



KNIHOVNA

—
AKADEMIE VĚD ČR

Strategie digitalizace pro období 2023-2025

Knihovna AV ČR, v. v. i.

leden 2023
verze 1.0

Obsah

1 Současný stav digitalizace v KNAV	3
1.1 Fond a digitalizace	3
1.2 Personální zabezpečení a technické vybavení procesu digitalizace	3
1.2.1 Kompetence jednotlivých oddělení	3
1.2.2 Skenery (DC)	4
1.2.3 Softwarové vybavení přímo související s digitalizací (DC, ODK, OZČ)	5
1.3 Výběr dokumentů k digitalizaci	6
1.3.1 Novodobý fond	6
1.3.2 Historický knižní fond	7
1.3.3 Born digital dokumenty	7
1.3.4 Kapacita	7
2 Strategie digitalizace na období 2023-2025	8
2.1 Cíle	8
2.2 Závazná kritéria výběru dokumentů k digitalizaci	8
2.3 Závazná kritéria výběru elektronických dokumentů ke zpracování	9
2.4 Postupy při digitalizaci specifických dokumentů	10
Seznam zkratk	11

1 Současný stav digitalizace v KNAV

1.1 Fond a digitalizace

Digitalizace v KNAV se v současnosti zaměřuje na periodika a monografie z fondu Knihovny AV ČR, v. v. i., především pak knihy a časopisy z produkce ústavů AV ČR a jejich předchůdců.

Fond KNAV je obecně zaměřen na dokumenty obsahující informace vědecké, výběrově populárně naučné, encyklopedické a biografické a na vědeckou publikační činnost zaměstnanců AV ČR, a to v češtině a základních světových jazycích. Důraz je kladen také na doplnění nejnovější zahraniční vědecké literatury.

Kromě digitalizace tištěných publikací jsou pro Digitální knihovnu AV ČR zpracovávány také dokumenty, které byly vytvořeny přímo v digitální podobě. Jde o elektronické monografie vydávané pracovišti AV ČR a především novější články z odborných periodik - v digitální knihovně se tak nacházejí tituly periodik, jejichž starší ročníky jsou tvořeny naskenovanými stránkami, zatímco novější ročníky jsou ve formátu PDF dodávaných přímo vydavateli.

1.2 Personální zabezpečení a technické vybavení procesu digitalizace

1.2.1 Kompetence jednotlivých oddělení

- **Digitalizační centrum (DC)**
 - skenování, tvorba metadat, organizační zajištění,
 - kromě digitalizace fondu KNAV zpracovává DC také externí digitalizační zakázky a podílí se na vývoji nástroje ProArc.
- **Oddělení pro digitální knihovnu (ODK):**
 - výběr dokumentů k digitalizaci, kontrola kvality, správa digitální knihovny, správa dlouhodobého uložení,
 - kromě digitalizace se ODK podílí na vývoji nástrojů ProArc, Kramerius, ARCLib a dalších.
- **Oddělení zpracování časopisů (OZČ)**
 - zpracování plných textů časopisů v PDF na úroveň článků,
 - oddělení se především věnuje jmenné a věcné katalogizaci časopiseckého fondu KNAV.
- **Knihovědné oddělení (KVO)**
 - metadatové zpracování starých tisků v systému ProArc,
 - KVO se primárně věnuje výzkumu české knižní kultury a shromažďování pramenných zdrojů k produkci českých a moravských tiskáren 16.–18. století.

