

Organizace digitálního zpracování dokumentů

Studijní text

Autor: Mgr. Zdeněk Hruška

Konzultant: Bc. Martina Smetánková

Ostrava 2019



VISK VEŘEJNÉ
INFORMAČNÍ SLUŽBY
KNIHOVEN



Odborná způsobilost	Organizace digitálního zpracování dokumentů
Profesní kvalifikace	Samostatný knihovník správce digitální knihovny (kód: 72-018-R)
Kvalifikační úroveň	6
Klíčová slova	Digitalizace, skenery, metadata, postprocessing obrazů, dlouhodobá archivace, digitální knihovna, OCR, zpřístupnění, čČNB, ISSN, Registr digitalizace, standardy digitalizace
Anotace	Tento studijní text je určen pro zájemce o problematiku digitalizace, digitálních knihoven a souvisejících procesů. Provede čtenáře jednotlivými kroky digitálního zpracování dokumentů od skenování, přes zveřejnění v digitální knihovně až po uložení v dlouhodobém repozitáři. Také popisuje různé typy skenerů a představuje národní databáze jako je Registr digitalizace, Česká národní bibliografie a báze ISSN.

Tvorba studijních textů je realizována za finanční podpory Ministerstva kultury České republiky v rámci projektu Veřejné informační služby knihoven (VISK 1).

Zkouška z odborné způsobilosti **Organizace digitálního zpracování dokumentů** má 5 částí, zkoušený musí splnit tato kritéria hodnocení:

Kritéria hodnocení		Způsob ověření
a	Charakterizovat jednotlivé fáze procesu digitalizace	Ústní ověření
b	Ověřit dokument v centrálních registrech digitalizovaných dokumentů	Praktické předvedení
c	Vyhledat číslo ISSN (International Standard Serial Number – Mezinárodní standardní číslo seriálových publikací) ke konkrétnímu českému pokračujícímu dílu	Praktické předvedení
d	Vyhledat a přiřadit k záznamu konkrétní monografie nebo pokračujícího díla číslo České národní bibliografie	Praktické předvedení
e	Popsat typy skenerů využívaných při digitalizaci	Ústní ověření

Tento text pro vás připravil Mgr. Zdeněk Hruška, Moravská zemská knihovna v Brně, zdenek.hruska@mzk.cz

Text konzultovala Bc. Martina Smetánková, Moravská zemská knihovna v Brně, martina.smetankova@mzk.cz

OBSAH

RYCHLÝ NÁHLED STUDIJNÍHO TEXTU.....	5
1 PROCES DIGITALIZACE	5
1.1 Příprava dokumentu před skenováním	5
1.2 Skenování.....	6
1.3 Úprava obrazových dat.....	6
1.4 OCR – rozpoznávání znaků.....	7
1.5 Metadata.....	9
1.5.1 Dělení metadat dle jejich účelu	9
1.5.1.1 Popisná metadata	10
1.5.1.2 Strukturální metadata.....	11
1.5.1.3 Technická metadata.....	11
1.5.1.4 Administrativní a archivační metadata	12
1.5.1.5 Metadatové standardy Národní Digitální Knihovny (NDK)	12
1.5.1.6 Formátový výbor NDK.....	13
1.6 Zpřístupnění.....	13
1.6.1 Zpřístupnění dokumentů v digitální knihovně (autorský zákon).....	14
1.7 Dlouhodobá ochrana digitálních dokumentů.....	15
2 CENTRÁLNÍ REGISTR DIGITALIZOVANÝCH DOKUMENTŮ.....	17
2.1 Číslo České národní bibliografie (čČNB)	18
2.2 Rozpoznání duplicity.....	18
3 VYHLEDÁNÍ ČÍSLA ČESKÉ NÁRODNÍ BIBLIOGRAFII	21
4 ČÍSLO ISSN U POKRAČUJÍCÍCH DĚL.....	22
5 TYPY SKENERŮ.....	23
5.1 Stolní skenery	23
5.2 Planetární skenery.....	24
5.3 Ostatní skenery.....	25
6 SHRUTÍ STUDIJNÍHO TEXTU.....	27
7 POUŽITÁ LITERATURA.....	28

RYCHLÝ NÁHLED STUDIJNÍHO TEXTU

ÚVOD

Proces digitalizace se skládá z jednotlivých dílčích kroků, z nichž každý má svůj význam. Text, který právě čtete, tyto kroky popíše a vysvětlí jejich důležitost. I když správce knihovny nepracuje přímo se skenery nebo s metadatovým editorem, je nutné, aby měl alespoň základní přehled o tom, které činnosti předchází nahrání dokumentu do digitální knihovny – ať už se jedná o vyhledání dokumentu v některé z centrálních databází, nebo o ořez obrazů a editaci metadat. Znalost jednotlivých kroků vám pomůže lépe poznat celý digitalizační proces a také se snadněji dorozumět s kolegy, kteří mají na starosti jeho konkrétní části.

PO PROSTUDOVÁNÍ STUDIJNÍHO TEXTU BUDETE UMĚT

- popsat jednotlivé části digitalizačního procesu a jejich význam,
- ověřit dokument v centrálním registru digitalizace,
- vyhledat číslo České národní bibliografie k příslušnému dokumentu,
- vyhledat ISSN pro pokračující dílo,
- rozlišit různé typy skenerů, které se k digitalizaci v knihovnách využívají.

KLÍČOVÁ SLOVA STUDIJNÍ OPORY

Digitalizace, skenování, postprocessing obrazových dat, editace metadat, OCR, zpřístupnění, digitální knihovna, dlouhodobá archivace, čČNB, ISSN, skenery, autorský zákon, Národní digitální knihovna, standardy digitalizace

1 PROCES DIGITALIZACE

Digitalizací se v knihovnách rozumí převod analogového dokumentu do digitální podoby za pomoci odpovídajícího zařízení (v tomto případě skeneru). Digitalizace v knihovnách pokrývá několik různých potřeb:

- **Snazší přístup uživatelů k dokumentu** – např. v podobě digitalizace skript nebo jiné žádané literatury, dokumentů uložených v externím skladu (odpadá nutnost dopravy) nebo rozměrných svazků (slovníky, encyklopedie),
- **Ochranná digitalizace** – týká se dokumentů, kterým hrozí poškození nebo dokonce krádež ze strany uživatelů (např. stará periodika tištěná na kyselém papíře, staré tisky a rukopisy a další vzácné svazky).

Oba tyto důvody se mohou samozřejmě překrývat a doplňovat. Obecně je primárním účelem digitalizace v knihovnách zpřístupnit dokumenty uživatelům, ale vytvoření digitální kopie umožňuje uložit originál do odpovídajících podmínek, již ho nezapůjčovat a tím ho lépe chránit.

Celý proces digitalizace se skládá z jednotlivých kroků, které si níže podrobněji popíšeme a charakterizujeme:

- **příprava před skenováním,**
- **skenování,**
- **úprava obrazových dat,**
- **OCR – rozpoznání znaků,**
- **tvorba metadat,**
- **zpřístupnění,**
- **archivace dat.**

1.1 Příprava dokumentu před skenováním

Před samotným skenováním je nutné provést kontrolu stavu dokumentu, a to především s ohledem na:

- **fyzický stav** – dokument se nerozpadá, vazba by měla vydržet namáhání při skenování (je dostatečně pružná, nehrozí její zlomení),
- **úplnost** – nechybí žádné strany, obálka či přílohy,
- **čistota** – dokument by neměl být zaprášený, plesnivý, se skvrnami od nápojů apod.

Špinavý dokument je potřeba vhodně očistit, protože prach a další nečistoty mohou zanést skener nebo znehodnotit skeny (šmouhy, drobná smítka apod.).

Pokud je dokument ve špatném fyzickém stavu, případně neúplný, je nutné vyhledat ve fondu jiný exemplář a k digitalizaci vybrat ten nejvhodnější. Jestliže není žádný v lepším stavu k dispozici, musíme při skenování postupovat se zvýšenou opatrností, aby nedošlo k dalšímu poškození. Když chybí některá část dokumentu, je možné požádat některou z jiných knihoven o její naskenování a zaslání.

