EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Podpořeno projektem: Zvýšení kvality vzdělávání na UK a jeho relevance pro potřeby trhu práce (CZ.02.2.69/0.0/0.0/16_015/002362)

Metodika digitalizace pro fakulty a další pracoviště UK

Verze 1.1

14. 9. 2023

Digitalizační centrum UK
Oddělení repozitářů a digitalizace

UNIVERZITA KARLOVA
Ústřední knihovna

Obsah

Úvod

Tato metodika slouží pro interní účely Univerzity Karlovy (UK). Definuje standardy a popisuje procesy digitalizace tištěných dokumentů na univerzitě v rámci sedmi dílčích kapitol. Standardy digitalizace přitom vycházejí ze standardů definovaných Národní digitální knihovnou (již zaštiťuje Národní knihovna ČR) a chtějí tak dostát potřebné kvality jak pro archivaci digitálních dokumentů, tak pro potřeby čtenářů, studentů a všech uživatelů univerzitní digitální knihovny Kramerius. Metodika popisuje procesy digitalizace, které na univerzitě běží a fungují. Chce tyto procesy sjednotit, aby právě jejich rámcové sjednocení utřídilo a ucelilo univerzitní digitalizační workflow, aby tak ve výsledku došlo k ulehčení práce a šetření času všem zúčastněným stranám. V neposlední řadě se z takto nastavených procesů může vycházet také při zavádění jejich další automatizace, k níž by mělo digitalizační workflow na univerzitě do určité míry směřovat.

Doufám, že metodika bude ku prospěchu jak jednotlivým fakultám a dílčím digitalizačním pracovištím na UK, tak Digitalizačnímu centru UK provozovanému Ústřední knihovnou UK. Její přílohou je pak elementární grafické znázornění digitalizačního workflow pro vytvoření si představy o jednotlivých procesech a jako referenční rámec pro jednotlivá digitalizační pracoviště na UK.

S poděkováním za spolupráci

Jáchym Šenkyřík

V Praze, 14. 9. 2023

1. Příprava digitalizace

Před zahájením digitalizace vybraných dokumentů, které chce fakulta či jiné pracoviště UK digitalizovat, je nejprve nutné zhodnotit, zda je jejich digitalizace vhodná a žádoucí.

1.1. Kontrola, zda dokument není již digitalizovaný

Tato kontrola probíhá primárně pomocí Registru digitalizace (RD; www.registrdigitalizace.cz), ideálně pomocí identifikátorů čČNB, případně ISBN/ISSN. V registru je možné vyhledávat i pomocí dalších bibliografických údajů. Existují však digitalizovaná bohemika zaznamenaná i mimo tento Registr digitalizace nebo také v zahraniční digitalizaci. Kontrola digitalizovaných titulů tedy probíhá i přes jiné databáze, pokud má člověk podezření, že by v nich konkrétní titul mohl být nalezen.[1]

Pokud vybraný titul je digitalizovaný, splňuje standardy digitalizace na UK (viz podkapitola 1.4 a kapitola 4) a fakulta či jiné pracoviště UK jej chce mít v univerzitní digitální knihovně (Kramerius), není nutné jej opětovně digitalizovat na univerzitě, ale požádáme o jeho replikaci tu instituci, která dokument digitalizovala. Takový titul nahlaste Digitalizačnímu centru UK pomocí tabulky k replikaci (viz kapitola 3). Pokud replikace titulu nebude z jakéhokoliv důvodu možná, lze jej případně digitalizovat na UK. Výjimku tvoří pak ty tituly, které jsou v cizích digitálních knihovnách veřejnosti volně dostupné (nepodléhají již autorskému zákonu) nebo jsou dostupné jako DNNT (Díla nedostupná na trhu) – replikace takových titulů není v obecném zájmu UK (více k DNNT viz podkapitola 1.3), pokud neexistuje pádný důvod duplikovat takový titul také v univerzitní digitální knihovně (např. kvůli doplnění určité tematické sbírky nebo periodika atd.).

