitého systému evidence digitalizace
- Použitý grafický formát a jeho vlastnosti
- Použitý barevný model
- Kontrolní hashovací řetězec zajišťující autentičnost dokumentu
- Rozměry obrázku
- ICC profil
- Zařízení na kterém byl sken pořízen a jeho základní parametry
- Název a verzi skenovacího softwaru a grafického editoru
- Výčet transformací pro úpravu obrázku
- Subjekt který sken pořídil
- Datum naskenování



Partnerské projekty:

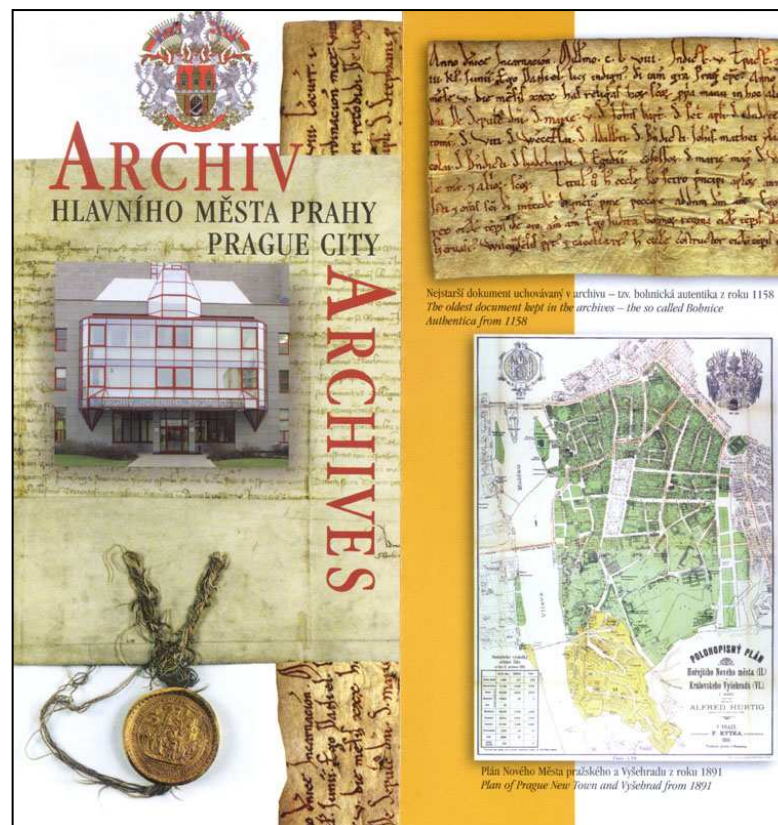
- Společný projekt Národního archivu a Archivu hlavního města Prahy: Možnosti a formy zpřístupnění archivních fondů nebo jejich součástí veřejnosti v elektronické podobě.
- Praga Publica Historica
Systém pro zpřístupnění archiválií v Archivním katalogu na Internetu
- Koordinační komise pro projekty schválené v první výzvě FM EHM/Norsko.
Výměna zkušeností, společná propagace

členové komise:

- Archiv hlavního města Prahy
- Městská knihovna v Praze
- Národní knihovna ČR

koordinátor:

- odbor umění a knihoven Ministerstva kultury ČR.





Podpořeno grantem z Norska
prostřednictvím
Norského finančního mechanismu

Zítra ve středu 16.6.2010 exkurze na pracovištích digitalizace:

- Městské knihovny v Praze
 - Archivu hlavního města Prahy
 - Národní knihovny ČR
 - Knihovny Akademie věd ČR
- V Archivu hlavního města Prahy je kromě exkurze na pracovišti digitalizace připravena pro účastníky konference doprovodná výstava výstupů projektu Ad fontes a ozvučená počítačová prezentace.

Tomáš Hanousek

Email: tomas.hanousek@cityofprague.cz

Archiv hlavního města Prahy

<http://www.ahmp.cz>

<http://www.ahmp.cz/adfontes>