1.2 Skenování

V českém knihovnickém prostředí je používání standardů v digitalizaci na vysoké úrovni, a to především díky projektu **Národní digitální knihovna**, kde vznikly **standardy pro digitalizaci** (o nich podrobněji viz kapitola Metadata),¹ a také díky dotačnímu programu VISK (konkrétně VISK 7)², který dodržování těchto standardů podmiňuje přidělením finančních prostředků.

Dle standardu NDK se monografie a periodika skenují minimálně na 300 DPI (Dots Per Inch – tedy bodů na palec čtvereční), ve 24bitové barevné hloubce a ukládají se ve formátu JPEG 2000. Skenery běžně skenují do formátu TIFF (ale též do jiných – např. JPG, PNG nebo jsou schopny vytvářet PDF soubory), takže dle standardu NDK probíhá konverze z TIFF do JPEG 2000.

Mapy a další rozměrné či detailní předlohy se doporučují skenovat na vyšší rozlišení 600 DPI.

Dle typu skenovaného dokumentu a dostupných zařízení je dokument přiřazen na nejvhodnější stroj, kde dojde k jeho naskenování. Nastavení dokumentu tak, aby proběhlo skenování v rovině textu, usnadní práci v další fázi digitalizace (úprava obrazových dat).

1.3 Úprava obrazových dat

Úprava obrazových dat, někdy též označována jako postprocessing, je série úkonů, která následuje po skenování a má za cíl **upravit sken do výsledné podoby pro uživatele**. Jedná se především o **ořez obrazu** a **natočení v rovině textu** (ne vždy je text správně vytištěn, případně může dojít ke špatnému natočení strany během skenování).

U novodobé produkce (cca po roce 1800) jde o tzv. **ořez vnitřní** (dovnitř strany, kdy se vynechá několik milimetrů), u starých tisků, rukopisů, map a grafik se pak

¹ Standardy digitalizace. Dostupné z: <https://www.ndk.cz/standardy-digitalizace>

² VISK 7. Dostupné z: <https://visk.nkp.cz/visk-7>

provádí **ořez vnější** (kvůli zachycení stavu celého dokumentu, tedy i okrajů – což je důležité např. pro badatele).

S ohledem na typ skeneru a charakter dokumentu může docházet v této činnosti také k rozdělení dvoustran na jednostrany (pokud tato činnost neproběhla automaticky během skenování).

Pokud je nutné, může docházet k dalším úpravám skenů:

- **Změna barevnosti** – i když standard NDK jasně hovoří u monografií a periodik o 24bitové barevné hloubce, někdy je pro uživatele vhodnější, pokud dostane skeny v jiné úpravě. Např. v odstínech šedi nebo černobílé – to může odstranit některé problémy, jako je např. prosvítající text z druhé strany, případně barevné skvrny na papíře. Takové obrazy jsou pak také vhodnější k tisku.
- Různé softwarové nástroje určené pro úpravu skenů nabízejí i další možnosti – umí **odstranit průsvit, šum** a další nedostatky, kterým se nedalo během skenování předejít.

1.4 OCR – rozpoznávání znaků

Optickým rozpoznáváním znaků (Optical Character Recognition) se rozumí **převod textu ze skenu** (v podobě obrázku) **do textového formátu** (typicky TXT, PDF nebo specializovaný formát ALTO³). Zatímco v obrázku nejde fulltextově vyhledávat (počítač nerozezná, co je písmeno a co jiná část obrázku – v podstatě jsou to všechno pixely), v textovém souboru lze najít zadaný text. To je klíčové pro uživatele, kteří mohou prohledávat fulltextově celou digitální knihovnu při pátrání po konkrétních slovech či slovních spojeních.⁴ OCR je tedy přidanou hodnotou, kterou digitalizace pro uživatele nabízí, a mělo by být vždy přítomno v rámci digitalizace (pokud je to možné a máme k tomu vše potřebné – o limitech OCR viz níže).

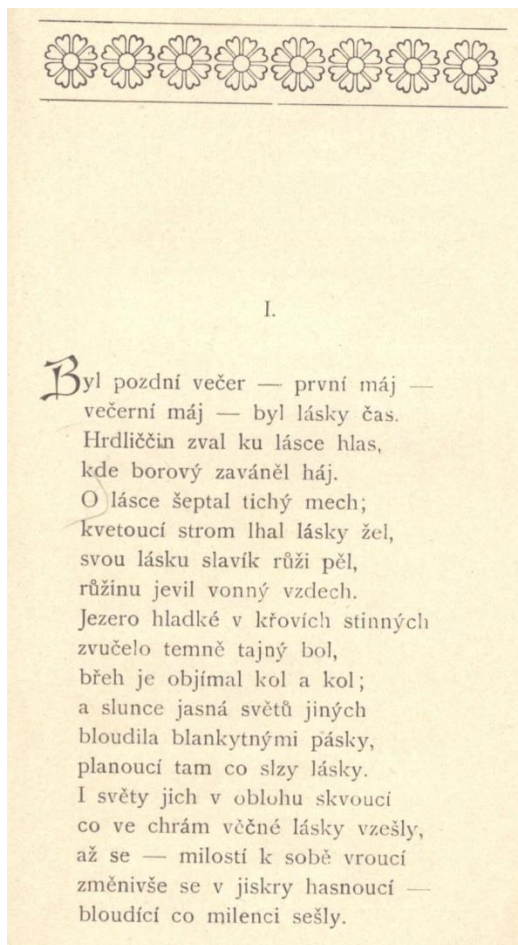
U moderních dokumentů je úspěšnost rozpoznání znaků vysoká a pohybuje se v řádu 98–100 % (tedy použitý software rozpozná téměř všechny znaky na stránce), u starší produkce pak klesá (noviny z přelomu 19. a 20. století mají úspěšnost 80 % a méně). Schopnost rozpoznání znaků také snižuje průsvitný či zažloutlý papír, nekvalitní tisk, deformovaný text (např. zahnutý u středu stran) apod.

U rukopisů bylo donedávna nemožné OCR provést, protože software nebyl schopný rozpoznat psané písmo. S nastupující umělou inteligencí bude možné

³ ALTO: Technical Metadata for Layout and Text Objects. Dostupné z: <https://www.loc.gov/standards/alto/>

⁴ Viz např. specifické pátrání: BROŽEK, Aleš. Najdeme informace o nevěstincích na Vysočině v systému Kramerius? AKM 2017. Dostupné z: <https://bulletin.skipcr.cz/prezentace/archivy-2017/Brozek.pdf>

rozpoznávat znaky i u těchto dokumentů (viz např. PERO, společný projekt MZK a VUT).⁵



I. Byl pozdní večer —
první máj — večerní máj —
byl lásky čas. Hrdliččin zval
ku lásce hlas, kde borový
zaváněl háj. O lásce šeptal
tichý mech; kvetoucí strom
lhal lásky žel, svou lásku
slavík růži pěl, růžinu jevil
vonný vzdech. Jezero hladké
v křovích stinných zvučelo
temně tajný bol, břeh je
objímal kol a kol ; a slunce
jasná světů jiných bloudila
blankytnými pásy, planoucí
tam co slzy lásky. I světy jich
v oblohu skvoucí co ve
chrám věčné lásky vzešly, až
se — milostí k sobě vroucí
změnivše se v jiskry
hasnoucí bloudící co milenci
sešly.

vlevo sken jedné strany, vpravo textový přepis za pomoci OCR

⁵ Projekt PERO [online]. Dostupné z: <https://pero.fit.vutbr.cz/>

1.5 Metadata

Metadata obecně jsou **data, které popisují jiná data** a tím nám pomáhají v jejich **identifikaci a zařazení, podporují vyhledávání** a další aspekty **uspořádání a uchovávání** informací. V oblasti knihovnictví metadata popisují knihovní dokumenty (ať už digitální nebo analogové) a jejich vlastnosti.

