Standardy digitalizace a budování digitální knihovny Kramerius

Konference Historické fondy UK

Milan Janíček
milan.janicek (at) ruk.cuni.cz

Eliška De Herdt
eliska.de.herdt (at) ruk.cuni.cz

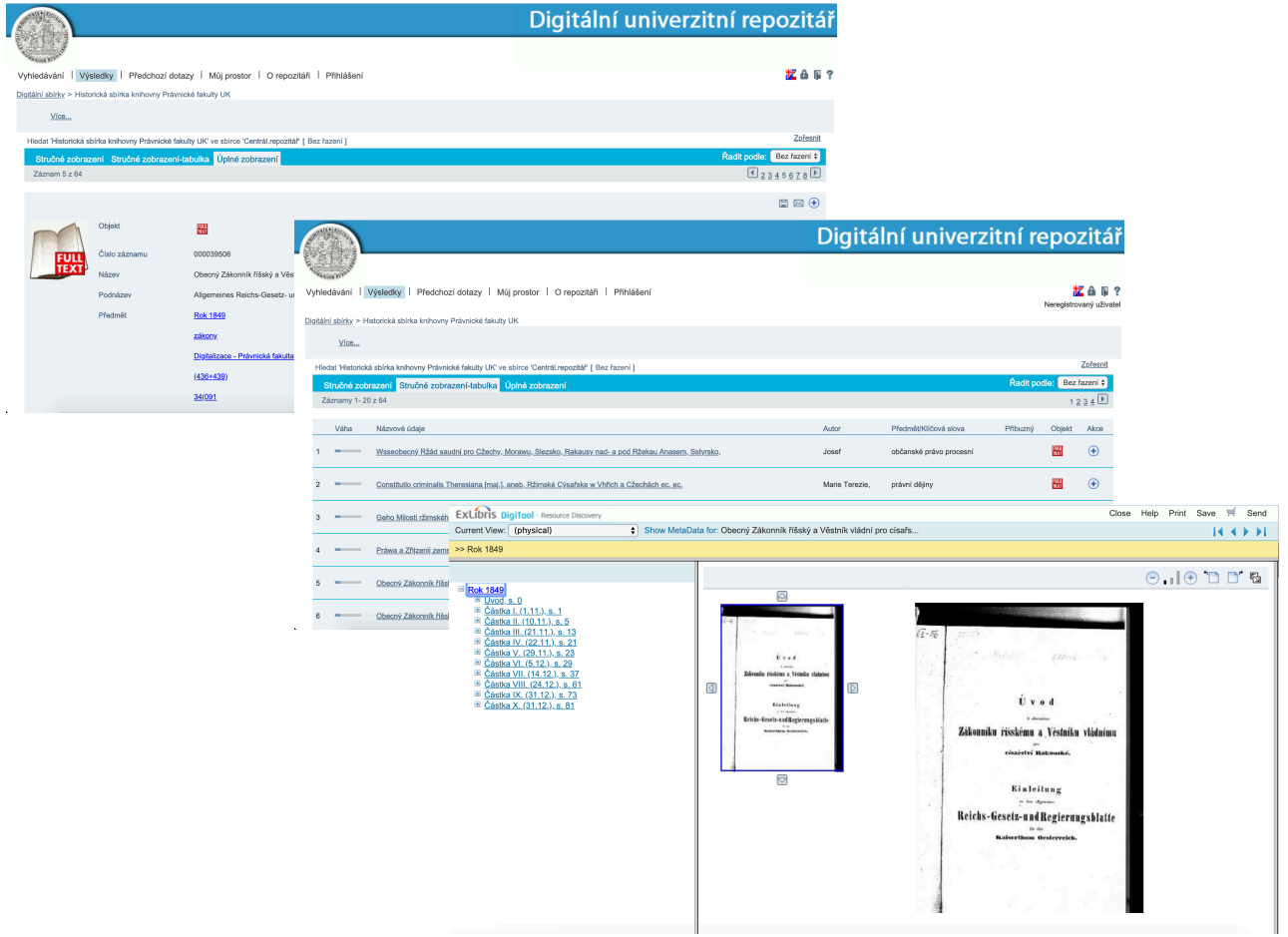
UNIVERZITA KARLOVA
Ústřední knihovna Univerzity Karlovy

Historie digitalizace na UK

- digitalizace probíhala řadu let v rámci různých projektů
  - mapy a další dokumenty Mapové sbírky
  - historická produkce univerzity
    - ročenky, studijní katalogy...
  - historické časopisy vydávané na fakultách
  - tematické sbírky
    - Říšský zákoník, Knihovna osobností
    - teologické časopisy
- od roku 2005 se digitalizované dokumenty zpřístupňují v systému DigiTool provozovaným Ústřední knihovnou UK

DigiTool

- komerční software (od dodavatele Alephu)
- „Cílem repozitáře je zejména dlouhodobé uchování a ochrana dokumentů...“
- vývoj skončil v roce 2009
- UK používá dodnes



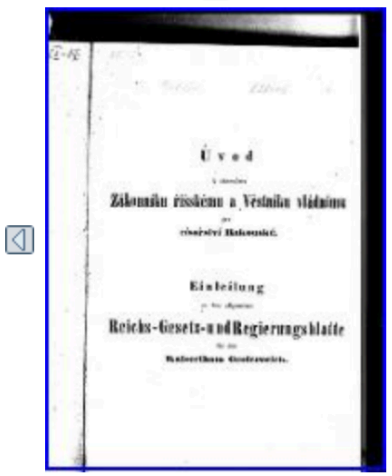
ExLibris DigiTool - Resource Discovery

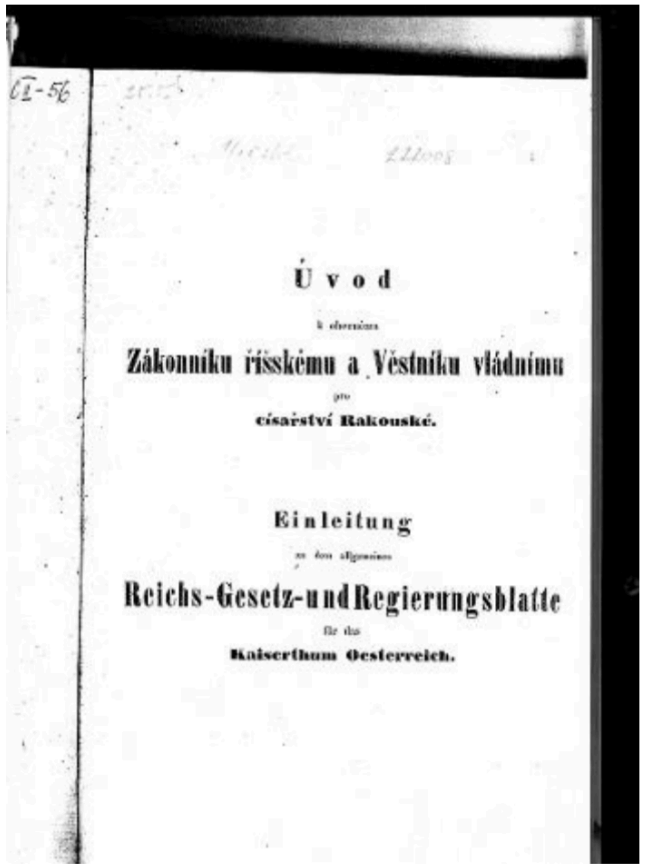
Close Help Print Save Send

Current View: (physical) Show MetaData for: Obecný Zákonník říšský a Věstník vládní pro císařs...