- **Oddělení pro správu informačních technologií (OSIT),
Oddělení pro uživatelskou podporu informačních technologií (OPIT)**
 - správa systémů Kramerius, ProArc, ARCLib a dalších, správa datového úložiště,
 - pracovníci OPIT poskytují přímou podporu všem zaměstnancům KNAV v budově na Národní třídě i na detašovaném pracovišti v Jenštejně, pro digitalizační centrum zabezpečují správu softwaru LIMB,
 - OSIT má na starost zabezpečení provoz a rozvoje IT infrastruktury pro interní chod samotné organizace, pro dlouhodobé ukládání dat spravuje a postupně rozvíjí datové úložiště.
- **Oddělení Akvizice (OA)**
 - výběr dokumentů k digitalizaci (novinky)
 - oddělení především zajišťuje akvizici zahraničních knih, českých knih a předplatného periodik.
- **Oddělení zpracování knih a elektronických nosičů (OZKEN)**
 - kontrola a úprava katalogizačních záznamů před zpracováním v ProArcu,
 - oddělení především zpracovává veškerý knižní fond přicházející do KNAV, vytváří katalogizační záznamy v systému Aleph, opatřuje dokumenty jmenným a věcným popisem, adjustací a signováním.

1.2.2 Skenery (DC)

- **Fujitsu**
 - průtahový skener, na kterém je možné skenovat jednotlivé listy do formátu A3, bere i tvrdší papír (například kartotéční lístky), má široké možnosti nastavení
- **Plustek**
 - použití pro skenování běžné produkce do rozměru 42x29, skenuje až do hřbetu předlohy.
- **Suprascan A1**
 - vhodný zejména ke skenování historického fondu, například starých tisků větších formátů (např. plány apod.), skenuje až do rozměru A1+.
- **Zeutschel OS 12002 V**
 - skener vhodný ke skenování starých tisků a dalších dokumentů s pevnější vazbou, které není možné plně rozevřít do rozměru A4+ každá strana, nelze ho použít ke skenování dokumentů tvořených jednotlivými listy.
- **Zeutschel OS 15 000**
 - použití ke skenování běžné produkce do formátu A4+ pro každou stranu nebo ke skenování dokumentů tvořených jednotlivými listy formátu A3.

1.2.3 Softwarové vybavení přímo související s digitalizací (DC, ODK, OZČ)

- **LIMB**
 - aplikace pro zpracování a úpravu naskenovaných obrazů,
 - v aplikaci je integrována také služba ABBYY FineReader Server pro optické rozpoznávání znaků (OCR).
- **ProArc**
 - open source systém pro produkci digitálních dokumentů,
 - umožňuje rychlou částečně automatizovanou produkci digitálních dokumentů zahrnující popisná, strukturální a archivační metadata i OCR/ALTO,
 - produkovaná data jsou kompatibilní se systémem Kramerius a podporuje publikované standardy Národní knihovny ČR pro digitalizaci¹.
- **Kramerius**
 - open source softwarové řešení pro zpřístupnění digitálních dokumentů,
 - primárně je Kramerius určen pro digitalizované knihovní sbírky, monografie a periodika, využít může být ale také ke zpřístupnění dalších typů dokumentů např. map, hudebnin a starých tisků, případně částí dokumentů jako jsou články a kapitoly, použít ho lze i pro born digital dokumenty,
 - systém poskytuje rozhraní pro přístup koncových uživatelů, zajišťující vyhledávání v metadatech a v plných textech, generování vícestránkových PDF dokumentů z vybraných stran a další operace nad uloženou sbírkou digitálních dokumentů².
- **ARCLib**
 - LTP (Long Term Preservation) řešení na bázi open source, zajišťující dlouhodobou ochranu knihovních digitálních dat s kompletní podporou OAIS,
 - podporuje příjem archivních balíčků ze softwaru ProArc, NDK SIP charakterizovaných Národní knihovnou ČR i dat exportovaných z digitální knihovny Kramerius³.