V případě digitalizace a digitálních knihoven jsou metadata zapsána v podobě **XML**, což je **obecný značkovací jazyk**, který jednoduše umožňuje tvorbu konkrétních dalších značkovacích jazyků pro nejrůznější použití.⁶ To umožňuje pomocí XML vytvářet další metadatové struktury a zapisovat je tak, aby byly čitelné jak lidským okem, tak i strojově.

Obecný postup během digitalizace je, že jsou načtena popisná metadata z knihovního katalogu (za pomoci jednoznačného identifikátoru dokumentu, jako je čárový kód, sysno apod.) do metadatového editoru. Zároveň se nahrají skeny a v dalších krocích jsou k jednotlivým stranám přiřazeny údaje, které pomáhají uživateli digitální knihovny v orientaci v dokumentu. Zpravidla se jedná o čísla a typ stran (obálka, obsah, rejstřík, tabulky, mapy, ilustrace a další). U periodik ještě dochází k dělení na jednotlivá čísla a ročníky – periodika se běžně vážou do vazby po celých rocích (měsíčníky), ale i po několika měsících (noviny vycházející každý den). Může tedy dojít k tomu, že je naskenován celý svazek periodik, což může být i 1 000 skenů. V takovém množství by se uživatel jen těžko orientoval (nehledě na to, že ve vazbě mohou být jednotlivá čísla přeházená), takže je nutné přiřadit skeny ke správným číslům a dodat informaci, kdy číslo vyšlo. Důležité je také jednotlivá čísla napojit logickou strukturou k příslušným ročníkům a ty k mateřskému dokumentu (periodiku), aby vše bylo v digitální knihovně vyhledatelné a jednoduše zobrazitelné.

1.5.1 DĚLENÍ METADAT DLE JEJICH ÚČELU

Protože metadata pokrývají řadu různých potřeb v rámci digitalizace i dlouhodobé ochrany digitálních dokumentů, dělí se na několik typů, které si zde podrobněji popíšeme. Metadatové formáty, které jsou níže zmíněny, jsou považovány za mezinárodní standardy v oblasti digitalizace a digitální archivace a jsou využívány i v projektu Národní digitální knihovny a dalších tuzemských digitalizačních projektech.

⁶ Extensible Markup Language. Dostupné z: https://cs.wikipedia.org/wiki/Extensible_Markup_Language

1.5.1.1 Popisná metadata

Jsou to ty údaje, které **pomáhají popsat, identifikovat a snadněji vyhledávat dokument** – jedná se o název, autora, rok a místo vydání, stejně tak sem patří i další identifikátory, jako je ISBN, ISSN, signatura, čárový kód, čČNB a další. Při digitalizaci se tato metadata přejímají z knihovního katalogu, výjimkou je identifikátor dle standardu URN:NBN. Ten je přidělován během digitalizace systémem ČIDLO (Český systém pro identifikaci a lokalizaci dokumentů digitálního kulturního dědictví), který provozuje Národní knihovna ČR.⁷ URN:NBN slouží pro identifikaci digitálních dokumentů a také podporuje jejich zpřístupnění a dlouhodobé uchovávání.

Dalším identifikátorem, který se přiřazuje během digitalizace, je UUID.⁸ Tento alfanumerický řetězec se používá v rámci digitálních repozitářů k označení jednotlivých objektů, protože se jedná o jedinečný identifikátor (jedinečnost se dána délkou řetězce i způsobem generování dle algoritmu – teoreticky lze najít dvě stejná UUID čísla, ale pravděpodobnost je velmi nízká), takže nachází využití všude tam, kde je potřeba identifikovat velké množství jednotek (řádově miliony až desítky milionů). V digitální knihovně Kramerius má své vlastní UUID každá digitalizovaná stránka, stejně jako nadřazený dokument (monografie, ale i každé číslo, ročník a titul periodika).

41a78cb0-e34c-11e9-84ef-005056827e51

Příklad UUID – jedná se o řetězec 32 znaků, kombinace písmen a čísel

Dublin Core (DC)

Standard vznikl původně v 90. letech pro popis elektronických zdrojů na webu. Postupně se rozšířil i do dalších oblastí, jako je např. digitalizace. Umožňuje jednoduchý popis pomocí patnácti základních metadataových elementů, z nichž žádný není povinný, a které se dají dále rozšiřovat.⁹

MODS (Metadata Object Description Schema)

Jedná se o metadataový standard zapsaný v XML a spravovaný Library of Congress. Byl vytvořen pro potřeby knihoven, kdy nabízí možnost bohatšího popisu než Dublin Core (ten je v řadě případů příliš obecný a nedostačující pro potřeby knihovní katalogizace). Obsahuje 20 základních prvků, které lze doplnit podprvky (všechny jsou nepovinné a opakovatelné, ale alespoň jeden element musí být vyplněn) a atributy (ty jsou neopakovatelné).

⁷ CUBR, Ladislav a Zdeněk VAŠEK. Identifikátory digitálních dokumentů se zaměřením na systém URN:NBN v ČR. Čtenář: Měsíčník pro knihovny [online]. Dostupné z: <https://svkkl.cz/en/ctenar/clanek/1254>

⁸ Universally unique identifier. Dostupné z: https://en.wikipedia.org/wiki/Universally_unique_identifier

⁹ Dublin Core. Dostupné z: http://webserver.ics.muni.cz/dublin_core/index.html

1.5.1.2 Strukturální metadata

Strukturální metadata **popisují strukturu intelektuálních nebo fyzických objektů**. Digitální objekty se skládají z různých částí – například skeny v archivní a uživatelské kopii, OCR v TXT a ALTO a další metadata. Je tedy nutné zapsat, kde se v balíčku jednotlivé části nacházejí (fyzické mapování) a jaké mají k sobě vazby (intelektuální mapování). U monografie to bude řazení a číslování stránek či kapitol, u periodik pak i rozdělení čísel a ročníků. Strukturální popis umožňuje správné zobrazení dokumentu při jeho prohlížení uživatelem a udržuje i složité objekty v logickém uspořádání.

METS (Metadata Encoding and Transmission Standard)¹⁰

METS je metadatový standard pro zápis popisných, administrativních a strukturálních metadat. Jedná se o kontejner, do něhož můžou být vloženy jiné metadatové formáty (DC, MODS, MIX, aj.). Je zapsán pomocí XML a o jeho správu se stará Library of Congress.

FOXML (Fedora Object XML)

Jedná se o interní formát repozitáře Fedora a je tedy součástí digitální knihovny Kramerius. Ve FOXML jsou zapsána ostatní metadata, jedná se tedy také o kontejnerový formát, podobně jako METS. Existuje řada převodníků, které umožňují import a export metadat do jiných systémů.¹¹

1.5.1.3 Technická metadata

Popisují **technické charakteristiky digitálních dokumentů**. Může sem spadat záznam o použitém skeneru (např. typ, výrobce, výrobní číslo) a softwaru (název, verze), nebo popis formátu souborů (verze, rozlišení a barevnost u skenů, bit-rate záznamu u zvukových dokumentů apod.).

MIX (Metadata for Images in XML Schema)¹²

Slouží pro zápis technických metadat obrazových souborů. Může obsahovat informace o formátu obrazu a jeho verzi, velikosti, rozlišení, barevném profilu, stejně jako údaje o skeneru.

ALTO (Analyzed Layout and Text Object)

Jedná se o metadatový standard zapsaný v XML, který spravuje Library of Congress. Umožňuje záznam rozvržení textu získaného díky OCR a jeho mapování na skeny stránek při zobrazení v digitální knihovně.

¹⁰ METS. Dostupné z: <http://www.loc.gov/standards/mets/>

¹¹ FOXML. Dostupné z: <https://wiki.lyrasis.org/pages/viewpage.action?pageId=66585857>

¹² MIX. Dostupné z: <http://www.loc.gov/standards/mix/>

1.5.1.4 Administrativní a archivační metadata

Popisují **okolnosti vzniku digitálních dokumentů**, které entity se na vzniku podílely (např. skenerista nebo software pro úpravu obrazů apod.) a k jakým akcím s daty došlo (např. převod z jednoho formátu na jiný).