>> Rok 1849

- Rok 1849
  - Úvod, s. 0
  - Částka I. (1.11.), s. 1
  - Částka II. (10.11.), s. 5
  - Částka III. (21.11.), s. 13
  - Částka IV. (22.11.), s. 21
  - Částka V. (29.11.), s. 23
  - Částka VI. (5.12.), s. 29
  - Částka VII. (14.12.), s. 37
  - Částka VIII. (24.12.), s. 61
  - Částka IX. (31.12.), s. 73
  - Částka X. (31.12.), s. 81





Úvod

k obecnému

Zákoníku říšskému a Věstníku vládnímu

pro

císařství Rakouské.

Einleitung

zu dem allgemeinen

Reichs-Gesetz-und Regierungsblatte

für das

Kaiserthum Oesterreich.

Open Source pro digitální knihovnu

- DSpace & Kramerius

  + rozsáhlé spolupracující komunity

    DSpace – univerzitní prostředí

    Kramerius – české knihovní prostředí

  + pokračující vývoj, možnost vlastních změn nebo ovlivnění směřování systému

  + nejsou licenční poplatky

  - (z hlediska předchozího směřování repozitáře) větší důraz na uživatelskou prezentaci, ne na LTP

  - větší náročnost personálního zajištění

Kramerius

- známý především z českých knihoven (a několika muzeí)
- dlouholetý vývoj - 2003 → 2008+ → 2019...
- vývojová skupina – zástupci knihoven
  - zejména KNAV, MZK, NK
- využívá se v projektu Národní digitální knihovny
- postupně vznikají další části ekosystému
  - ProArc – produkční a archivační systém
  - ArcLib – řešení dlouhodobé archivace
- nedávno významný update uživatelského rozhraní

Digitální knihovna 1.LF UK

Hledat v digitální knihovně ...

Sbírky Procházet Informace English

Výsledky: 1 - 25 / **25**

22 3 0 0 0 0 0 0

Řazení: Nově přidané

Dostupnost

pouze veřejné	24
pouze neveřejné	1
Vše	**25**

Klíčové slovo

Lékařské vědy. Lékařství	11
Staré tisky	8
Lékařství	6
Anatomie člověka a srovnávací anatomie	5
Anatomie	4
Chemie	4

Zobrazit více

Autor

Horbaczewski, Jan, 1854-1942	4
Adachi, Buntaro	1
Adachi, Buntaro, 1865-1945	1
Arnold, Friedrich, 1803-1890	1
Froriep, Robert , 1804-1861	1

LE SPECTACLE DE LA NATURE, OU ENTRETIENS

Pluche, Antoine, 1688-1761

1733

Kniha

Katalog des Lesemuseums des medicinischen

1888

Kniha

Sborník lékařský =: Archives bohémes de

Univerzita Karlova Lékařská

1908-2003

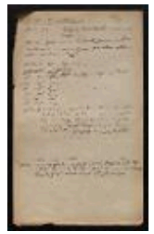
Noviny a časopisy

Fragment der Rede Ilg's gelegentlich der Eröffnung

Ilg, Johann Georg (1771-

1877

Kniha



Acta facultatis medicae Pragensis

[1900-1950?]

Kniha

Das Arteriensystem der Japaner

Kniha

Das Lymphgefässsystem der Japaner

Kihara, Takusaburo

1953

Kniha

Sborník klinický: časopis pro pěstování vědy

1900-1907

Noviny a časopisy

Das Venensystem der Japaner

Adachi, Buntaro, 1865-1945

1933

Kniha

Kramerius z pohledu UK

- není tak obecný jako DigiTool
  - omezené množství používaných formátů dokumentů (knihy, periodika, mapy, grafiky...)
  - zpřístupnění (typicky) po digitalizovaných stránkách
- není moc vhodný pro kategorii nejstarších historických dokumentů
  - Manuscriptorium má pro tento typ obsahu i lepší nástroje (jazyk TEI)
    - spolupráce Kramerius & Manuscriptorium je v současnosti předmětem diskuzí

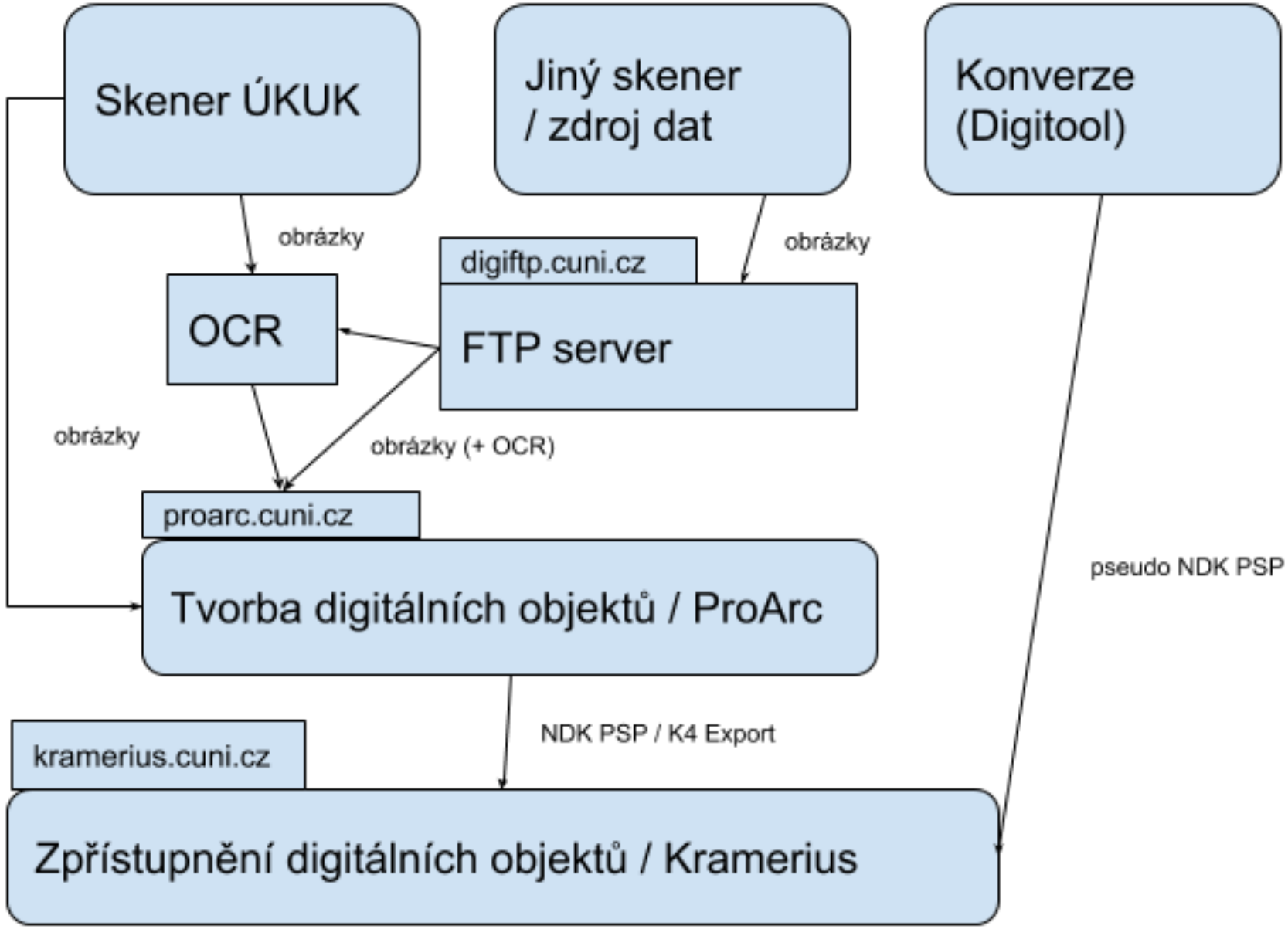
Kramerius z pohledu UK (2)