Pokud vybraný titul není digitalizovaný, ale jeho digitalizaci plánuje nebo už zpracovává jiná instituce (podle záznamu v RD), také zažádáme o jeho replikaci (pokud tedy nebude volně dostupný nebo jako DNNT, viz předchozí odstavec). Titul opět nahlaste Digitalizačnímu centru UK pomocí tabulky k replikaci (viz kapitola 3) a do sloupce „Poznámky" k němu uveďte, že jeho digitalizace ještě není hotová: je tedy ve zpracování nebo se plánuje. Může však nastat situace, kdy pro pedagogické, studijní či vědecké účely je důležité digitalizovat vybraný dokument v co nejrychlejší možné lhůtě. Pokud zjistíte od instituce,

[1] Srov. např. heraldické časopisy, digitalizace Zeměměřického úřadu, samizdatová a exilová literatura, kramářské listy, nebo zahraniční knihovny jako Österreichische Nationalbibliothek, Bayerische Staatsbibliothek (odkazy na tyto databáze lze dohledat na stránkách RD, viz http://registrdigitalizace.cz/rdcz/info/digbohemika). Dále také srov. eSbírky (https://www.esbirky.cz/), databázi digitalizovaných mapových sbírek (http://dms.euweb.cz/databaze.php) či světovou databázi digitalizovaných mapových sbírek Old Maps Online (https://www.oldmapsonline.org/). Obzvláště pak u zahraničních titulů je dobré kontrolovat také databáze jako Google books, Internet Archive, Europeana aj., zda vybraný titul byl digitalizován a je veřejně či jinak dostupný, a následně zhodnotit, zda je jeho digitalizace na UK žádoucí.

která již digitalizaci takového dokumentu zpracovává, že finální digitální balíček bude pro UK dostupný k replikaci až za „příliš dlouhou“ dobu, lze se zkusit s dotyčnou institucí domluvit, že digitalizaci dokumentu provede místo ní UK. Informaci o takové domluvě také vepište do sloupce „Poznamka“, ale nyní do tabulky s hlášením digitalizace do RD (viz kapitola 2). (Jen ve vzácných a odůvodněných případech lze provést i duplicitní digitalizaci, ale nejde o vhodné řešení a chceme se ho vyvarovat.)

1.2. Kontrola, zda dokument není separátem nebo přílohou z již digitalizovaného titulu, mateřského dokumentu

Informace o separátech či přílohách a jejich vazbu na mateřský dokument lze v mnohých případech vyhledat pomocí Souborného katalogu ČR (SK ČR; https://aleph.nkp.cz/cze/skc) nebo knihovního katalogu UK (také při zobrazení zdrojového záznamu MARC 21).

Pokud vybraný dokument je separátem či přílohou z již digitalizovaného titulu, je nutné zhodnotit, zda lze žádat o replikaci tohoto mateřského dokumentu a zda je taková replikace žádoucí:

1. Pokud je mateřský dokument v cizí digitální knihovně volně přístupný nebo jako DNNT, o replikaci zpravidla nežádáme.
2. Primárním zájmem UK je kdyžtak replikace celého titulu, ne jeho konkrétní části.
3. UK však musí mít celý požadovaný titul ve svém vlastnictví, ve svém knihovním fondu (nebo prokazatelně vědět, že jej vlastnila).
4. Pokud není splněna předchozí podmínka, lze zkusit zažádat o replikaci pouze té části digitalizovaného mateřského dokumentu, která obsahuje požadovaný separát či přílohu (např. článek z periodika) nebo kterou má UK ve svém fondu (např. speciální číslo periodika).
5. Pokud není splněna podmínka v předchozím bodu (3.) a nelze žádat o replikaci separátu či přílohy (4.), může se separát digitalizovat.[2]
6. Separát či příloha se může digitalizovat také v případech, kdy se výrazně liší od exempláře v mateřském dokumentu nebo je autorsky unikátní (např. rukopisné poznámky, podpis, ex libris aj.). Unikátnost takového dokumentu posoudí dotyčná fakulta. (Tuto unikátnost separátu či přílohy, a jak se projevuje [např. rukopisnými poznámkami, podpisem, ex libris aj.], je pak záhodno doplnit do bibliografického záznamu dokumentu v systému Alma.)

