



Digitalizační linka s využitím AI



EXON s.r.o.

Sídlo: Vrážská 73/10, 153 00 Praha

Pobočka Plzeň: Teslova 1242/5c, 301 00, Plzeň

Jan Valenta, Projektový manažer

E: jan.valenta@exon.cz, **M:** +420 602 483 971

- Komplexní zpracování dokumentů se 17 letou tradicí
- Projekty v České republice, Slovensku, Německu, Maďarsku, Polsku
- Partner světových leaderů v oblasti řízení dokumentů a digitalizace



- Vlastní produkty v oblasti zpracování dat a digitalizace



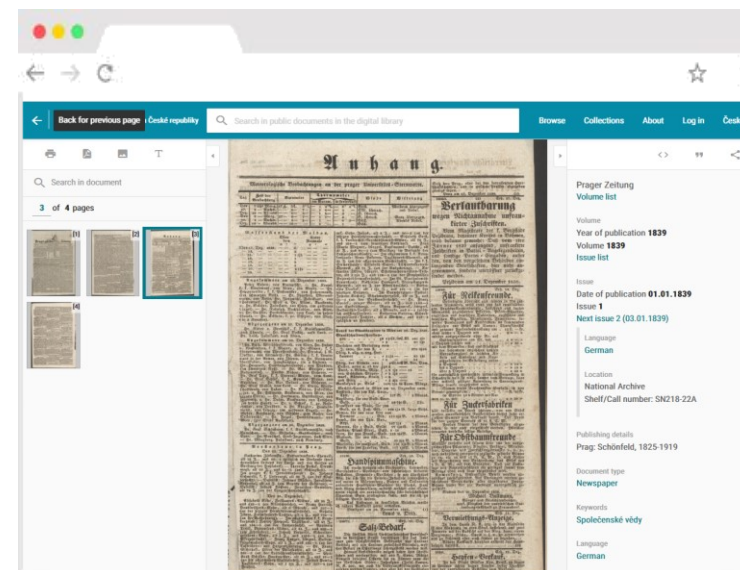
KAITOS



KAITOS – automatizace procesu digitalizace



KAITOS



Typické kroky digitalizačního procesu

skenování

příprava
skenů

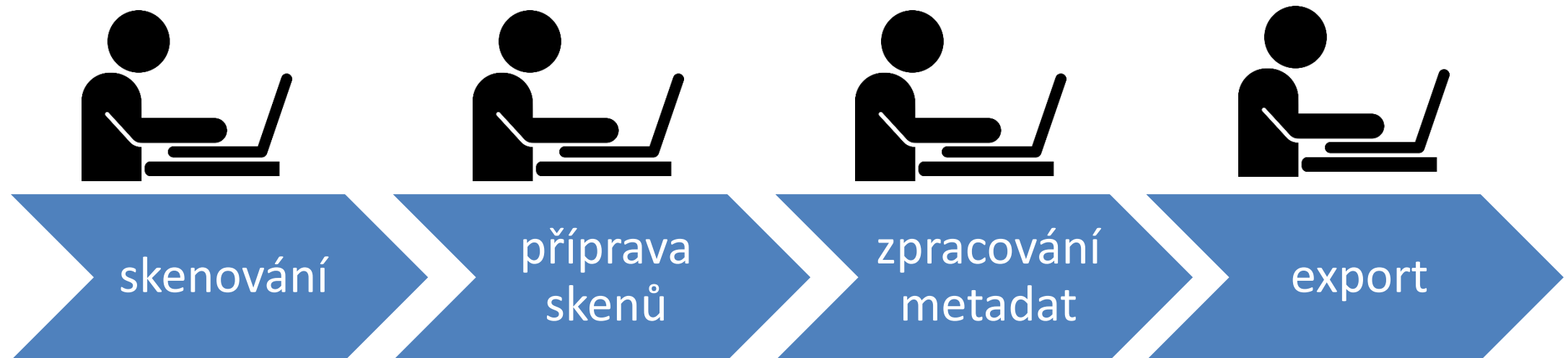
zpracování
metadat

export

Proces digitalizace dnes

Zpracování probíhá **ručně**

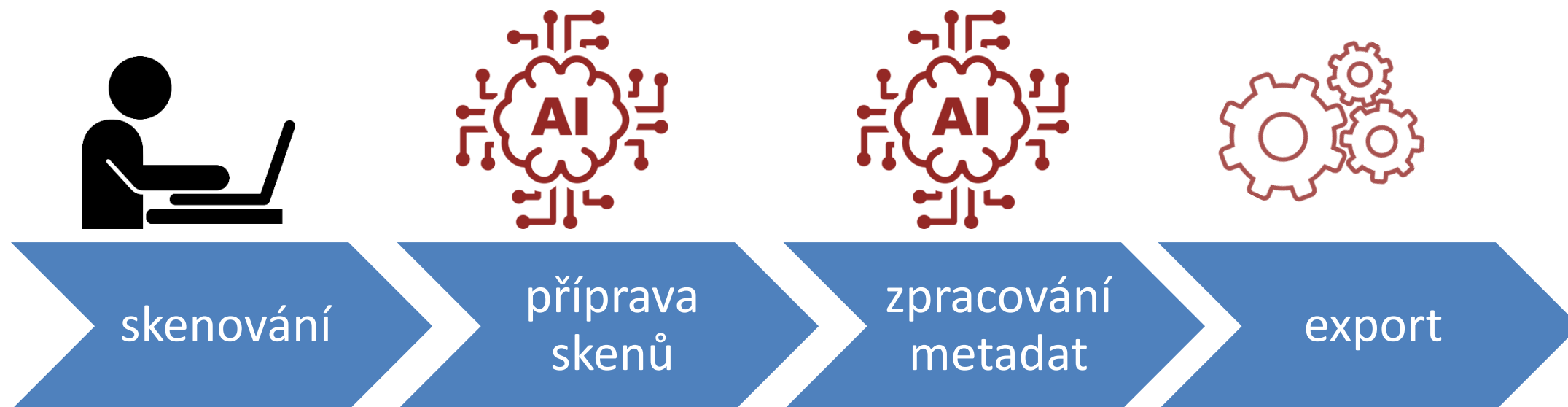