- Kramerius je svým způsobem kompromis
  + příjemné uživatelské rozhraní pro běžného uživatele
  + velmi dobré zobrazování dokumentů, propojení s Imageserverem (= zoomovatelné obrázky), ALTO OCR
  + de facto český standard
  + dobrá architektura (z počítačového pohledu), potenciál dalšího rozšiřování funkcí
  + pokračují práce na dalších důležitých nástrojích (LTP)
  → dobrý poměr cena/výkon, ideální pro „typické“ digitalizované dokumenty: 1800 → současnost

Podpora digitalizace na UK

- pracuje se na vytvoření digitalizačního pracoviště
- vytvoření centrální instalace Krameria (a ProArcu) v ÚK UK

- nová digitalizace – Ústřední knihovna nabízí podporu fakultním knihovnám:
  - specifikace standardů a požadavků
  - vyřešení „společného technického základu“, řešení problémů
  - možnost efektivně zveřejnit digitalizované dokumenty
  - workflow →


Skener ÚKUK
Jiný skener / zdroj dat
Konverze (Digitool)
obrázky
obrázky
digiftp.cuni.cz
OCR
FTP server
obrázky
obrázky (+ OCR)
proarc.cuni.cz
Tvorba digitálních objektů / ProArc
pseudo NDK PSP
NDK PSP / K4 Export
kramerius.cuni.cz
Zpřístupnění digitálních objektů / Kramerius

Dělba práce...

- (typický) příklad:
  - požadavky na skeny: **ústřední** knihovna
  - skenování: **fakultní** knihovny
  - záznamy v knihovním systému: **fakultní** knihovny
  - uložení skenů do systémů: **ústřední** knihovna
  - zpracování skenů (ProArc): **fakultní** knihovny
  - zpřístupnění v Krameriovi: **ústřední** knihovna
  - další služby zpřístupnění: **ústřední** knihovna

Požadavky na digitalizaci – ÚK UK

- jednoduchá verze pro součásti UK:

Konzultujte s námi prosím, než začnete vytvářet data určená pro naše systémy.

Požadavky na digitalizaci – ÚK UK (2)

- vycházíme z požadavků Národní digitální knihovny
  - NDK je významný projekt
  - Kramerius a ProArc podporují standardy NDK
  - naše data budou lépe interoperabilní

  - nejsme k použití standardů vyloženě vázáni
  - nicméně je chceme využít z praktických důvodů

➔ u nové digitalizace chceme NDK

Požadavky: kvalita obrazu

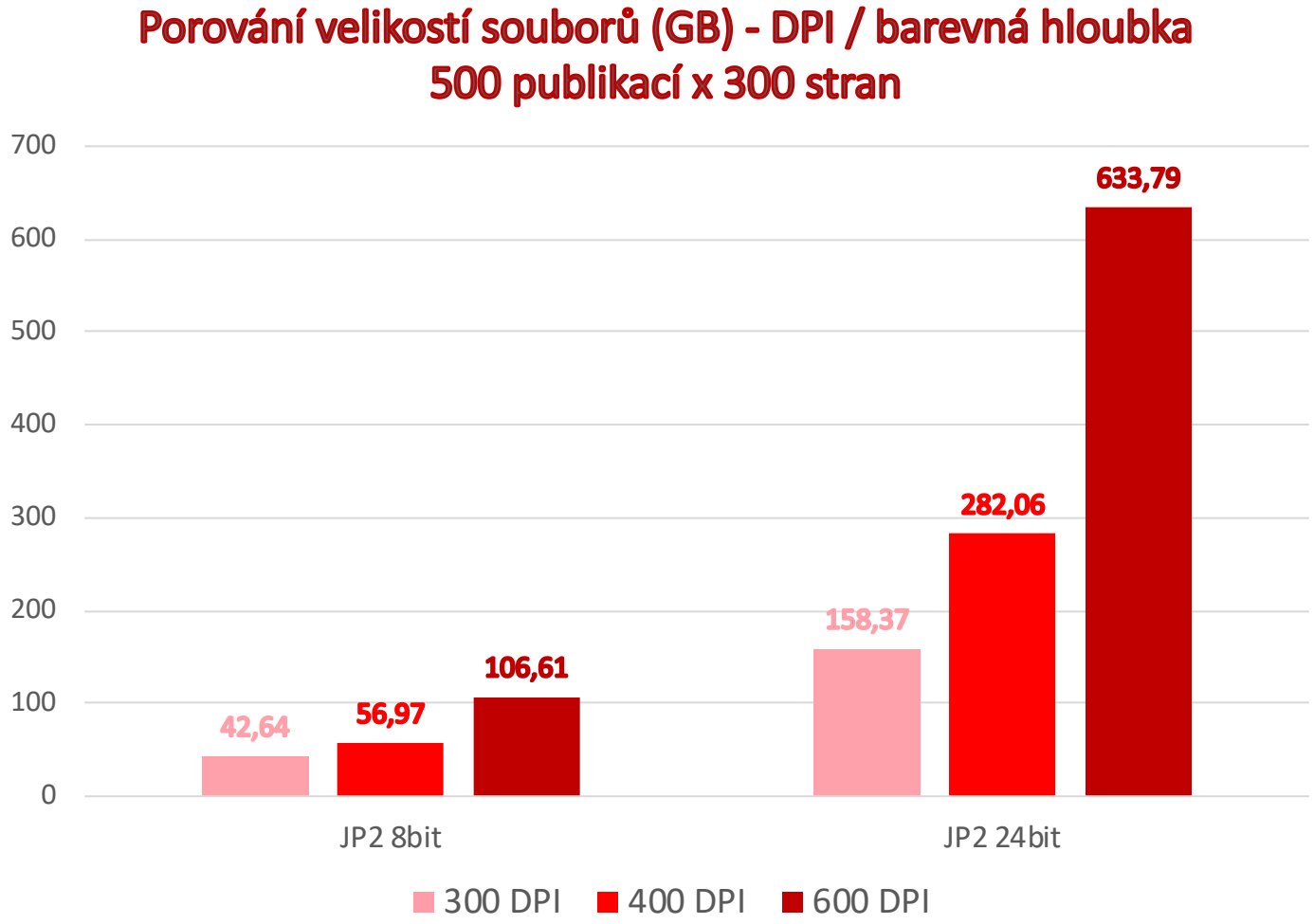
- nastavení skeneru
  - formát : TIFF* (JP2)
  - komprese: bezztrátová
  - ICC profily: Ano
  - DPI: 300 DPI (pref. 400 DPI)
  - barevná hloubka: 8 bit***, 24 bit