PREMIS (Preservation Metadata Implementation Strategies)

Metadatový formát spravovaný Library of Congress slouží k podpoře dlouhodobého uchování digitálních dokumentů. Umožňuje popsat prostředí vzniku (hardware, software a zapojené osoby), práva (autorské právo a podmínky zveřejnění) a další informace, které podporují dlouhodobé uložení digitálních dokumentů.¹³

1.5.1.5 Metadatové standardy Národní Digitální Knihovny (NDK)

Pro digitalizaci **v českých knihovnách existuje několik standardů digitalizace pro různé typy knihovních dokumentů**, které vytváří a spravuje Oddělení pro standardy Národní knihovny ČR.

Momentálně existují standardy pro monografie a periodika, e-publikace a některé druhy zvukových dokumentů (gramofonové desky a fonografické válečky). Do budoucna by se mělo portfolio standardů dále rozšiřovat o zvuková CD, kroniky a další dokumenty, které se vyskytují ve fondu tuzemských knihoven a jejichž digitalizace by měla teprve následovat.¹⁴

Standardy NDK využívají výše uvedené metadatové formáty (MODS, DC, METS, ALTO, PREMIS) a popisují datový a metadatový balíček, který obsahuje tyto části:

- **popisná metadata** (Dublin Core, MODS),
- **administrativní metadata** (PREMIS),
- **technická metadata** (MIX),
- **strukturální metadata** (METS),
- **archivní verze digitálního dokumentu** (JPEG 2000 u obrazových dat, WAVE u zvukových dokumentů, PDF/A a EPUB pro e-publikace),
- **uživatelská verze digitálního dokumentu** (JPEG 2000 u obrazových dat, MP3 u zvukových dokumentů),
- **OCR** ve formě ALTO a TXT souborů,
- **kontrolní součty**.

¹³ CAPLAN, Priscilla. UNDERSTANDING PREMIS [online]. Dostupné z: <http://www.loc.gov/standards/premis/understanding-premis-rev2017.pdf>

¹⁴ Aktuální výčet lze nalézt zde: <https://www.ndk.cz/standardy-digitalizace/metadata>

1.5.1.6 Formátový výbor NDK

Formátový výbor¹⁵ je **poradní orgán Národní knihovny ČR v otázkách dat a metadat** v oblasti digitalizace a digitálních dokumentů. Na jeho půdě se řeší standardy NDK, jejich úprava, případně tvorba nových standardů. Aktuálně má tři pracovní skupiny – pro digitalizaci klasických papírových dokumentů, pro e-publikace a pro zvukové dokumenty. Členy Formátového výboru jsou zástupci Národní knihovny ČR, Moravské zemské knihovny, krajských knihoven a dalších knihoven, které digitalizují podle standardů NDK a mají zájem se podílet na jejich rozvoji.

1.6 Zpřístupnění

Jednou z posledních instancí v celém procesu digitalizace je **import dat do digitální knihovny a zpřístupnění dokumentů uživatelům**. Digitální knihovna je tedy jak **úložiště dat**, tak i **prezentační vrstva**. Nejrozšířenější digitální knihovnou v knihovním prostředí v ČR je **system Kramerius**, na univerzitách je to repozitář DSpace. V zahraničí je pak k vidění celá plejáda různých systémů s více či méně povedeným uživatelským rozhraním a funkcemi. Obecně může být jednoduchou digitální knihovnou i prosté zveřejnění naskenovaných dokumentů na webu (např. ve formě JPG souborů jednotlivých stran nebo v PDF). Takové jednoduché rozhraní bude nicméně jen těžko nabízet moderní funkce pokročilých systémů digitálních knihoven.

Prezentace dokumentů v digitální knihovně přináší řadu výhod:

- dokument je v tu samou chvíli dostupný většímu množství uživatelů,
- pokud je dokument autorsky volný, je jeho prohlížení možné bez ohledu na čas a prostor,
- originál může být uložen v odpovídajících podmínkách a chráněn před poškozením, krádeží a dalšími riziky,
- uložení digitální verze v repozitáři slouží jako záloha fyzického originálu,
- knihovna může dokument využít pro svoji prezentaci, např. sdílet odkaz na sociálních sítích,
- přidaná hodnota pro uživatele – při existenci OCR je možné fulltextově prohledávat nejen celý dokument, ale rovněž i celou digitální knihovnu,
- autorsky volné dokumenty si uživatelé mohou stahovat a dále je využívat pro svoje potřeby (badatelské, umělecké, aj.),

¹⁵ Formátový výbor NDK. Dostupný z: <https://www.ndk.cz/formatovy-vybor-ndk>

- některé systémy nabízejí vytváření uživatelských účtů, kde si čtenáři mohou vytvářet své soukromé sbírky, vkládat do dokumentů záložky a využívat další pokročilé funkce, které moderní technologie nabízejí (např. on-line slovníky či odborné znalostní databáze).

PRO ZÁJEMCE

Vybrané digitální knihovny a repozitáře v ČR i zahraničí.

Porovnejte si grafické rozhraní, vyhledávání i další funkce:

- *Digitální knihovna.cz*: <http://www.digitalniknihovna.cz>
- *rozcestník DSpace repozitářů v ČR*: <https://www.dspace.cz/dspace-v-cr>
- *Manuscriptorium*: <http://www.manuscriptorium.com/cs>
- *Polona*: <https://polona.pl/>
- *Rakouská národní knihovna*: https://search.onb.ac.at/primo-explore/search?vid=ONB&lang=de_DE
- *Gallica*: <https://www.bnf.fr/en/gallica-bnf-digital-library>
- *Europeana*: <https://www.europeana.eu/portal/cs>
- *Google Books*: <https://books.google.cz/>

1.6.1 ZPŘÍSTUPNĚNÍ DOKUMENTŮ V DIGITÁLNÍ KNIHOVNĚ (AUTORSKÝ ZÁKON)

Zpřístupnění dokumentu v digitální knihovně se řídí zákonem č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).¹⁶ Z toho hlediska lze v digitální knihovně nalézt dva druhy dokumentů – **autorsky volné** a **nezpřístupněné**. Uživatel vidí a může vyhledat záznamy všech dokumentů a může si zobrazit jejich metadata a obálku. Avšak pouze **v budově knihovny si může prohlížet i ty nepřístupné. Autorsky volné lze zpřístupnit i mimo budovu knihovny.**

S držiteli práv lze případně uzavírat smlouvy, které umožní zpřístupnění dokumentů i mimo budovu knihovny.

¹⁶ Zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon). Dostupné z: <https://www.zakonyprolidi.cz/cs/2000-121>

Pro potřebu zveřejňování obsahu v digitální knihovně jsou důležitá ustanovení kolem majetkových práv. Dokumenty lze zveřejnit, pokud jsou splněny následující podmínky:

- Uplynulo **70 let** od smrti autora a rovněž všech spoluautorů:
 - též překladatele, jedná-li se o překlad,
 - ilustrátora, pokud má dokument ilustrace,
 - editora, pokud má významnou roli při pořádání díla (může být rovněž autorem předmluvy či doslovu apod.).
- Uplynulo **70 let** od zveřejnění v případě, že jde o dílo anonymní (není uveden autor ani instituce) či pseudonymní, a nelze zjistit totožnost autora. Nebo se jedná o dílo kolektivní, kde nejsou autoři uvedeni (ale je uvedena instituce).
- Uplynulo **50 let** od vydání nakladatelem.
- Uplynulo **25 let** od zveřejnění volného díla, které dosud nebylo zveřejněno (majetková práva patří v tomto případě tomu, kdo dílo zveřejnil).

Důležité je, že 70 let od úmrtí autora lze počítat dle autorského zákona až od 1. ledna následujícího roku po roce úmrtí. Tedy díla autorů zemřelých v roce 1949 mohou být zveřejněna až v roce 2020.