[2] Digitalizovat separáty či přílohy jsme však oprávněni jen tehdy, kdy jde o „oficiální“ separáty či přílohy. Tedy, kdy např. časopis poslal autorovi výtisk jen jeho článku z konkrétního čísla časopisu. Pokud nejde o „oficiální“ separát (autor např. „vytrhl“ svůj článek z časopisu), nejsme oprávněni jej digitalizovat (ani jakkoliv jinak šířit), pokud titul stále spadá pod ochranu autorského zákona (zákon č. 121/2000 Sb.).

1.3. Kontrola DNNT (Díla nedostupná na trhu)

DNNT (https://dnnt.cz/) jsou díla, která jsou chráněna autorským právem, která však nejsou dostupná na trhu a byla zařazena do seznamu DNNT spravovaného Národní knihovnou ČR. Digitální verze těchto titulů jsou pak často dostupné online pro uživatele s univerzitním účtem prostřednictvím vzdáleného přístupu (přihlášení přes eduID / CAS UK).[3] Takto zpřístupněné dokumenty pro studenty i zaměstnance UK není v obecné rovině zapotřebí replikovat.

Ve výjimečných a odůvodněných případech však lze zažádat i o replikaci titulu, který je dostupný jako DNNT, např. v situaci, kdyby dílo doplnilo určitou tematickou sbírku v univerzitní digitální knihovně Kramerius nebo kdy je dílo zařazeno pouze do terminálového přístupu k DNNT. V takovém případě je příslušné odůvodnění potřeba uvést v tabulce s tituly k replikaci, kterou posíláte Digitalizačnímu centru UK (viz kapitola 3).

Zároveň je dobré nahlásit titul vybraný k digitalizaci, který digitalizovaný není, jako DNNT, pokud jako DNNT veden být může a ještě není (ať jej lze prohlížet třeba i prostřednictvím vzdáleného přístupu). O této možnosti informujte ÚK UK, která dotyčný titul nahlásí do DNNT. Informování probíhá ideálně pomocí tabulky s hlášením digitalizace do RD, kde do sloupce „Poznamka“ vepište žádost o nahlášení titulu do DNNT (viz kapitola 2). Zároveň však okolnost, zda se podaří titul zapsat do seznamu DNNT nebo ne, nijak nebrání hlášení titulu do RD a jeho případné digitalizaci.

1.4. Kontrola, zda případný digitalizovaný dokument splňuje standardy digitalizace na UK

Pokud zjistíte, že dokument, který chcete digitalizovat, je již digitalizovaný jinou institucí, je nutné před jeho replikací ještě zhodnotit, zda splňuje standardy digitalizace na UK (viz především tabulka v kapitole 4). Ty se obecně řídí standardy Národní digitální knihovny (NDK, https://old.ndk.cz/standardy-digitalizace). Proto digitální dokumenty v rámci NDK budou s největší pravděpodobností vždy tyto standardy splňovat (v záznamu RD je u takových titulů ve sloupci „Financováno“ uvedeno „NDK“, „IOP“ nebo „IOP-NDKU“).

Údaje o zpracování jednotlivých digitálních dokumentů lze zjistit v jejich technických metadatech, ta však často nebývají publikovaná. Pokud takové údaje nelze zjistit, je dobré obrátit se na instituci, která dokument vlastní, s žádostí o technickou specifikaci naskenovaného dokumentu (údaje o rozlišení, barevné hloubce a souborovém formátu digitálního

[3] Tedy ne zcela všechna DNNT jsou dostupná přes vzdálený přístup. Existuje část DNNT (konkrétně novodobější knihy z let 2003 až 2007), která jsou dostupná pouze na terminálech v prostorách knihoven. Univerzita Karlova nemá v současné chvíli zařízený terminálový přístup k těmto DNNT.

dokumentu, který je možné poskytnout, replikovat). Pokud instituce neposkytne technickou specifikaci digitálního dokumentu, je nutné zhodnotit jeho kvalitu pomocí lidského oka. A to, zda je text rovný a dobře čitelný včetně případných grafů, tabulek, obrázků, zda je tedy vhodný k replikaci.