- * - odchylky oproti standardům NDK
  - * TIFF si zkonvertujeme do JP2 sami
  - ** ...

Požadavky: barevná hloubka

- 24 bitů – 16 777 216 barev
- 8 bitů – 256 barev
- významně ovlivňuje velikost souborů

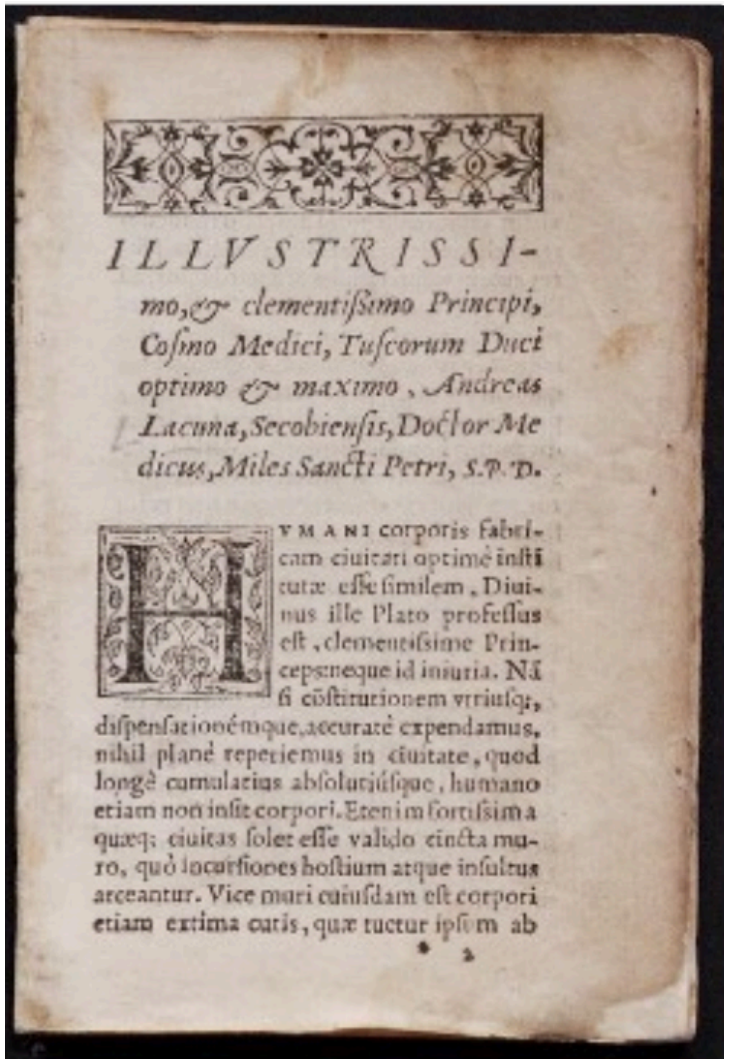
- existují dokumenty, kde není potřeba velká barevná hloubka



Požadavky: obraz

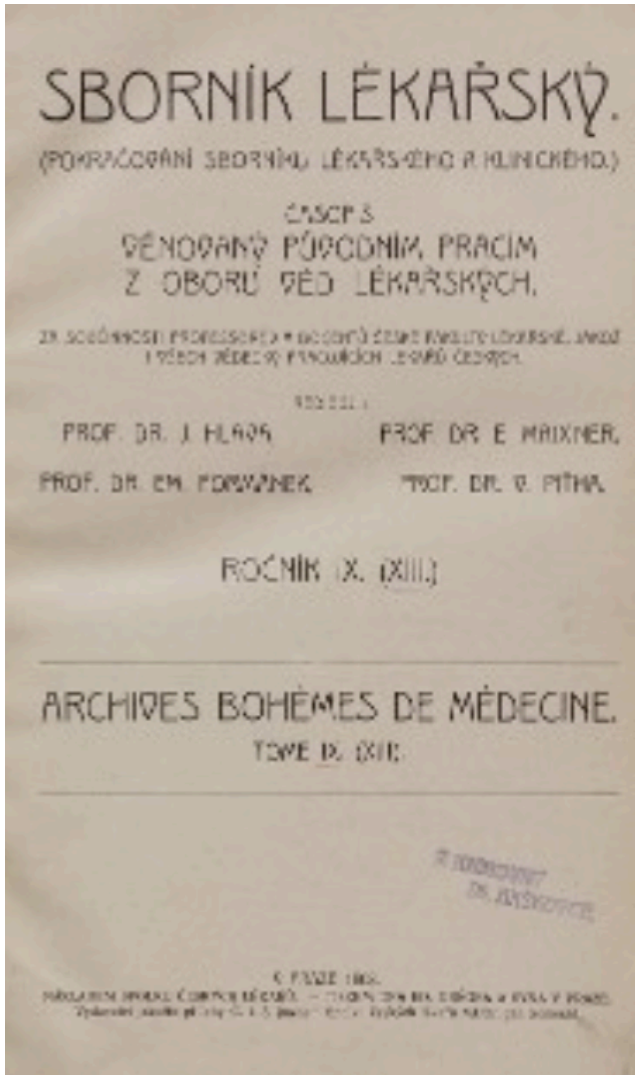
- nepodcenit kalibraci skeneru
  - špatná kalibrace může způsobit
    - rozmazání
    - barevné deformace
    - prostorové deformace
    - barevný posun oproti originálu
- ořez
  - novodobá digitalizace – dovnitř
  - historická digitalizace – ven, případně i s přesahem do druhého listu

Ořezy – vnější, vnitřní



ILLVSTRISSI-
mo, & clementissimo Principi,
Cosmo Medici, Tuscorum Duci
optimo & maximo, Andreas
Lacuna, Secobiensis, Doctor Me
dicus, Miles Sancti Petri, S.P.D.