Při rozhodování, zda je dílo možné zveřejnit dle autorského zákona, doporučujeme prohlédnout přinejmenším následující části dokumentu: titulní stranu, tiráž, obsah (autorství kapitol), předmluvu a doslov.

Doporučujeme si vést vlastní databázi (např. ve formě excelovské tabulky) zveřejněných dokumentů a také těch, které budou v budoucnu zveřejnitelné (až budou splněny zákonné podmínky).

Kde lze hledat informace ohledně autorů:

- databáze Národních autorit ČR (<http://authority.nkp.cz/>),
- databáze VIAF – zejm. pro zahraniční autory (<http://viaf.org/>),
- internet – Wikipedia, Google, ...

1.7 Dlouhodobá ochrana digitálních dokumentů

Poté, co dokument projde digitalizační linkou, jsou data nahrána do digitální knihovny, kde jsou přístupná uživatelům, a zároveň by měla být uložena do dlouhodobého úložiště, aby byla zajištěna jejich ochrana do budoucna.

Digitální archivace, v angličtině se často používá zkratka **LTP (Long Term Preservation)**, je množina aktivit, která má za cíl **zajištění životního cyklu digitálních dokumentů tak, aby se svou životností vyrovnaly analogovým**. Nezaměřuje se pouze na výsledky digitalizace, ale též na tzv. digital born

dokumenty (tedy takové, které již v digitální podobě vznikly). Řada dokumentů nemá svoji analogovou verzi – vědecká data a informace, umění a kultura, software aj. jsou dnes často vytvářena jen v digitální podobě.

Digitální archivace má ještě jednu funkci, a tou je **zajištění ochrany investic** (finančních, lidských a dalších), který byly věnovány digitalizaci. Je tedy nutné zabezpečit, aby nepřišly vniveč a výsledky snažení řady lidí byly přístupné i v budoucnosti.

Dlouhodobý digitální repozitář (LTP repozitář) není to samé, co digitální knihovna. Cílem LTP repozitáře je dlouhodobě uchovávat digitální dokumenty, zatímco digitální knihovna je prezentační vrstvou, která slouží uživatelům k přístupu k datům. V praxi to může být tak, že nad dlouhodobým úložištěm je postavena digitální knihovna, ale také mohou být tyto dva systémy odděleny a fungovat nezávisle.

LTP repozitář se skládá z technologií, lidí a postupů a společně tyto vrstvy tvoří jeden systém, který **zajišťuje autenticitu a integritu uložených dat**. Důležitý je **mandát**, který má repozitář od svého zřizovatele (tento mandát může být vyjádřený legislativním dokumentem v případě národních institucí, zřizovací listinou nebo jiným způsobem) a dlouhodobé **financování**. Je nezbytné vědět, jaké **formáty** jsou v repozitáři uloženy a jaká **rizika** jim hrozí, kdo jsou uživatelé repozitáře (tzv. **určená komunita**) a jaké mají znalosti a dovednosti pro práci s daty.

V České republice je aktuálně několik dlouhodobých digitálních repozitářů:

- LTP repozitář Národní knihovny ČR (projekt Národní digitální knihovna),
- Národní digitální archiv (NDA),
- projekt ARCLib, který počítá s pilotním provozem v KNAV v roce 2020,
- repozitáře vědeckých dat Univerzity Karlovy v Praze:
 - Repository of Charles University in Prague, Map Collection,
 - LINDAT-Clarin – Centre for Language Research Infrastructure in the Czech Republic,
 - Český sociálněvědný datový archiv,
- Centrální datový archiv (CDA) v Bratislavě (Slovensko).

2 CENTRÁLNÍ REGISTR DIGITALIZOVANÝCH DOKUMENTŮ

V českém prostředí funguje jako centrální registr služba Registr digitalizace (RD)¹⁷, který je dostupný na webové adrese: <http://registrdigitalizace.cz/>.

Původně vznikl, aby se zabránilo duplicitní digitalizaci na pracovištích Národní knihovny ČR a Knihovny Akademie věd (KNAV). Od té doby se počet digitalizujících institucí výrazně zvýšil a vzrostla důležitost registru. Aktuálně provoz zajišťuje Národní knihovna, na rozvoji služby se podílí Moravská zemská knihovna a KNAV. Jednou z podmínek pro přidělení peněz z programu VISK 7 je hlášení výsledků digitalizace do RD.

Primární funkcí Registru digitalizace je ověření, zda dokument nebyl již digitalizován, nebo se jeho digitalizace v některé instituci neplánuje. Sekundárně se pomocí této služby dá zjistit, ve které digitální knihovně je dokument dostupný. Služba také nabízí přehled o tom, jaké knihovny jsou aktivní v digitalizaci,¹⁸ kolik dokumentů mají naskenováno, jaké typy dokumentů to jsou a další informace.

Do RD se dají hlásit tři stavy:

- Plánovaná digitalizace – instituce ohlásí svůj záměr a dodá základní bibliografické údaje o dokumentu. Funguje též jako rezervace titulu.
- Probíhající digitalizace – záměr může a nemusí být ohlášen, instituce dodá bibliografické údaje (mohou být podrobnější než u fáze plánování).
- Hotová digitalizace – instituce zpět ohlásí provedenou digitalizaci, dodá nejen bibliografické údaje, ale též odkaz na dokument.

Data se do RD dají zasílat několika cestami:¹⁹

- Importní dávka v MARCXML, která je umístěna na FTP, nebo odeslána přes webovou službu,
- Formát XLS(X) z MS Excel, který se hodí pro menší jednorázové dávky,
- Přes protokol OAI-MPH – vyžaduje, aby knihovna měla OAI provider, který vytvoří požadovaný XML soubor a ten je následně importován do RD,
- Následně je zpátky na e-mail zaslán výsledek importu:

¹⁷ Registr digitalizace [online]. Dostupné z: <http://registrdigitalizace.cz/>

¹⁸ Přehled institucí zapojených do registru digitalizace. Dostupné z: http://registrdigitalizace.cz/rdcz/info/prehled_instituci

¹⁹ Jak zasílat data. Dostupné z: <http://registrdigitalizace.cz/rdcz/info/data>

Pokud je v hlášení duplicita, knihovna se může rozhodnout, že duplicitní dokument z digitalizace vyřadí, nebo jej může i přesto digitalizovat. V obou případech je nutné v RD provést další krok,

- v případě nepokračování digitalizace musí být dokument z RD vyřazen,
- V případě digitalizace i přes hlášení duplicity se dokument přesune do stavu „plánovaná digitalizace“ nebo „ve zpracování“.

2.1 Číslo České národní bibliografie (čČNB)

Jedná se o **identifikátor, který byl vyvinut kvůli digitalizaci** – nejnovější dokumenty mají ISBN a ISSN, které je spolehlivě identifikují, ale to se bohužel nedá říct o starší produkci. A právě zde nastupuje čČNB – to se přiděluje všem dokumentům po roce 1800 (tzv. **novodobá bohemikální produkce**, netýká se tedy starých tisků a rukopisů). Bohužel ne všechny dokumenty čČNB mají, a přidělování ze strany Národní knihovny bývá pomalé, takže se do digitalizace a RD dostávají i dokumenty bez čČNB, což zhoršuje rozpoznání duplicit.

Proto Registr digitalizace kontroluje nejen čČNB, ale i další identifikátory – ISSN a ISBN a porovnává u záznamů údaje o autorovi, názvu a roku vydání. Pokud má systém podezření na duplicitu, zašle tuto informaci e-mailem do instituce, která import záznamů do RD zaslala. Odpovědný pracovník pak musí rozhodnout, zda se opravdu o duplicitu jedná.

2.2 Rozpoznání duplicity

Duplicita se dá kontrolovat dvěma způsoby:

1. Uživatel má přístup do účtu své instituce, která do RD zasílá importní dávky. V takovém případě bude duplicity rozpoznávat skrze webovou aplikaci Relief, která má pokročilé možnosti a zjednodušuje uživateli práci.
2. Uživatel přístup nemá a musí tedy využít webové rozhraní RD. I touto cestou se dá bez problémů duplicita rozeznat.