Pokud již digitalizovaný dokument jinou institucí nesplňuje standardy digitalizace na UK, může fakulta či jiné pracoviště UK digitalizovat vybraný dokument znovu.

1.5. Kontrola, zda dokument není separátem nebo přílohou z titulu, který není digitalizovaný a UK jej má ve svém fondu

Informace o separátech či přílohách a jejich vazbu na mateřský dokument lze v mnohých případech vyhledat pomocí Souborného katalogu ČR (SK ČR; https://aleph.nkp.cz/cze/skc) nebo knihovního katalogu UK (také při zobrazení zdrojového záznamu MARC 21).

Pokud má UK ve svém fondu titul, mateřský dokument, z kterého separát či příloha pochází, je v zájmu UK digitalizovat tento titul, ne tedy onen separát/přílohu.

Pokud takový titul není veden jako DNNT, ale přitom by být mohl, zkusíme jej do DNNT nahlásit. O této možnosti informujte ÚK UK ideálně pomocí tabulky s hlášením digitalizace do RD (více viz podkapitola 1.3 a kapitola 2).

Pokud se titul nachází na jiné fakultě nebo pracovišti, než které chce titul digitalizovat, je zapotřebí se domluvit s tímto pracovištěm o možném vypůjčení titulu za účelem digitalizace.

1. Výpůjčka pak probíhá ideálně pomocí Meziknihovní výpůjční služby (MVS), skrze níž se výpůjčka zaznamená.
2. O výpůjčku zažádá to pracoviště, kde se bude daný titul digitalizovat. (Pokud se tedy fakulta či jiné pracoviště UK domluví s Digitalizačním centrem UK na digitalizaci v rámci prostor a služeb centra, o MVS zažádá centrum, tedy ÚK UK.)
3. V rámci hlášení digitalizace uveďte v tabulce do sloupce „Poznamka“ (viz kapitola 2), na jaké fakultě, v kterém knihovním fondu, se titul nachází.

Pokud jde o separát z periodika, lze případně digitalizovat jen konkrétní ročník nebo i jen číslo periodika. Zbylá čísla a ročníky lze do Krameria nahrávat dodatečně.

Navzdory výše řečenému lze separát či přílohu digitalizovat v těch případech, kdy se výrazně liší od exempláře v mateřském dokumentu nebo je autorsky unikátní (např. rukopisné poznámky, podpis, ex libris aj.).[4] Unikátnost takového dokumentu posoudí dotyčná fakulta. (Tuto unikátnost separátu či přílohy, a jak se projevuje [např. rukopisnými poznámkami, podpisem, ex libris aj.], je pak záhodno doplnit do bibliografického záznamu dokumentu v systému Alma.)

Po provedení kontroly vybraných dokumentů je možné přejít k následujícím dvěma krokům – k hlášení záměru digitalizace do Registru digitalizace a hlášení titulů k replikaci.

[4] Digitalizovat separáty či přílohy jsme však oprávněni jen tehdy, kdy jde o „oficiální" separáty či přílohy. Tedy, kdy např. časopis poslal autorovi výtisk jen jeho článku z konkrétního čísla časopisu. Pokud nejde o „oficiální" separát (autor např. „vytrhl" svůj článek z časopisu), nejsme oprávněni jej digitalizovat (ani jakkoliv jinak šířit), pokud titul stále spadá pod ochranu autorského zákona (zákon č. 121/2000 Sb.).

2. Hlášení do Registru digitalizace (RD)

Univerzita hlásí do RD již prvotní záměr digitalizace, tedy vybrané tituly, které chce digitalizovat. Samotná práce na digitalizaci započne až po kontrole tohoto záměru ze strany RD. Digitalizační centrum UK informuje příslušnou fakultu či pracoviště o výsledcích této kontroly.