HVMANI corporis fabricam ciuitati optimè institutæ esse similem, Diuinus ille Plato professus est, clementissime Princeps: neque id iniuria. Nã si cõstitutionem vrbium, dispensationémque, accuratè expendamus, nihil planè reperiemus in ciuitate, quod longè cumulatius absolutiúsque, humano etiam non insit corpori. Etenim fortissima quæq; ciuitas solet esse valido cincta muro, quò incursiones hostium atque insultus arceantur. Vice muri cuiusdam est corpori etiam extima cutis, quæ tuetur ipsum ab



SBORNÍK LÉKAŘSKÝ.

(POKRAČOVÁNÍ SBORNÍKU LÉKAŘSKÉHO A KLINICKÉHO.)

ČÁSTEČNĚ

VĚNOVANÝ PŮVODNÍM PRACÍM
Z OBORU VĚD LÉKAŘSKÝCH.

PROF. DR. J. HLAVA PROF. DR. E. MAIXNER.

PROF. DR. EM. PURKYNĚ. PROF. DR. V. PÍTHA.

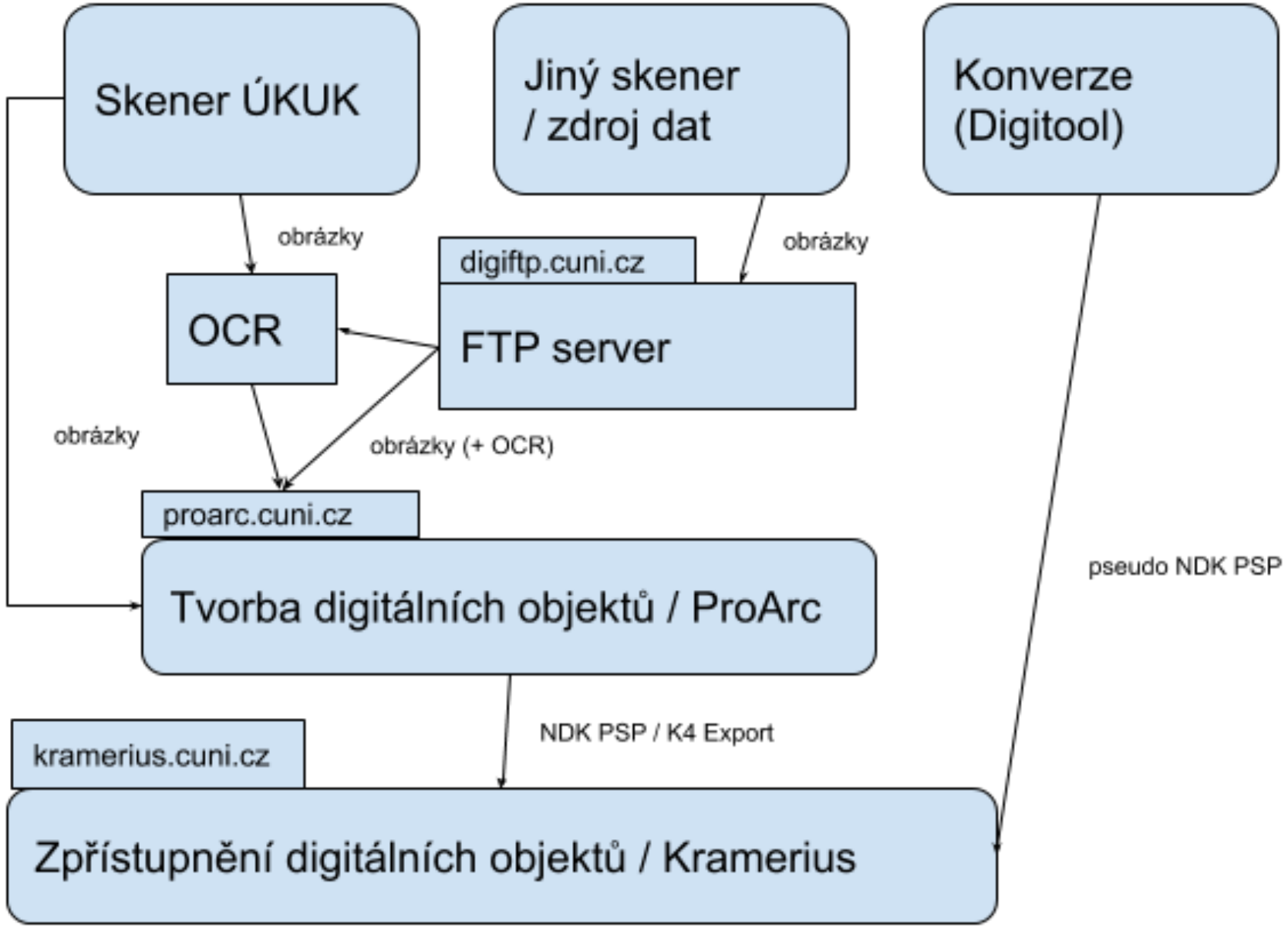
ROČNÍK IX. (XIII.)

ARCHIVES BOHÊMES DE MÉDECINE.

TOME IX. (XIII.)

Požadavky: soubory

- názvová konvence
  - sysno_název_nepovinnýText_###
    → lepší orientace v souborech
- kontrolní součet MD5
  - MD5 – „otisk“ souboru
  - umožňuje kontrolu integrity dat
- OCR
  - rozpoznávání textu
  - ideálně TXT a ALTO XML
  - u historických není vždy možné