¹ <https://github.com/proarc/proarc/wiki>

² <https://github.com/ceskaexpedice/kramerius>

³ <https://github.com/LIBCAS/ARCLib>

1.3 Výběr dokumentů k digitalizaci

1.3.1 Novodobý fond

Výběr dokumentů k digitalizaci v rámci novodobého fondu vychází z několika zdrojů:

Uživatelé

Uživatelé zadávají tipy na digitalizaci dokumentů prostřednictvím formulářů, které jsou umístěny u konkrétních záznamů v knihovním katalogu. Následně proběhne ze strany ODK ověření, zda je dokument možné digitalizovat, uživatel je o rozhodnutí informován. Navrhnout digitalizaci je možné v rámci projektů EOD⁴ a EODOPEN⁵. EOD představuje placenou službu, skrze kterou si mohou registrovaní i neregistrovaní uživatelé objednat z fondu KNAV digitalizaci jakékoliv monografie, která byla vydaná v letech 1500–1910 a není již chráněna autorským zákonem. Na tuto službu navazuje projekt EODOPEN, který si klade za cíl zpřístupnit publikace z 20. a 21. století v digitální podobě pro širokou veřejnost zdarma.

OA

Oddělení akvizice vybírá z fondu KNAV každý měsíc 10 zajímavých nových přírůstků, které jsou digitalizovány a zpřístupněny na počítačích ve studovně KNAV. Čtenáři KNAV se tak dostanou k novému titulu i pokud je fyzický nosič vypůjčen.

ODK

V Oddělení pro digitální knihovnu probíhá výběr dokumentů, které jsou digitalizovány a následně zveřejněny v rámci projektu EODOPEN, dále zde probíhá výběr dokumentů z fondu KNAV, které vydávají pracoviště AV (signatura AK).

Současné nastavení priorit pro digitalizaci novodobého fondu:

- tituly objednané uživateli prostřednictvím služby EOD,
- dokumenty vhodné pro zveřejnění v rámci projektu EODOPEN,
- novinky z fondu KNAV (každý měsíc 10 monografií - vybírá OA),
- monografie vhodné pro zveřejnění v rámci licence DNNT,
- dokumenty "K dokončení" - dříve naskenované tituly různého původu a v různých jazycích, které ještě neprošly metadatovým zpracováním,
- monografie se signaturou "AK" - jedná se o publikace vydávané pracovišti AV ČR.

Zároveň se při výběru dokumentů zohledňuje:

- autorsko-právní hledisko výběru (zda může být digitalizované dílo zpřístupněné volně, v rámci licence DNNT, nebo pouze v budově knihovny),
- jazykové hledisko, upřednostňují se dokumenty zejména v češtině a angličtině.

⁴ <https://www.lib.cas.cz/sluzby/e-knihy-na-objednavku/>

⁵ <https://www.lib.cas.cz/o-knihovne/projekty/eodopen/>

1.3.2 Historický knižní fond

Paralelně probíhá postupně digitalizace historického knižního fondu KNAV⁶. Výběr dokumentů provádějí pracovníci KVO podle stanovených kritérií:

- množství dochovaných exemplářů,
- unikátnost konkrétního exempláře,
- úplnost exempláře,
- existence digitální kopie,
- stav exempláře.

1.3.3 Born digital dokumenty

- **elektronická periodika**
 - pro Digitální knihovnu je zpracováváno přes 65 titulů periodik až na úroveň jednotlivých článků,
 - dále dochází k uzavírání nových smluv na nová periodika,
 - plné texty časopisů jsou získávány na základě smluv přímo od redakcí, a to ve formátu PDF. Velmi často je ale nutné doskenovat některé části dokumentu (nejčastěji obálky) z fyzické předlohy. To zajišťuje Digitalizační centrum ve spolupráci s OZČ.
- **elektronické monografie**
 - zpracovávány jsou elektronické monografie vydané ústavu AV ČR.
 - Nutnou podmínkou je uzavření rámcové a následně licenční smlouvy s daným pracovištěm (k červenci 2022 má KNAV tuto smlouvu s 29 ústavu AV ČR).
 - Výběr titulů ke zpracování nyní neprobíhá, jsou zpracovávány všechny elektronické monografie, které jsou KNAV jednotlivými ústavu poskytnuty.
 - Od července 2021 do června 2022 bylo takto zpracováno cca 250 titulů.