Šablona tabulky, pomocí které se tituly hlásí do RD, je dostupná na intranetu knihoven UK (zde). Jméno šablony je ve formátu „abd162_z_RRRRMMDD_ZkrFakulty“, kde „RRRRMMDD“ označuje datum zaslání tabulky do Digitalizačního centra UK a kde „ZkrFakulty“ označuje zkratku fakulty či jiného pracoviště.

V tabulce je zapotřebí vyplnit všechny sloupce kromě posledních dvou – sloupce „URL“ a „Poznamka“. Poslední sloupec však můžete využít k připsání důležitých informací ohledně digitalizace konkrétního titulu. Šablona obsahuje poznámky a instrukce k vyplnění tabulky. Podrobnější návod je případně dostupný na stránkách RD (http://registrdigitalizace.cz/rdcz/info/data/excel):

- *DruhDok (Druh dokumentu)*: BK = kniha; SE = seriál, periodikum; MP = kartografický dokument; MU = hudebnina; RP = starý tisk, rukopis; GP = grafika.
- *CCNB (Číslo ČNB)*: Pole 015 ve zdrojovém záznamu MARC 21.
    - Zjišťuje se případně v bázi NK ČR zde: http://aleph.nkp.cz/cze/cnb
    - čČNB se přiděluje jen dokumentům, jejichž záznamy jsou v SK ČR. Pokud zde záznam chybí, je nutné tuto okolnost vepsat do sloupce „Poznamka“.
    - čČNB nemají a nemohou mít zahraniční dokumenty (uvádí se „ZL“) a nepublikované dokumenty „šedé literatury“ (uvádí se „GL“).
    - Pokud dokument nemá čČNB, ale měl by jej mít a je zapsán v SK, požádáme o přidělení čČNB (provádí ÚK UK). Tuto okolnost také vepište do poznámky k titulu.
    - Více k čČNB viz https://wiki.alma.cuni.cz/index.php?title=P%C5%99id%C4%9Blov%C3%A1n%C3%AD_%C4%8C%C3%ADsla_%C4%8Cesk%C3%A9_n%C3%A1rodn%C3%AD_bibliografie.
- *Pole001* = MMS ID (pole 001 ve zdrojovém záznamu MARC 21).
- *Autor*: Vyplnit pouze jedno jméno i v případě více autorů (ve tvaru „Příjmení, Jméno“).
- *Název*: Vyplnit i s podtitulem.
- *MistoVyd*: Místo vydání.
- *Vydavatel*.

- *RokVyd*: Rok(y) vydání. Uvádí se konkrétní rok nebo rozmezí let, které odpovídá celkovému rozmezí, kdy dotyčný titul vycházel (nejde tedy o rok vydání např. konkrétního svazku vícesvazkové monografie, ale celého titulu).
- *ISBN*: „U vícesvazkových děl bez vlastních názvů se použije ISBN souboru. Pokud jde o svazek vícesvazkové monografie s vlastním názvem (a vlastním záznamem), pak se uvádí ISBN svazku“ (Registr digitalizace, https://registrdigitalizace.cz/rdcz/info/data/excel).
- *Sigla* a *Sigla2*: Pod první siglu patří sigla ÚK UK (ABD162) jakožto správce digitalizace na UK a digitální knihovny UK Kramerius. Pod druhou siglu (nepovinný údaj) můžete dopsat siglu Vaší knihovny jakožto té, z jejíhož fondu pochází dokument, který se bude digitalizovat.

Do sloupce „Poznamka“ doplňte případné informace o titulech pro ÚK UK. Např. o jaké separáty jde, z čeho; nebo proč je nutné dokument digitalizovat i přes to, že už jako digitalizovaný existuje; zda plánujete digitalizaci celého seriálu nebo jen jeho části; zda titul lze nahlásit do DNNT; zda titul, který má být digitalizován, se nachází ve fondu jiné fakulty UK atd. Do tohoto sloupce napište i případné nedostatky, např. že titul nebyl nalezen v SK ČR; titul nemá přidělené čČNB apod.