Kvalita a rychlost výstupu silně **závisí na kvalitě a rychlosti operátora**



Proces digitalizace **zítra**

Data z dokumentů jsou získávána **strojově**

Operátor **pouze kontroluje** výsledky a zpracovává výjimky



Příprava skenů s využitím AI

Co dokážeme vyřešit automaticky:

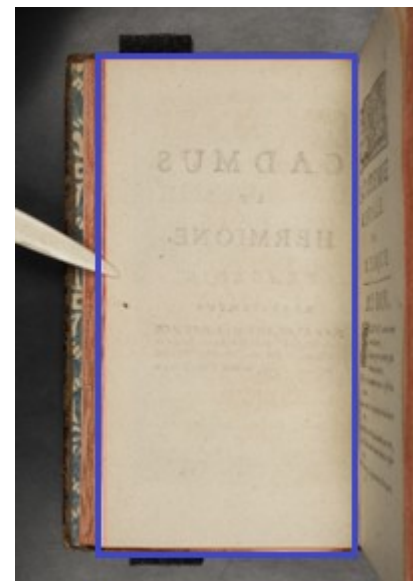
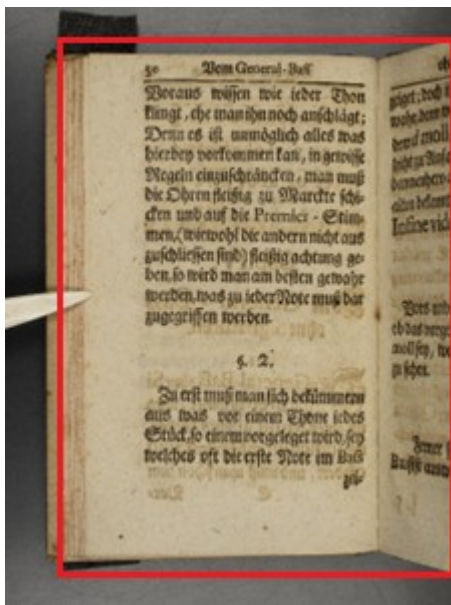
- identifikace jednotlivých stran na skenu
- hrubé zarovnání (rotace o 90/180/270 stupňů)
- jemné zarovnání podle sklonu textu
- nalezení přesné hranice stránky a ořezu stránky (vnitřní/vnější ořez + parametry ořezu)

Co budeme řešit v blízké budoucnosti:

- kontrola kontinuity skenovaných stránek (podle rozpoznaných čísel stran na skenech)
- kontrola ořezů (zařiznutí textu)
- kontrola vizuální kvality skenů – stíny, odlesky, ostrost, oříznutí textu ...