1.3.4 Kapacita

V rámci projektu EODOPEN se od jeho počátků v roce 2020 do srpna 2022 zdigitalizovalo 675 monografií. V rámci systematické digitalizace je každý rok digitalizováno přibližně 800 svazků monografií, z nichž cca 200 představují staré tisky. Průběžně se také zpracovávají nová čísla přibližně 65 titulů elektronických periodik.

⁶ <https://kvo.lib.cas.cz/historicky-fond/>

2 Strategie digitalizace na období 2023-2025

2.1 Cíle

Dlouhodobý cíl: Digitální knihovna by měla obsahovat všechny dokumenty, které vydala některá ze součástí AV ČR. Primárně se jedná o ty dokumenty, které jsou vedeny pod signaturou AK.

Dílčí cíle na období let 2023-2025:

- počet nově digitalizovaných svazků monografií (včetně starých tisků): 2400,
- počet nově digitalizovaných stran (monografie i periodika): 160000,
- počet zpracovaných born digital dokumentů:
 - 900 elektronických monografií,
 - 900 čísel elektronických periodik.

2.2 Závazná kritéria výběru dokumentů k digitalizaci

1) Výběrová digitalizace

- **EOD - Požadavky čtenářů**, placená služba
- **Novinky**
 - každý měsíc (kromě letních prázdnin) 10 nově vydaných monografií,
 - vybírá OA.
- **projekt EODOPEN**,
 - dokumenty vhodné pro zveřejnění v rámci projektu - publikace z 20. a 21. století,
 - výběr titulů provádí čtenáři skrze požadavky v knihovním katalogu a pracovník ODK.

2) Ochranná digitalizace

Digitalizace těch děl, u nichž hrozí vážné riziko ztráty originálu.

3) Systematická digitalizace

Do digitalizace budou postupně zařazována všechna díla vydaná v rámci AV ČR (monografie na signatuře AK a periodika).

- bude vytvořen přehled všech dokumentů na signatuře AK a podle něj bude vytvořena fronta dokumentů čekajících na digitalizaci,
- při sestavování fronty mohou být některé dokumenty upřednostněny z těchto důvodů:
 - hrozba rizika ztráty originálu (krádež, rozpad),
 - možnost vzdáleného zpřístupnění dokumentu v digitální knihovně,
 - velký zájem o vypůjčení titulu,
 - příslušnost dokumentu ke sbírce/kolekci,
 - možnost využít destruktivní digitalizaci,
 - jazykové hledisko (přednost budou mít dokumenty v češtině a angličtině)

- Dokumenty spadající do této kategorie, které již zdigitalizovala jiná instituce, se nebudou digitalizovat znovu, pokud bude možné s danou institucí vyjednat jejich replikaci do Digitální knihovny AV ČR.

4) Destruktivní digitalizace

V případě vyřazování děl z fondu knihovny se vždy individuálně zváží, zda není vhodné daná díla digitalizovat a zařadit do Digitální knihovny, umožní-li to kapacita.

Historický knižní fond

Výběr dokumentů k digitalizaci bude probíhat podle výše popsaných podmínek beze změny, posuzovat se tedy bude:

- množství dochovaných exemplářů,
- unikátnost konkrétního exempláře tzn. vlastnické a provenienční znaky, vazba, rukopisné přípisky,
- úplnost exempláře (nekompletní tisky se nedigitalizují, jen pokud jde o unikáty či je počet dochovaných exemplářů malý, a to primárně na území ČR),
- existence digitální kopie (zda, kolikrát a jak již je kniha digitalizovaná)
- stav exempláře (digitalizace je v tomto bodě často spojena s restaurováním, a to dle stavu tisku buď před nebo po digitalizaci)

2.3 Závazná kritéria výběru elektronických dokumentů ke zpracování

Elektronické monografie

Postupně budou zpracovávány všechny formálně vydané elektronické publikace z produkce jednotlivých ústavů AV ČR. Za účelem jejich získání budou ústavy pravidelně oslovovány ze strany KNAV.