Některé záznamy v knihovním katalogu, obzvláště pak starých publikací či separátů, nejsou úplné, nebo se objevují třeba duplicity. Je dobré v takových případech záznam v Almě obohatit o nalezené údaje či opravit. Některé vícesvazkové monografie mají zase přiděleno jen jedno čČNB (zejména, pokud jsou zpracovávány tzv. „shora“) a je zapotřebí záznam v Almě deduplikovat. Kontrolu správnosti takovýchto záznamů konzultujte případně se svými fakultními katalogizátory ještě před posláním tabulky do Digitalizačního centra UK.

Vyplněnou tabulku zašlete Digitalizačnímu centru UK (jachym.senkyrik@ruk.cuni.cz), které Vás bude informovat o výsledku hlášení záměru digitalizace do RD a zda je možné s digitalizací vybraných dokumentů začít. Následné opětovné hlášení do RD o hotové digitalizaci vybraných titulů (tedy ve chvíli, kdy jsou tituly nahrané v Krameriovi) již plně zajišťuje Digitalizační centrum UK.

Pokud náhodou digitalizace vybraných dokumentů již probíhá nebo je již hotová (bez nahlášení původního záměru digitalizace), informujte o tom Digitalizační centrum UK (jachym.senkyrik@ruk.cuni.cz), rádi bychom ale hlásili právě už ten prvotní záměr digitalizace.

3. Hlášení titulů k replikaci

Replikovat lze jen ty tituly, které má UK ve svém vlastnictví, ve svém knihovním fondu (nebo o kterých UK prokazatelně ví, že je vlastnila). Zároveň není v obecném zájmu univerzity replikovat tituly volně dostupné v jiné digitální knihovně nebo vedené jako DNNT, k nimž se uživatelé UK dostanou i prostřednictvím vzdáleného přístupu. Mohou se však vyskytnout výjimky a i např. replikace DNNT může být v jistých případech žádoucí (viz podkapitola 1.3).

Tituly určené k replikaci se hlásí přímo Digitalizačnímu centru UK prostřednictvím vyplněné tabulky, jejíž šablona je dostupná na intranetu knihoven UK (zde). Jméno šablony je ve formátu „sablona-replikace--interni_RRRRMMDD_ZkrFakulty“, kde „RRRRMMDD“ označuje datum zaslání tabulky do Digitalizačního centra UK a kde „ZkrFakulty“ označuje zkratku fakulty či jiného pracoviště.

V tabulce se vyplňují veškeré sloupce až po „URL“:

- *Vlastník*: Instituce, která vlastní digitální dokument, u které tedy lze žádat o replikaci.
- *Požadované ročníky/čísla*: Zpravidla není potřeba vyplňovat, pokud UK vlastní celý požadovaný titul. UK má totiž zájem o replikaci celého titulu. Pokud lze žádat jen o určité ročníky nebo čísla, protože UK nevlastní (ani prokazatelně nevlastnila) celý titul, je nutné tuto skutečnost zde napsat, tedy o které ročníky/čísla lze zažádat k replikaci.
- *UUID*, *Úroveň popisu UUID*, *URL*: údaje digitálního dokumentu, který chceme replikovat.
    - *UUID*: identifikátor ve tvaru „uuid:vvvvvvvv-wwww-xxxx-yyyy-zzzzzzzzzzzz“, který jednoznačně označuje dokument určený k replikaci. Lze ho najít v URL odkaze na daný dokument, případně v metadatech dokumentu.
        - Př. (URL): https://kramerius5.nkp.cz/uuid/uuid:4d324f80-bcdb-11e4-b2e2-005056827e52
        - Př. (metadata): <mods:identifier type="uuid">4d324f80-bcdb-11e4-b2e2-005056827e52</mods:identifier>; <dc:identifier>uuid:4d324f80-bcdb-11e4-b2e2-005056827e52</dc:identifier>
    - *Úroveň popisu UUID*: Vlastní unikátní identifikátor UUID má jak titul, tak také jednotlivé ročníky nebo čísla periodika, dokonce i jednotlivé stránky. V tomto sloupci tak vyplňte, zda do tabulky zapsané UUID odkazuje právě na celý titul (napište „**titul**“), ročník (napište „**ročník**“) nebo číslo (napište

„**číslo**“), či na svazek vícesvazkové monografie (napište „**svazek**“). V zájmu UK je žádat o replikaci celého titulu, pokud je to možné (pokud jej tedy má UK ve svém knihovním fondu).