Příprava skenů s využitím AI



- automatické zpracování problematických skenů
- velmi vysoká úspěšnost automatického ořezu



Zpracování metadat s využitím AI

Co dokážeme vyřešit automaticky:

- automatická identifikace čísel stran
- detekce typů stran (normální strana, titulní strana, obsah, reklama...)
- detekce významných bloků textu – titul, ročník, nadpis, tabulka, ilustrace, popisek...
- vytěžení textu (OCR)

Co budeme řešit v blízké budoucnosti:

- rozdělení na jednotlivá vydání a identifikace vnitřních jednotek
- metadata pro vydání a vnitřní jednotky (ročník, nadpis, podnadpis...)
- spojení bloků textu, které k sobě logicky patří – typicky rozdělené novinové články



Zpracování metadat s využitím AI

Příklad klasifikace typů stran a číslování s informací o míře jistoty detekce

Popisná struktura		Vnitřní jednotky	OCR	Proces			
Index	Název souboru	Mods ID	Číslo strany	Type	Stav	[%]	X
1	PS_0019.tif		35	Normální strana	PROCESSED	81	Edit
2	PS_0019.tif		34	Normální strana	PROCESSED	83	Edit
3	PS_0031.tif		58	Normální strana	PROCESSED	81	Edit
4	PS_0031.tif		59	Normální strana	PROCESSED	79	Edit
5	PS_0032.tif		61	Normální strana	PROCESSED	81	Edit
6	PS_0032.tif		60	Normální strana	MANUAL	94	Edit
7	PS_0034.tif		64	Normální strana	PROCESSED	7	Číslování: 81% Typ strany: 91% Ořez: 90%
8	PS_0034.tif		65	Normální strana	PROCESSED	8	
9	PS_0039.tif		75	Normální strana	PROCESSED	9	
10	PS_0039.tif		74	Normální strana	PROCESSED	71	Edit
11	PS_0043.tif		82	Normální strana	PROCESSED	76	Edit
12	PS_0043.tif		83	Normální strana	PROCESSED	85	Edit
13	PS_0047.tif		91	Normální strana	PROCESSED	81	Edit
14	PS_0047.tif		90	Normální strana	PROCESSED	85	Edit
15	PS_0050.tif		96	Normální strana	PROCESSED	76	Edit
16	PS_0050.tif		97	Normální strana	PROCESSED	81	Edit
17	PS_0061.tif		119	Normální strana	PROCESSED	76	Edit
18	PS_0061.tif		118	Normální strana	PROCESSED	82	Edit

skenování

příprava
skenů

zpracování
metadat

export

Zpracování metadat s využitím AI

Příklad identifikace významných oblastí stránky s vyznačením typu obsahu



skenování

příprava
skenů

zpracování
metadat

export

InkCapture – zpracování ručně psaného textu

Hlavní funkce:

- Vytěžení textu z ručně psaného dokumentu
- export textu ve formátu použitelném pro digitální publikaci
- vyhledávání textu v dokumentu - vyhodnocuje se podobnost slov a pravděpodobnost shody, nikoliv pouze přesná shoda
- vyhledávání na základě textu i na základě obrazové podoby slov