Elektronické časopisy

Zpracovávány jsou takové elektronické časopisy, jejichž vydavatel nebo spoluvydavatel je některý z ústavů AV ČR nebo jeho předchůdců.

2.4 Postupy při digitalizaci specifických dokumentů

Vícedílné monografie

Při digitalizaci dokumentů, u kterých existuje pochybnost o tom, zda se jedná o vícedílnou monografii, nebo ne, je potřeba před zahájením zpracování nechat katalogizátory daný dokument zkontrolovat a adekvátně upravit katalogizační záznam/záznamy.

Staré tisky

Při skenování rukopisů, prvotisků a starých tisků (t.j. tisky vydané do roku 1800) se vždy používá podložení jednotlivých listů černým papírem a skenuje se s přesahem kolem celé strany, tedy i na protilehlou stranu. Skenuje se hřbet a všechny ořízky. Před zahájením skenování celé předlohy se udělá sken speciální barevné kalibrační tabulky, která slouží k následnému vytvoření ICC profilu. Dále se skenuje měřítko u titulní strany a u ilustrací, iniciál nebo jiných zajímavých prvků. Při skenování konvolutů se naskenuje celý konvolut do jednoho souboru.

Skeny se nijak graficky neupravují. Při zpracování v ProArcu se používá model pro STT. Nyní se z důvodu nízké výstupní kvality nepoužívá OCR. Konvoluty se zpracovávají samostatně. Desky, ořízky, hřbet, měřítko a tabulka se přiřazují k hlavnímu dílu. Jednotlivé konvoluty jsou v Krameriovi samostatně.

Přívazky

Existuje-li přívazek i jako samostatná monografie, měla by být digitalizována ta. Pokud ne, pak se z každého přívazku vytvoří samostatný digitalizát, který bude začínat titulní stranou a končit poslední stranou textu nebo prázdnou stranou (stranami), pokud nejsou desky vevázané. K přívazkům, které desky nemají, se vždy zkusí desky sehnat skrze MVS. Ke každému přívazku bude existovat samostatný záznam v katalogu, který se použije při metadatovém zpracování.

Přílohy monografie

- **CD/DVD** - přílohy, které se nedigitalizují pro Digitální knihovnu, v poznámce v metadatech by ale vždy měla být informace o existenci CD/DVD. Pro archivaci příloh typu CD/DVD se používá projekt CDArcha⁷
- **Mapy** - jako samostatné dokumenty se v KNAV nedigitalizují, neboť až na výjimky nejsou ve fondu knihovny zastoupeny. Mapy, které jsou přílohou monografie, se digitalizují stejným způsobem jako zbytek publikace.
- **Erratum**, vložené **poznámky** - řeší se poznámkou v elementech "note" a "extent".

⁷ <https://github.com/moravianlibrary/CDArcha-client>

Seznam zkratek

zkratka	význam
ALTO	Analyzed Layout and Text Object (XML schéma popisující rozložení a obsah textu na stránce)
AV ČR	Akademie věd ČR
DC	Digitalizační centrum
DK	Digitální knihovna
DNNT	Díla nedostupná na trhu
EOD	eBooks-On-Demand
EODOPEN	eBooks-On-Demand-Network Opening Publications for European Netizens
KNAV	Knihovna Akademie věd ČR
KVO	Knihovědné oddělení
LTP	Long-term preservation (dlouhodobá ochrana digitálních dokumentů)
NDK	Národní digitální knihovna
OA	Oddělení akvizice
OCR	Optical Character Recognition (Optické rozpoznávání znaků)
ODK	Oddělení pro digitální knihovnu
OPIT	Oddělení pro uživatelskou podporu informačních technologií
OSIT	Oddělení pro správu informačních technologií
OZČ	Oddělení zpracování časopisů
OZKEN	Oddělení zpracování knih a elektronických nosičů
SIP	Submission Information Package (balíček dat a metadat)
STT	Staré tisky