- *URL*: Internetový odkaz na příslušný dokument v digitální knihovně instituce, kterou budeme žádat o replikaci. Např.: https://krameriu5.nkp.cz/uuid/uuid:vvvvvvvv-wwww-xxxx-yyyy-zzzzzzzzzzzz

Do sloupce „Poznámky“ můžete doplnit další údaje, které s replikací konkrétních titulů souvisejí (např. že digitalizace dotyčného titulu se teprve zpracovává apod.).

Vyplněnou tabulku zašlete Digitalizačnímu centru UK (jachym.senkyrik@ruk.cuni.cz), které Vás bude informovat o výsledcích replikace.

UK má v současné chvíli uzavřené replikační smlouvy s NK ČR, MZK a KNAV (disponujících největším počtem digitalizovaných dokumentů). To však neznamená, že UK nechce nebo nemůže uzavřít takovou smlouvu i s jinými knihovnami a institucemi. Pokud však budeme chtít replikovat titul od instituce, s níž zatím uzavřenou smlouvu nemáme, může se celý proces protáhnout. Nejprve je totiž zapotřebí uzavřít onu replikační smlouvu (zajišťuje ÚK UK).

4. Skenování

Samotné skenování vybraných dokumentů nastává až ve chvíli, kdy jsou tyto dokumenty zdárně zapsány do RD. O tomto zapsání informuje fakulty a pracoviště UK Digitalizační centrum UK. Pro skenování je pak zapotřebí vybrat ty nejlepší možné, nejzachovalejší exempláře, ideálně bez vpisků (tužkou vepsané poznámky můžete případně vygumovat).[5] Zároveň je dobré myslet i na to, jak dobře se bude daný exemplář skenovat. Pokud hrozí např. zkreslení skenů u hřbetu knihy, stojí za zvážení povolit vazbu knihy pro skenování, pokud je to technicky možné.

Fakulta nebo jiné pracoviště UK se rozhodne, zda bude titul skenovat na svém pracovišti nebo zda jej přepraví ke skenování do Digitalizačního centra UK:

- V současné chvíli nemá Digitalizační centrum UK personální kapacitu k tomu, aby mohlo naplňovat pravidelnou poptávku po skenování. Zároveň však nabízí jednotlivým digitalizačním pracovištím UK své vybavení za účelem jejich vlastní digitalizace (informace o vybavení lze najít zde: https://digitalizace.cuni.cz/DKU-86.html). Nabízejí se tak především dva způsoby, jak lze využít skenery v ÚK UK k digitalizaci vybraných dokumentů:
    - První varianta: Fakulta či jiné pracoviště UK vyšle, po předchozí domluvě s Digitalizačním centrem UK, svého zástupce do ÚK UK, kde bude tento zástupce zaškolen (bez poplatku) a sám pak bude schopen naskenovat vybrané tituly pomocí vybavení Digitalizačního centra UK. Pracovník Digitalizačního centra UK bude tomuto zástupci k dispozici při řešení technických či jiných problémů, které v průběhu skenování nastanou.
    - Druhá varianta: Po předchozí domluvě lze dopravit jednotlivé dokumenty do ÚK UK, kde je pracovník Digitalizačního centra UK naskenuje. Za tuto službu si Digitalizační centrum účtuje poplatek podle daného ceníku (zde: https://digitalizace.cuni.cz/DKU-99.html).

Skenování musí splňovat následující standardy digitalizace na UK, které vycházejí ze standardů Národní digitální knihovny:

- Formát jednotlivých souborů (jednotlivých naskenovaných stran): **TIFF** (bez komprese nebo s bezztrátovou kompresí LZW).
- Parametry skenování, viz tabulka:

[5] V případě potřeby zvažte i možnost restaurování či konzervaci dokumentů.