Webová aplikace - zaměření na okamžité a snadné využití kýmkoliv

Webová služba – funkce integrovatelná do jakéhokoliv systému třetí strany

InkCapture – vlastní kolekce dokumentů

inkCAPTURE

Prohlížet dokument

Přidat dokument

Velikost miniatur

Navigátor

Moje dokumenty

Historické rodinné dokumenty

Školní poznámky

Česká literatura po roce 1945

Světová poválečná literatura

Zápisky ze schůzek

Spis notářský 1921

Česká literatura po roce 1945

Hromadné úpravy

1. strana

2. strana

3. strana

4. strana

5. strana

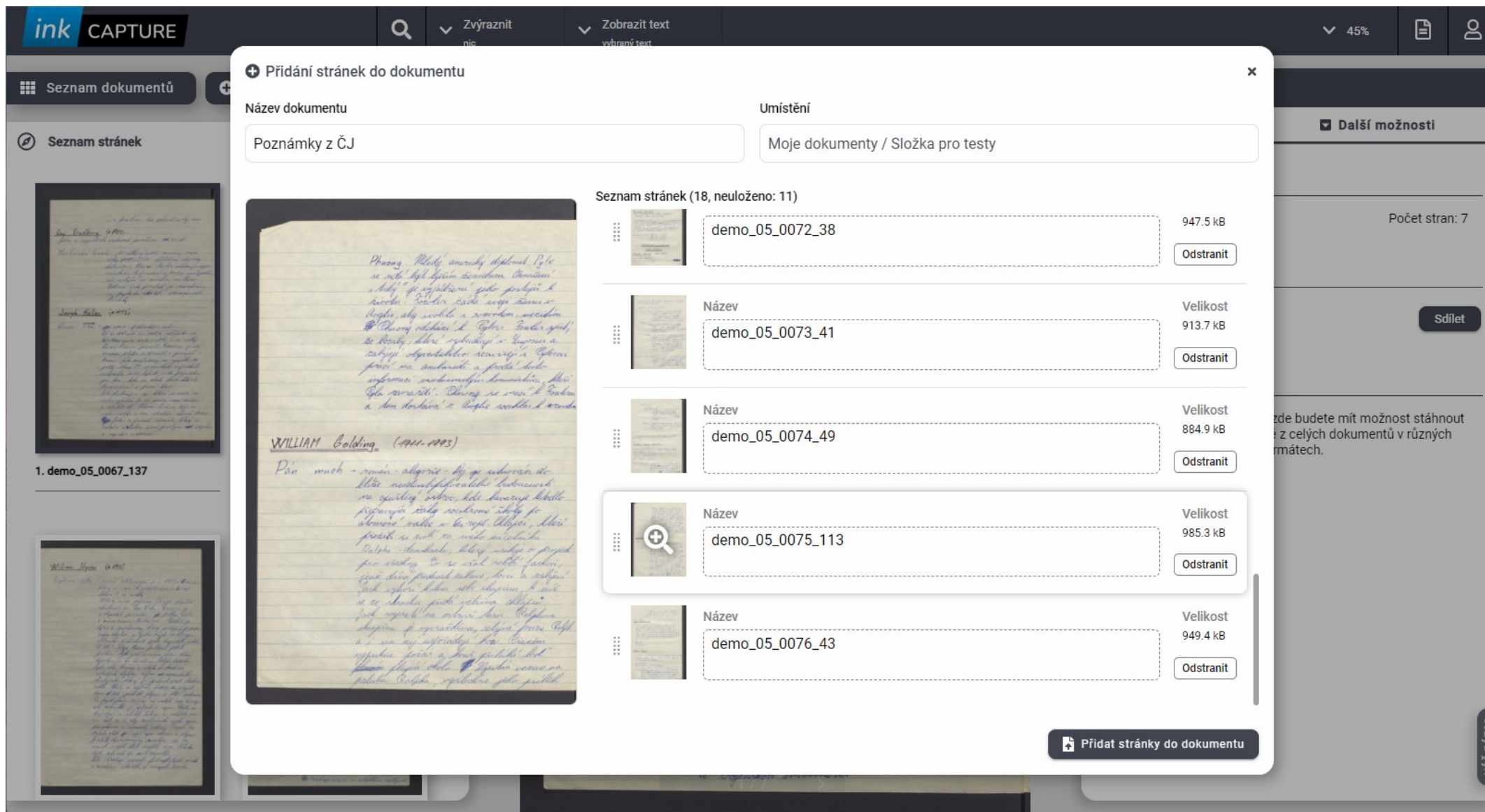
6. strana

7. strana

8. strana

váš názor

InkCapture – vlastní kolekce dokumentů



InkCapture – práce s textem dokumentu

inkCAPTURE

Zvýraznit
nic

Zobrazit text
vybraný text

50%

Seznam dokumentů

Přidat stránku

Seznam stránek

1. strana
Nahráno: 23. 5. 2022

2. strana
Nahráno: 23. 5. 2022

3. strana
Nahráno: 23. 5. 2022

4. strana
Nahráno: 23. 5. 2022

číslo 103.
b.) na poli číslo katastru 536/2 u Košíka ve vložce číslo 502.
c.) na poli číslo katastru 361/1 u Košíka ve vložce číslo 511.
d.) na domku číslo popisné 74 v Košíku ve vložce číslo 582.

O tom sepsal jsem tento spis notářský, přečetl jsem jej oběma účastníkům, ti schválili celý obsah a po té spis tento přede mnou vlastnoručně podepsali.
Josef Tesař _____ Paula Tesařová _____
L. J. Jan Komers, _____
notář.

Snímek tento určený pro paní Paulu Tesařovou, manželku rolníka v Košíku číslo popisné 33, srovnává se doslovně se spisem notářským, který jest v mých spisech pod číslem jednacím 5.819 založen, na

3. strana

TextExport

Strana třetí.
číslo 103,
b.) na poli číslo katastru 536/2 u Košíka ve vložce číslo 502,
c.) na poli číslo katastru 361/1 u Košíka ve vložce číslo 511,
d.) na domku číslo popisné 74 v Košíku ve vložce číslo 582
O tom sepsal jsem tento spis notářský, přečetl jsem jej oběma účastníkům, ti schválili celý obsah, a po té spis tento přede mnou vlastnoručně podepsali.