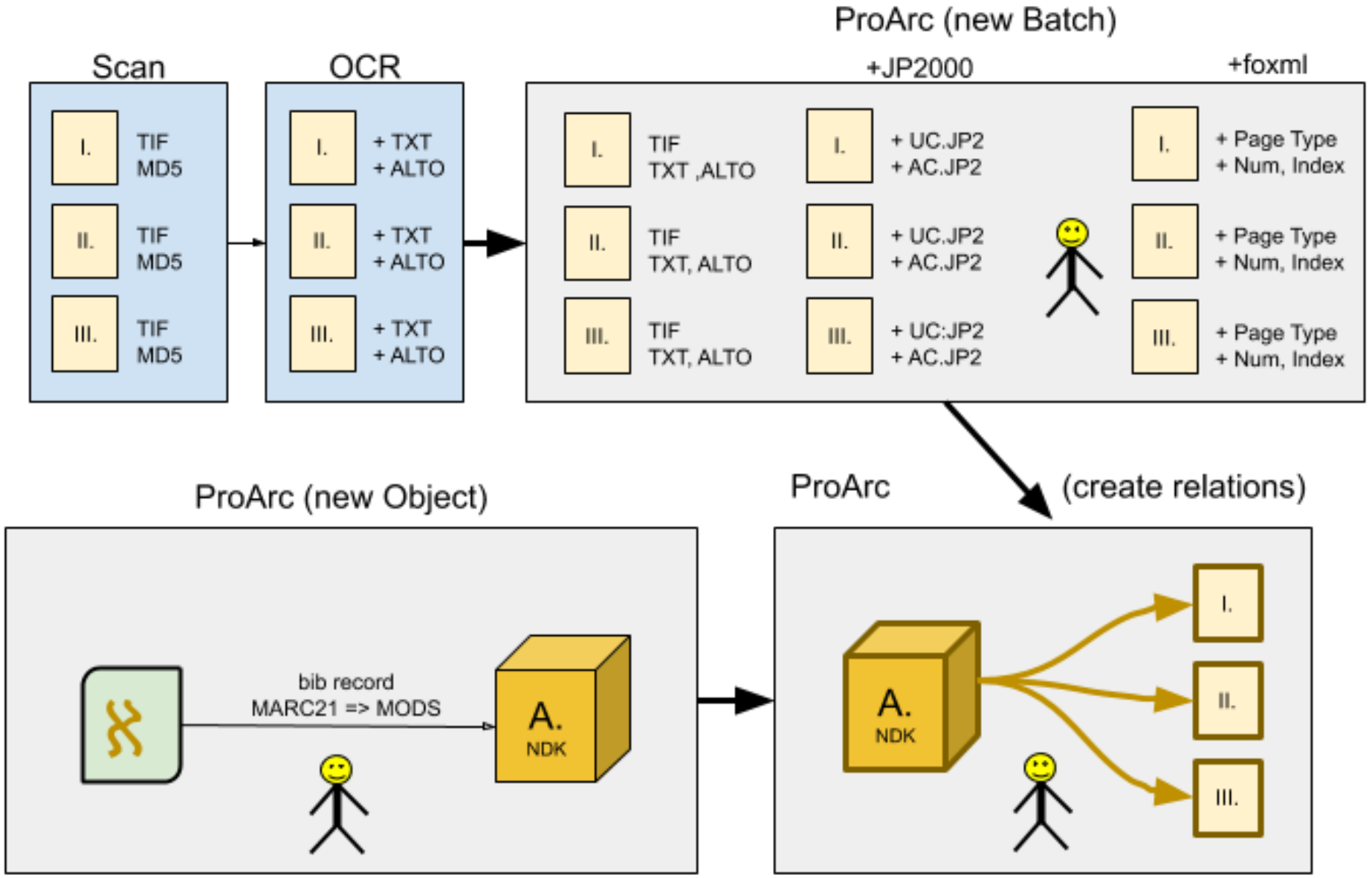

Skener ÚKUK
Jiný skener / zdroj dat
Konverze (Digitool)
obrázky
obrázky
digiftp.cuni.cz
OCR
FTP server
obrázky
obrázky (+ OCR)
proarc.cuni.cz
Tvorba digitálních objektů / ProArc
pseudo NDK PSP
NDK PSP / K4 Export
kramerius.cuni.cz
Zpřístupnění digitálních objektů / Kramerius

ProArc

- ProArc – **pro**dukční a **arc**hivační systém
- systém vytvářející data pro Kramerius (a potenciálně další systémy)

- zpracovávají se
  - metadata dokumentu
  - metadata stran

- metadata jsou důležitá pro vyhledávání(!)
  - v našich i cizích systémech


ProArc (new Batch)
Scan
OCR
+JP2000
+foxml
I.
TIF
MD5
II.
TIF
MD5
III.
TIF
MD5
I.
+ TXT
+ ALTO
II.
+ TXT
+ ALTO
III.
+ TXT
+ ALTO
I.
TIF
TXT ,ALTO
II.
TIF
TXT, ALTO
III.
TIF
TXT, ALTO
I.
+ UC.JP2
+ AC.JP2
II.
+ UC.JP2
+ AC.JP2
III.
+ UC:JP2
+ AC.JP2
I.
+ Page Type
+ Num, Index
II.
+ Page Type
+ Num, Index
III.
+ Page Type
+ Num, Index
ProArc (new Object)
bib record
MARC21 => MODS
A.
NDK
ProArc
(create relations)
A.
NDK
I.
II.
III.

Menu Česky proarc ProArc

Editor připojených objektů - NDK Svazek monografie: 75 let Evangelické bohoslovecké fakulty v Praze

Obnovit Editory Nový Smazat Validovat Označit řadu Náhled

Popis	Model
[1a], Přední desky (FrontCover)	Strana
[1b], Přední přídeští (FrontEndSheet)	Strana
[1], Prázdná strana (Blank)	Strana
[2]	Strana
[3], Titulní strana (TitlePage)	Strana
[4], Prázdná strana (Blank)	Strana
5	Strana
6	Strana
7	Strana
8	Strana
9	Strana
10	Strana
11	Strana
12	Strana
13	Strana
14	Strana
15	Strana
16	Strana
17	Strana
18	Strana
19	Strana

Metadata - Strana: [1a], Přední desky (FrontCover)

Další Metadata Uložit

Typ strany :
Přední desky (FrontCover)

Index strany :
1

Číslo strany :
[1a]

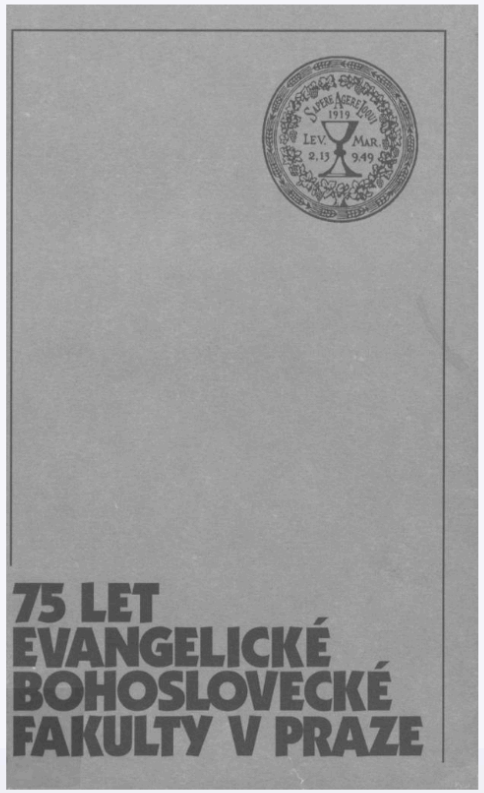
Identifikátor

Typ : UUID **Hodnota :** 05ca50d8-3869-4b1

Poznámka :

Editor digitalizovaných dat - Strana: [1a], Přední desky (FrontCover)

Další Přizpůsobit panelu Náhled



ProArc (2)

- na první pohled „není příliš uživatelský přívětivý“
  - je potřeba zaškolení
  - (plánuje se nové rozhraní)

- s podporovanými typy dokumentů se pracuje relativně dobře

- v tuhle chvíli se používá pro exporty dat do Krameria... v budoucnu bude sloužit jako zdroj dat pro LTP

Kramerius

- nedávno zveřejněno nové rozhraní
  - podpora přepínání mezi dig. knihovnami
    - 3 na UK: centrální Kramerius, 1.LF, FSV
  - moderní design
    - zoomování
    - citace (včetně textu)
    - možnosti výřezů
    - dvoustránkový režim
    - ...