Kontrola duplicity přes webové rozhraní je snazší, proto si ji zde ukážeme na příkladu monografického dokumentu:

Registr digitalizace
evidence dokumentů digitalizovaných v ČR

Home Nápověda Info Kontakt Přihlásit English

Hledej v registru digitalizace ... ☐ Test duplicity DALŠÍ ZDROJE

Nalezeno: 1 položek

Použité filtry: čČNB: cnb001488580

Vlastník dokumentu (2) Digitální knihovna (2) Stav (2) Dokončeno (1) Ve zpracování (1)

Sloučeno podle: ČNB počet výsledků: 10 Seřazeno podle: relevance

Jana z Hvězdy. 1. díl, Růže a lilie
Variantní název: Růže a lilie
Hlavní autor: MESSADIÉ, Gerald
Místo vydání: Praha
Nakladatel/vydavatel: Argo
Rok(y) vydání: 2004
Vydání: Vyd. 1.
Podrobné bib. údaje

Vlastník	Počet svazků
Národní knihovna České republiky (ABA001)	1
Moravská zemská knihovna, Brno (BOA001)	1

Seznam předloh (2)

Stav	Číslo RDCZ	Vlastník	Rok per.	Část/roč.	Číslo	Financováno	Signatura	Čár. kód
■	Pr000411896	ABA001				NDK-udržitelnost	1 434271	1001090499
■	Pr000590708	BOA001				NDK-udržitelnost	2-1144.995,1	2610238859

Digitální dokument (2)

<http://kramerius4.nkp.cz/search/handle/uuid:405a5c60-ac00-11e6-840e-005056827e52>
<http://www.digitalniknihovna.cz/mzk/uuid/uuid:405a5c60-ac00-11e6-840e-005056827e52>

Předlohy: 2 | čČNB: cnb001488580 | ISBN: 80-7203-593-2 SKRÝT DETAIL

MESSADIÉ, Gerald. *Jana z Hvězdy. 1. díl, Růže a lilie*. Praha: Argo, 2004. ISBN 80-7203-593-2. čČNB: cnb001488580, čárový kód v MZK: 2610238859, signatura v MZK: 2-1144.995,1.

V případě knihy *Jana z Hvězdy. 1. díl, Růže a lilie* se jedná o duplicitu, protože tento dokument byl již zdigitalizován Národní knihovnou (viz obrázek – položka stav a vlastník).

V Registru digitalizace se dá vyhledávat podle následujících parametrů:

- název,
- hlavní autor,
- rok(y) vydání,
- nakladatel/vydavatel,
- ISSN/ISBN,
- čČNB,
- signatura,
- čárový kód,
- číslo RDCZ,
- lokální identita (001),
- ČAN.

Jako další příklad si uvedeme periodikum:

Sloučeno podle: ČNB ▼ počet výsledků: 10 ▼ Seřazeno podle: relevance ▼

Religio

Podnázev: Revue pro religionistiku
 Variantní název: Revue pro religionistiku
 Místo vydání: Brno
 Nakladatel/vydavatel: Česká společnost pro studium náboženství
 Rok(y) vydání: 1993-
 Podrobné bib. údaje

Vlastník	Počet svazků
✓ Národní knihovna České republiky (ABA001)	2
✓ Moravská zemská knihovna, Brno (BOA001)	25

— Seznam předloh (27)

Stav	Číslo RDCZ	Vlastník	Rok per. ↓	Část/roč.	Číslo	Financováno	Signatura	Čár. kód
■	Pr000067552	BOA001	2009	17	1-2	NDK	03429-1018.080	2610450625
■	Pr000060172	ABA001	2010			ANL plus		
■	Pr000067553	BOA001	2010	18	1-2	NDK	03429-1018.080	2610485446
■	Pr000127768	BOA001	2011	19	1-2	NDK	03429-1018.080	2610520114
■	Pr000060173	ABA001	2011			ANL plus		
■	Pr000128022	BOA001	2012	20	1-2	NDK	03429-1018.080	2610561390
■	Pr000235103	BOA001	2013	21	1-2	NDK-udržitelnost	03429-1018.080	2610603615
■	Pr000352455	BOA001	2014	22	1-2	NDK-udržitelnost	03429-1018.080	2610657099
■	Pr000423889	BOA001	2015	23	1-2	NDK-udržitelnost	03429-1018.080	2610685390
■	Pr000544382	BOA001	2017	25	1-2	NDK-udržitelnost	03429-1018.080	2610762351
■	Pr000590705	BOA001	2018	26	1-2	NDK-udržitelnost	03429-1018.080	2610802927

+ Digitální dokument (2)

Předlohy: 27 | čČNB: cnb000358425 | ISSN: 1210-3640 SKRÝT DETAIL

Religio: revue pro religionistiku. Brno: Česká společnost pro studium náboženství, 2018, 26(1-2). ISSN 1210-3640. čČNB: cnb000358425, čárový kód v MZK: 2610802927, signatura v MZK: 03429-1018.080.

RD sice v tomto případě hlásí duplicitu, ale je to z důvodu, že pod jedním záznamem je u periodik celá řada jednotek. Zmíněný 26. ročník periodika duplicitou není, i když se tak na první pohled jeví.

3 VYHLEDÁNÍ ČÍSLA ČESKÉ NÁRODNÍ BIBLIOGRAFII

Databázi České národní bibliografie spravuje Národní knihovna ČR a je volně přístupná na webu: [Katalogy a databáze Národní knihovny ČR](#)

Databáze Národní knihovny ČR

Konec | Přihlášení | Databáze | Dílicí báze | Nastavení | Otázky | MVS | Nápověda

Vyhledávání / Rejstříky | Výsledky dotazu | Předchozí dotazy | Schránka | Historie

Základní vyhledávání | Rozšířené vyhledávání | Z víceází | Vyhledávání CCL

CNB - Česká národní bibliografie

Základní vyhledávání

Přehled novinek | Informace o České národní bibliografii

Vyberte údaj pro vyhledávání: Všechny údaje

Zadejte slovo nebo slovní spojení:

Druh dokumentu: pouze dokumenty v češtině

Blízkost slov? ☒ Ne ☐ Ano

OK | Vyčistit formulář

Pro další informace použijte tabulku kódů jazyků, tabulku kódů zemí a tabulku kódů druhů dokumentu.

Úvodní obrazovka pro vyhledávání v České národní bibliografii

CNB - Úplné zobrazení záznamu

Zvolte formát: Standardní -- Katalogizační záznam -- Stručný záznam -- MARC -- [Citace](#)

Záznam 3 z 10

Hlavní záhlaví	Messadié, Gerald, 1931-
Název	Jana z Hvězdy. 1. díl, Růže a lilie / Gerald Messadié ; [přeložila Danuše Navrátilová]
Vydání	Vyd. 1.
Nakl. údaje	Praha : Argo, 2004
Popis (rozsah)	262 s. ; 22 cm
	Digitalizovaný dokument
EXEMPLÁŘE	
Číslo nár.bibl.	cnb001488580
ISBN	80-7203-593-2 (váz.)
Název originálu	Jeanne d'Estaille
Poznámka	Přeloženo z francouzštiny
Poznámka	Obsahuje bibliografické odkazy

Záznam konkrétního dokumentu s vyznačeným čČNB

4 ČÍSLO ISSN U POKRAČUJÍCÍCH DĚL

ISSN (International Standard Serials Numbering) je **mezinárodní jednoznačný identifikátor pro tištěné i online pokračující zdroje** (noviny, časopisy, sborník, apod). Mezinárodní centrum ISSN sídlí v Paříži, České národní středisko ISSN pak v Národní technické knihovně v Praze (NTK). ISSN se pojí s hlavním názvem publikace, změní-li se název, mění se i ISSN. Jeho používání je nepovinné, i když usnadňuje identifikaci jak vydavatelům, prodejcům, tak i knihovnám. ISSN lze přiřazovat (na rozdíl od ISBN) i zpětně již nevycházejícím pokračujícím zdrojům.