Josef Tesař Paula Tesařová
L. S. Jan Komers,
notář.
Snímek tento určený pro paní Paulu Tesařovou, manželku rolníka v Košíku číslo popisné 33, srovnává se doslovně se spisem notářským, který jest v mých spisech pod číslem jednacím 5.819 založen, na jednom archu napsán a kolkem za dvě koruny - 2 K Opatřen. V Libáni, dne čtvrtého června roku

přidat do poznámek (Ctrl+Alt+C)

váš názor

InkCapture – sdílení dokumentů

inkCAPTURE

Zvýraznit

nic

Zobrazit text

vybraný text

50%

Seznam dokumentů

Přidat stránku

Seznam stránek

1. Strana 1

2. Strana 2

3. Strana 3

4. Strana 4

3. Strana 3

Text

Další možnosti

Dokument

Testovací dokument

Počet stran: 4

Sdílení

Dokument je sdílený

Změnit

Export

V dalších verzích aplikace zde budete mít možnost stáhnout text ze stránek nebo také z celých dokumentů v různých formátech.

váš názor

Sdílení dokumentu

Dokument: Testovací dokument

Sdílet dokument s: kýmkoliv

Každý, kdo bude mít odkaz na tento dokument, bude moci dokument zobrazit. Dokument budou moci s odkazem vidět i lidé, kteří nemají založený účet v aplikaci InkCapture.

Odkaz pro sdílení: <http://localhost:8100/browse-doc>

Zkopírovat

Nastavení úspěšně uloženo

Nové produkty EXON



KAITOS

www.kaitos.eu

- Příprava skenů
 - ořez a rotace
 - kontroly kvality
- Zpracování metadat
 - číslování stran
 - detekce typů stran
 - vytěžení tištěného textu
 - identifikace významných bloků textu
 - logické spojení rozdělených částí textu



www.inkcapture.com

- Vytěžení ručně psaného textu
- Vyhledávání ručně psaného textu
- Prezentace dokumentů na internetu



EXON

Intelligent Document Management

CULTURAL HERITAGE DIGITIZATION

BEDIGITAL

INTELLIGENT DOCUMENT MANAGEMENT

WTHUS

Děkuji Vám za pozornost

EXON s.r.o.

Sídlo: Vrážská 73/10, 153 00 Praha

Pobočka Plzeň: Teslova 1242/5c, 301 00, Plzeň

Jan Valenta, Projektový manažer

E: jan.valenta@exon.cz, **M:** +420 602 483 971