Sbírky Procházet Informace English

Digitalizace Univerzity Karlovy

Digitální knihovna 1.LF UK

Digitální knihovna FSV UK

Hledat v d

pouze veřejné

NAVŠTÍVENÉ NEJNOVĚJŠÍ DOPORUČENÉ

Andr. Siegm. Marggrafs Directors und Chymici

Marggraf, Andreas

1767

Kniha

Češíček : humoristicko-satyrický, illustrovaný

1866

Noviny a časopisy

Acta facultatis medicae Pragensis

[1900-1950?]

Kniha

Tisk, noviny a novináři : věstník

Inaugurace prvního děkana Husovy

Ročenka Husovy československé

kramerius.cuni.cz/uk

Digitalizace Univerzity Karlovy

Hledat v digitální knihovně ...

Sbírky Procházet Informace English

Výsledky: 1 - 60 /**163**

105 0 56 2 0 0 0 0

Řazení: Abecedně

Dostupnost

pouze veřejné	93
pouze neveřejné	70
Vše	**163**

Klíčové slovo

Regionální geografie	29
Náboženství	18
Učební osnovy. Vyučovací předměty. Učebnice	17
Města	13
Vojenská geografie	13
Regionální mapy	9

Zobrazit více

Autor

Husova československá bohoslovecká fakulta	21
Univerzita Karlova Matematicko-fyzikální fakulta	7

50 let Komenského evangelické bohoslovecké

Říčan, Rudolf, 1899-1975

1969

Kniha

75 let Evangelické bohoslovecké fakulty v

Univerzita Karlova

1994

Kniha

Accurater GrundRiss der Stadt und Vestung Ollmütz:

[1758?]

Mapa

Administrativ Karte der Hochfürstlich

Stifter, Johann

[1845]

Mapa

Alternativní školy a inovace ve vzdělávání

Průcha, Jan, 1934-

2012

Kniha

Analytická geometrie v afinních a eukleidovských

Kubát, Václav, 1946-

2011

Kniha

Analýza metod pro maticové výpočty: základní

Duintjer Tebbens, Erik Jurjen

2012

Kniha

Anatomie s topografií a klinickými aplikacemi

Petrovický, Pavel, 1937-

c2002

Kniha

Aplikovaná kartografie I: tematické mapy

Voženílek, Vít, 1965-

2001

Kniha

kramerius.cuni.cz/uk/view/uuid:ea7f0a82-85ae-470f-bbb6-304608f21...

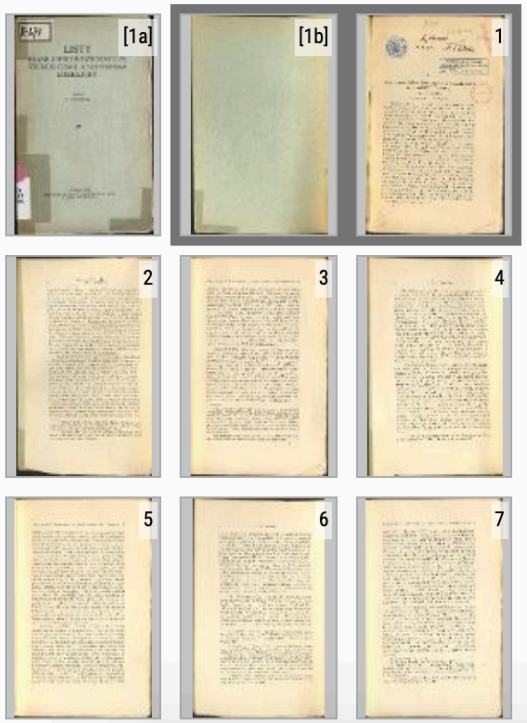
Digitalizace Univerzity Karlovy

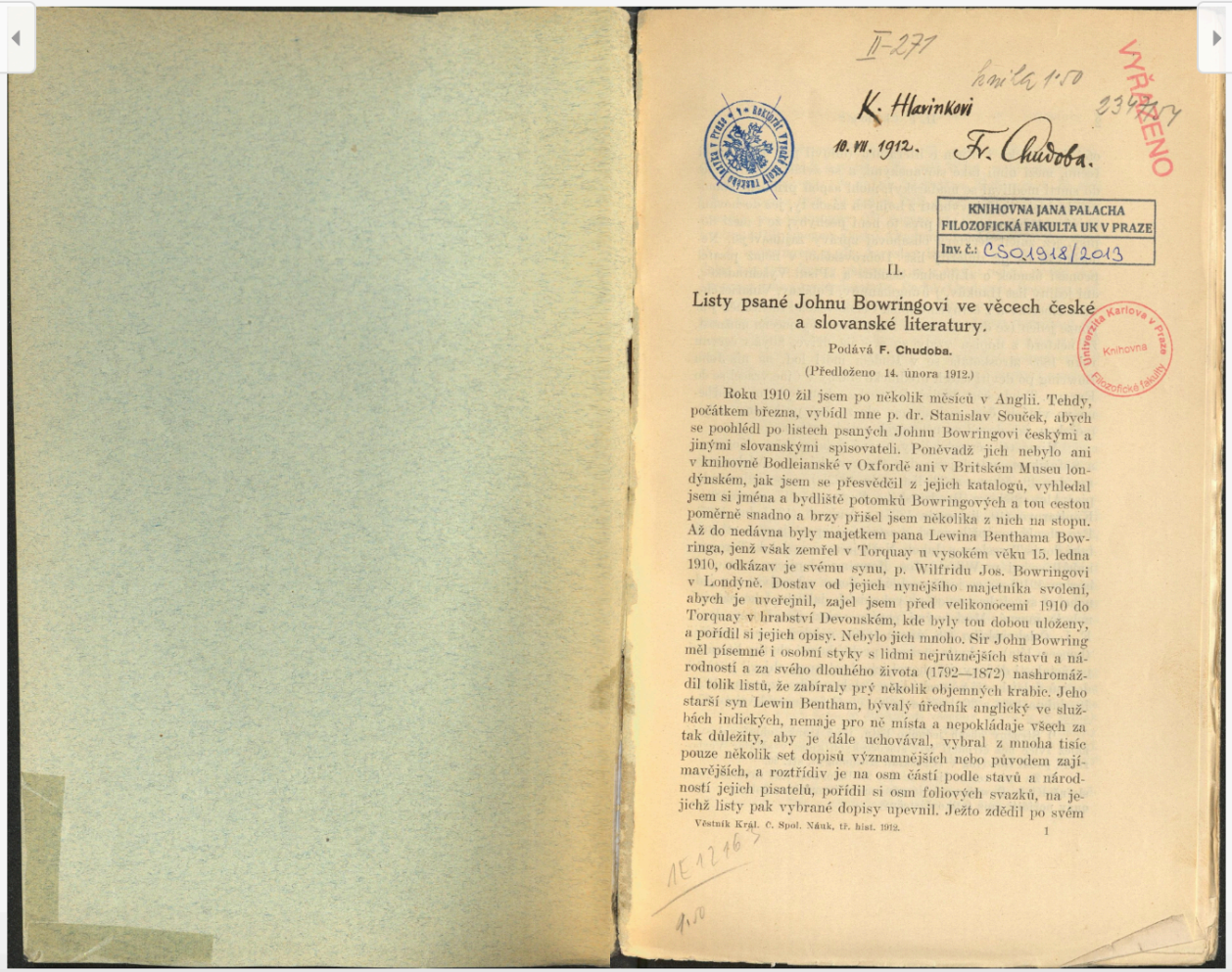
Hledat v digitální knihovně ...