ISSN číslo sestává ze dvou skupin číslic po čtyřech číslech, která jsou oddělena rozdělovníkem. Poslední číslice má kontrolní funkci a nabývá hodnot 0–9, případně X (ve smyslu číslovky 10).

NTK spravuje národní databázi ISSN, ve které lze snadno vyhledávat: http://aleph.ntkcz.cz/F/?func=find-b-0&local_base=stk02

Úvodní obrazovka pro vyhledávání v národní databázi ISSN

Zobrazení plného záznamu:

Vyberte formát zobrazení: [Standardní formát](#) [Katalogizační lístek](#) [Citace](#) [Plný záznam s tagy](#) [Záznam ve formátu MARC](#)

Záznam 2 z 2

Číslo záz.:	000003053
ISSN	ISSN 1210-3640
Kl.název	Religio (Brno, Print)
Název	Religio
Naklad.údaje	Brno : Česká společnost pro studium náboženství, 1993-2007 Brno : Česká společnost pro religionistiku o.s., 2007-
Jazyk	cze, slo, eng
Periodicita	2x za rok
Další variant.ná	Revue pro religionistiku
Vydáváno	Roč. 1, č. 1 (1993)-
Vydavatel	FF MU, Ústav religionistiky
Vyd.na j.nosiči	Religio (Brno, On-line) 2336-4475

Záznam konkrétního dokumentu s vyznačeným ISSN

5 TYPY SKENERŮ

Skener je zařízení pro převod analogové předlohy do digitální podoby. V knihovním prostředí se nejčastěji setkáme s papírovými předlohami (typicky knihy, periodika, mapy, grafiky, rukopisy a staré tisky), takže některé typy skenerů (ruční, bubnové nebo 3D) zde nenajdou uplatnění, a proto jsou z našeho popisu vynechány.

Knihovna by měla disponovat takovou sestavou zařízení, která jí umožní digitalizovat drtivou většinu fondu. Ze zkušeností Moravské zemské knihovny v rámci projektu Národní digitální knihovny vyplývá, že nejuniverzálnější se prokázaly manuální skenery (ať už stolní nebo planetární). Ty se sice vyznačují nižší rychlostí skenování, ale ve spojení se šikovností operátora jsou schopny skenovat většinu typů dokumentů.

Důležité parametry skeneru:

- typ (stolní/planetární, manuální/robotický),
- velikost snímané plochy (obvykle A3-A0),
- rozlišení (obvykle 100–600 DPI, u některých zařízení až 1600 DPI),
- rychlost (strany za hodinu),
- univerzálnost (kolik různých typů dokumentů zvládne stroj skenovat),
- úroveň rozevření dokumentu (110°-180°),
- složitost obsluhy,
- náročnost údržby (frekvence a složitost čištění a výměny opotřebovaných částí),
- výstupní formáty.

5.1 Stolní skenery

Jedná se o nejjednodušší zařízení, která lze často najít i v kancelářích nebo domácnostech. Ke skenování dokumentu slouží sklo na vrchní straně zařízení, na které se dokument přitiskne a snímač pod sklem provede sken. Pro zamezení přístupu vnějšího světla bývá snímací plocha opatřena víkem.

Pro profesionální použití v knihovní digitalizaci je vhodné mít skenery, které jsou schopny skenovat i dokumenty, které se kvůli špatné vazbě nemohou otevřít na 180°. Jednou z možností jsou stolní skenery, které mají sklo až na hranu stroje, což umožňuje skenování knih s těsnou vazbou, která by jinak byla příliš namáhána a hrozilo by její prasknutí.

Dokumentové skenery

Umožňují skenování jednotlivých listů papíru, vynikají vysokou rychlostí zpracování, ale v knihovním prostředí je jen málo dokumentů, které lze tímto skenerem digitalizovat (proto tento typ najdete spíše v kancelářích). Hodí se pro zpracování knih vyřazených z fondu, kterým lze uříznout hřbet a pak je skenovat po jednotlivých listech (tzv. destruktivní digitalizace)²⁰. Další využití najdou při digitalizaci lístkového katalogu nebo v akademických knihovnách, které se rozhodnou skenovat fond starších závěrečných prací.

Ukázka fungování dokumentového skeneru: <https://youtu.be/xzg3OxHr9VA>

5.2 Planetární skenery

Jedná se o specializované knižní skenery, určené pro šetrné skenování dokumentů ve vazbě. Mohou být opatřené kolébkou, která umožňuje rozevření knižní vazby v rozmezí 60°-120° (dle konkrétního zařízení), což méně namáhá knižní vazbu. Z vrchu je dokument snímán kamerou či fotoaparátem (u větších formátů pak dvojicí kamer nebo fotoaparátů tzv. „do kříže“ – tedy kamera na pravé straně snímá levou stránku a naopak). Součástí bývá i osvětlovací systém a zatemnění proti okolnímu světlu pro zajištění stálých podmínek. Dle velikosti zařízení lze skenovat až dokumenty s rozměrem jedné strany A2, což umožňuje digitalizaci i velkoformátových periodik.



Dva typy planetárních manuálních skenerů

Ukázka planetárního manuálního skeneru v praxi:

https://youtu.be/vdCUh_OGm_U

²⁰ INDRÁK, Michal. Destruktivní digitalizace v Moravské zemské knihovně v Brně. Dostupné z: <https://www.svkhk.cz/Pro-knihovny/Zpravodaj-U-nas/Clanek.aspx?id=20170311>

Robotické skenery

Robotické skenery jsou většinou podobné konstrukce jako běžné planetární skenery, disponují však automatickým otáčením stránek – systém může pracovat buď s tlakem vzduchu (což je šetrnější způsob) nebo s mechanickým systémem (různé způsoby, např. tzv. mechanický prst). Zařízení většinou disponují systémem na rozpoznání otočení více stránek, ale vždy je potřeba dohled operátora, protože ne vždy je systém spolehlivý na 100 % (stránky knihy mohou být slepeny nebo jinak držet u sebe). Robotické skenery nabízejí vyšší rychlost skenování, vždy je ale nutná přítomnost a kontrola operátora, který musí zajistit, aby nedocházelo k chybnému skenování. Může dojít k posunutí svazku, špatnému otočení strany či jinému problému a tom případně musí obsluha zasáhnout a problém odstranit.

Ukázka automatického skeneru s mechanickým otáčením:

<https://youtu.be/qcuMxzIlygc>

- Automatický skener s kolébkou do V a otáčením stránek pomoc podtlaku:
<https://youtu.be/SdipuAuWsEs>
- Automatický skener s kombinovaným otáčením stránek:
<https://youtu.be/RROgUub-cDQ?t=60>
- Automatický planetární skener s mechanickým prstem pro otáčení stránek:
<https://youtu.be/uX0g4aNynro?t=40>

Velkoformátové skenery

Jsou ideální pro skenování velkých předloh, jako jsou mapy, grafiky, plakáty apod. Většinou jsou ve formátu A1 až A0, ale ve světě jsou i větší přístroje na skenování rozměrných map (např. rozměr plochy 2xA0+ pro námořní mapy). Skládají se z plochy pro upevnění předlohy, přitlačného skla a snímacího zařízení (kamera nebo senzor). Disponují schopností snímat plochu ve vysoké kvalitě (typicky 300 až 1200 DPI), která je potřeba k digitalizaci mapových podkladů a dalších rozměrných dokumentů.

Ukázka velkoformátového (A0) skeneru: <https://youtu.be/EZbjfr16Geo>

5.3 Ostatní skenery

V prostředí knihoven se lze setkat také se skenery mikrofilmů a mikrofiší, které se využívají ke skenování negativů, na které byly uloženy kopie dokumentů během ochranného reformátování (především významných tuzemských periodik). Tyto skenery umožňují skenovat svitkové mikrofilmy 35 mm, 16 mm a mikrofiše. Výstupem jsou obrazy v 16bitové šedé škále.