Sbírky Procházet Informace English

Hledat v dokumentu ...

2 z 74 stránek





Listy psané Johnu Bowringovi ve věcech české a slovanské literatury

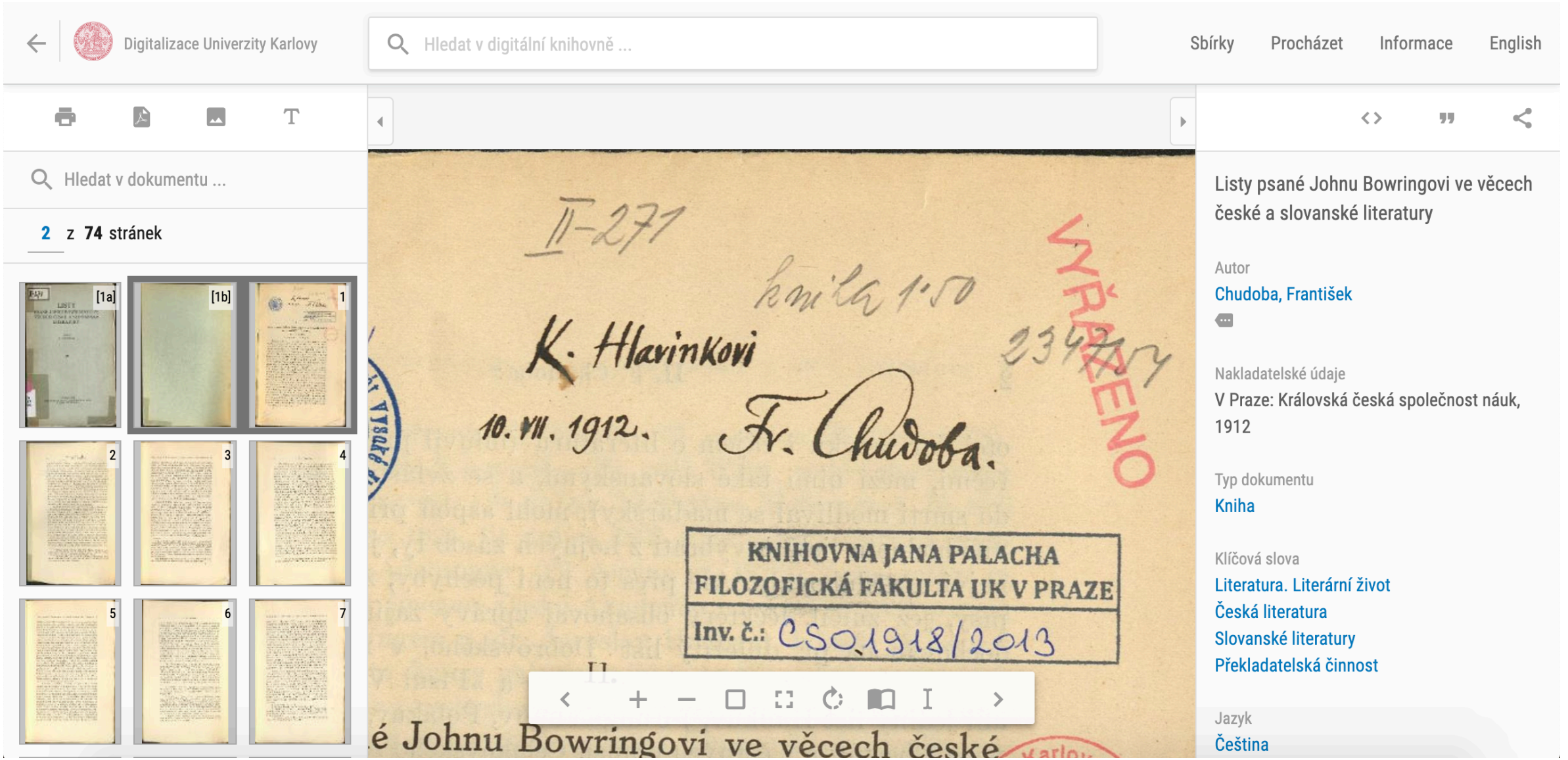
Autor
Chudoba, František

Nakladatelské údaje
V Praze: Královská česká společnost náuk, 1912

Typ dokumentu
Kniha

Klíčová slova
Literatura. Literární život
Česká literatura
Slovanské literatury
Překladatelská činnost

Jazyk
Čeština


Digitalizace Univerzity Karlovy
Hledat v digitální knihovně ...
Sbírky
Procházet
Informace
English
Hledat v dokumentu ...
2 z 74 stránek
II-271
kniha 1·50
K. Hlavinkovi
10. VII. 1912. Fr. Chudoba.
VYŘAZENO
KNIHOVNA JANA PALACHA
FILOZOFICKÁ FAKULTA UK V PRAZE
Inv. č.: CSO.1918/2013
Listy psané Johnu Bowringovi ve věcech české a slovanské literatury
Autor
Chudoba, František
Nakladatelské údaje
V Praze: Královská česká společnost náuk, 1912
Typ dokumentu
Kniha
Klíčová slova
Literatura. Literární život
Česká literatura
Slovanské literatury
Překladatelská činnost
Jazyk
Čeština

Kramerius – kam dál

- má / bude mít kvalitní rozhraní **pro stroje** (API)
  - sdílení (meta)dat
  - efektivní vyhledávání
  - sdílení obrázků (standardy IIIF)
- má potenciál pro tvorbu dalších nástrojů
  - co třeba čtečka integrovaná v jakékoli webové stránce (à la Google Maps)?
  - nástroje pro práci s dokumenty pro Digital Humanities?
- **digitalizací může dokumentu začít nový život ;-)**

Děkujeme za pozornost!

Milan Janíček
milan.janicek (at) ruk.cuni.cz

Eliška De Herdt
eliska.de.herdt (at) ruk.cuni.cz

UNIVERZITA KARLOVA
Ústřední knihovna Univerzity Karlovy