Mikrofilmový skener

6 SHRnutí STUDIjNíHO TEXTU

Digitalizace patří v současnosti k moderním službám knihovny – uživatel očekává, že se k dokumentu dostane snadno a rychle, nechce čekat na vyhledání a doručení dokumentu z fondu. Díky digitálním knihovnám nejsou čtenáři omezeni časem a prostorem (pokud odhlédneme od omezení daných zákony).

Proces digitalizace se skládá ze správného výběru dokumentů, jejich přípravy k digitalizaci, skenování, úpravě obrazů a metadat, importu do digitální knihovny a v neposlední řadě též v dlouhodobé archivaci, která zajišťuje ochranu investovaných zdrojů a také přístup k digitálním dokumentům v dlouhodobém časovém horizontu.

Při samotném zpracování se musí dbát na to, aby nedocházelo k duplicitní digitalizaci, proto se dokumenty ověřují pomocí čísla české národní bibliografie a dalších identifikátorů (ISSN, ISBN a další) v centrální databázi – Registru digitalizace.

Při skenování je důležité umístit dokument na vhodný skener, protože každé zařízení se hodí na určitý druh předlohy, kdy hraje roli typ vazby, gramáž papíru a další parametry, které ovlivňují kvalitu naskenovaných obrazů.

Při digitalizaci je důležitá spolupráce všech, kteří se na tomto procesu podílejí – od výběru vhodných knih, přes jejich skenování, úpravu obrazů, až po správce digitální knihovny. Je to však čtenář, na koho se nesmí zapomínat a měly by to být jeho potřeby, které by měla digitalizace a digitální knihovna naplňovat.

7 POUŽITÁ LITERATURA

ALTO: Technical Metadata for Layout and Text Objects. *The Library of Congress* [online]. 2019 [cit. 2019-12-01]. Dostupné z: <https://www.loc.gov/standards/alto/>

CAPLAN, Priscilla. *UNDERSTANDING PREMIS* [online]. Rev. 2017 [cit. 2019-11-23]. Dostupné z: <http://www.loc.gov/standards/premis/understanding-premis-rev2017.pdf>

CUBR, Ladislav a Zdeněk VAŠEK. Identifikátory digitálních dokumentů se zaměřením na systém URN:NBN v ČR. *Čtenář: Měsíčník pro knihovny* [online]. 2013, **65**(6) [cit. 2019-12-01]. Dostupné z: <https://svkkl.cz/en/ctenar/clanek/1254>

Dublin Core. *Masarykova univerzita* [online]. 2006 [cit. 2019-11-12]. Dostupné z: http://webserver.ics.muni.cz/dublin_core/index.html

DVOŘÁKOVÁ, Helena. Registr digitalizace. *Ikaros* [online]. 2014, ročník 18, číslo 3 [cit. 2019-11-27]. urn:nbn:cz:ik-14208. ISSN 1212-5075. Dostupné z: <http://ikaros.cz/node/14208>

ElarSCAN A2 Planetary Book Scanner. *iGuana* [online]. c2019 [cit. 2019-12-01]. Dostupné z: <https://iguana-idm.com/elarscan-a2-planetary-book-scanner/> Extensible Markup Language. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-12-01]. Dostupné z: https://cs.wikipedia.org/wiki/Extensible_Markup_Language

Fedora Object XML (FOXML). *Wiki Lyra* [online]. 2014 [cit. 2019-12-01]. Dostupné z: <https://wiki.lyrasis.org/pages/viewpage.action?pageId=66585857>

Formátový výbor NDK. *Národní digitální knihovna* [online]. 2019 [cit. 2019-11-12]. Dostupné z: <https://www.ndk.cz/formatovy-vybor-ndk>

INDRÁK, Michal. Destruktivní digitalizace v Moravské zemské knihovně v Brně. *U ná* [online]. **27**(3) [cit. 2019-12-01]. ISSN 2571-4597. Dostupné z: <https://www.svkhhk.cz/Pro-knihovny/Zpravodaj-U-nas/Clanek.aspx?id=20170311>

International Standard Serial Number. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-11-21]. Dostupné z: https://cs.wikipedia.org/wiki/International_Standard_Serial_Number

Jak zasílat data. *Registr digitalizace* [online]. c2017 [cit. 2019-12-01]. Dostupné z: <http://registrdigitalizace.cz/rdcz/info/data>

KRAVECOVÁ, Lenka. Kontrola autorských práv u neveřejných dokumentů v Krameriu MZK. Brno : Moravská zemská knihovna v Brně, 2019.

METS: Metadata Encoding & Transmission Standard. *The Library of Congress* [online]. 2019 [cit. 2019-12-01]. Dostupné z: <http://www.loc.gov/standards/mets/>

Mezinárodní registrační systémy: ISSN. *Národní knihovna ČR* [online]. 2019 [cit. 2019-12-01]. Dostupné z: <https://www.nkp.cz/sluzby/sluzby-pro/isbn-ismn-issn/oma#ISSN>

MIX: NISO Metadata for Images in XML Schema. *The Library of Congress* [online]. 2015 [cit. 2019-12-01]. Dostupné z: <https://www.loc.gov/standards/mix/>

Projekt PERO [online]. 2019 [cit. 2019-11-23]. Dostupné z: <https://pero.fit.vutbr.cz/>

Přehled institucí zapojených do registru digitalizace. *Registr digitalizace* [online]. c2017 [cit. 2019-12-01]. Dostupné z: http://registrdigitalizace.cz/rdcz/info/prehled_instituci

Qidenus Mastered Book Scanner. *iGuana* [online]. c2019 [cit. 2019-12-01]. Dostupné z: <https://iguana-idm.com/qidenus-mastered-book-scanner/>

Registr digitalizace [online]. c2017 [cit. 2019-12-01]. Dostupné z: <http://registrdigitalizace.cz/>

RYCHTÁŘOVÁ, Pavla, KRAVECOVÁ, Lenka a Markéta KRUTSKÁ: *Metadata II. E-learningový kurz Digitalizace v knihovnách* [online]. Brno : Moravská zemská knihovna v Brně, 2019 [cit. 2019-11-20]. Dostupné z: <https://kurzy.knihovna.cz/>

RYCHTÁŘOVÁ, Pavla, KRAVECOVÁ, Lenka a Markéta KRUTSKÁ: *Skenování. E-learningový kurz Digitalizace v knihovnách* [online]. Brno : Moravská zemská knihovna v Brně, 2019 [cit. 2019-11-20]. Dostupné z: <https://kurzy.knihovna.cz/>

SNÍŽKOVÁ, Martina. *METS* [online]. 2012 [cit. 2019-12-01]. Dostupné z: <http://wiki.knihovna.cz/index.php/METS>

Standardy digitalizace. *Národní digitální knihovna* [online]. 2019 [cit. 2019-11-12]. Dostupné z: <https://www.ndk.cz/standardy-digitalizace>

Standardy pro metadata. *Národní digitální knihovna* [online]. 2019 [cit. 2019-11-12]. Dostupné z: <https://www.ndk.cz/standardy-digitalizace/metadata>

Universally unique identifier. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-11-21]. Dostupné z: https://en.wikipedia.org/wiki/Universally_unique_identifier

VISK 7: Národní program digitálního zpřístupňování dokumentů ohrožených degradací kyselého papíru – Kramerius. *VISK: Veřejné Informační Služby*

Knihoven [online]. 2019 [cit. 2019-11-21]. Dostupné z: <https://visk.nkp.cz/visk-7>

What is microfilm/microfiche? Where is the microfilm/microfiche located?

Purchase College Library [online]. 2018 [cit. 2019-12-01]. Dostupné z: <http://purchase.libanswers.com/faq/36771>

Zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon). Ročník 2000, číslo 121.

Dostupné také z: <https://www.zakonyprolidi.cz/cs/2000-121>

Název:	Organizace digitálního zpracování dokumentů
Autor:	Mgr. Zdeněk Hruška
Jazyková korektura:	Markéta Roupcová
Počet stran:	30