Typ publikace	Rozlišení skenu	Barevná hloubka skenu
čistě textová publikace • Č/B | odstíny šedi	300 DPI	alespoň 8bitová hloubka odstínů šedi[6] // 24bitová
textová publikace obsahující tabulky, grafy • Č/B | odstíny šedi	300/400 DPI[7]	alespoň 8bitová hloubka odstínů šedi // 24bitová
Textová publikace obsahující tabulky, grafy • Barevně	300/400 DPI	24bitová
Převážně obrazová publikace • Č/B | odstíny šedi	400 DPI	alespoň 8bitová hloubka odstínů šedi // 24bitová
Převážně obrazová publikace • Barevně	400 DPI	24bitová

- Obálku, přední a zadní desky vždy skenujte barevně (ke skenování obálky srov. první a poslední obrázky této knihy https://ndk.cz/uuid/uuid:d1df9c80-d254-11ea-9c41-005056827e52).
- Pokud kniha obsahuje poznámky pod čarou, které jsou psány velmi malým písmem a kniha spadá do kategorie skenování do 300 DPI, je zapotřebí během skenování zkontrolovat, zda toto písmo bude ve výsledném obrazu stále dobře čitelné, případně zvyšte rozlišení skenů.
- Grafiky či mapy doporučujeme skenovat ve větším rozlišení – minimálně na 400 DPI, zvažte však i 600, případně i 800 DPI, pokud skener disponuje takovým optickým rozlišením. Jde o to, aby i ty nejmenší grafické detaily dokumentu byly reprezentovány alespoň několika body, aby tloušťka čar byla zachycena alespoň dvěma body.[8]
- Po skončení skenování je nutné zkontrolovat kvalitu naskenovaných stránek – zda jsou všechny, zda nevznikly duplikáty, zda jsou správně ořezány, zda je text na stranách rovně (aby tak mohlo zdárně proběhnout OCR zpracování) apod. Ořez a

[6] Je možné takovýto typ publikace skenovat také v 24bitové barevné hloubce (platí i pro ostatní 8bitové zmínky v tabulce), pokud takovou možnost digitalizační pracoviště preferuje.
[7] Pokud kniha obsahuje malé a detailní grafy či tabulky, u kterých by mohla při skenování na 300 DPI hrozit ztráta čitelnosti, skenujte dokument radši na 400 DPI. Pokud si nejste jisti, jakou hodnotu rozlišení zvolit, vyzkoušejte obě varianty, zkontrolujte je a rozhodněte se, zda je rozlišení 300 DPI dostatečné či ne.
[8] K tomu srov. Petr Žabička, „Metodika pro on-line zpřístupňování starých map a dalších grafických dokumentů pro paměťové instituce“, MZK, 2011, dostupné na: https://www.mzk.cz/sites/mzk.cz/files/metodika_pro_online_zpristupnovani_starych_map__1.pdf [16. 9. 2022].

narovnání stránek lze případně zajistit i jinými programy než jen v rámci nástrojů, které nabízejí softwary konkrétních skenerů.

Všechny vytvořené soubory je nutné pojmenovat podle jednotné názvové konvence. Takto je třeba mít pojmenovanou i složku, ve které se soubory nacházejí. Názvy se řídí následujícím formátem: „**MMSID_Nazev_Autor_ZkrFakulty**“. V názvech nepoužívejte háčky, čárky, jinou diakritiku, interpunkci nebo další speciální symboly. Místo mezer používejte spojovník (-). Návod na hromadné přejmenování souborů pomocí programu IrfanView je dostupný na intranetu knihoven UK (zde).

4.1. Postprocesy – ScanTailor a narovnání, ořez i jiné možné úpravy

Digitalizační centrum UK doporučuje pro úpravu jednotlivých skenů program ScanTailor (freeware, viz https://scantailor.org/). Program dokáže automaticky narovnat, oříznout a celkově sjednotit výsledné skeny v rámci jedné dávky, jednoho naskenovaného dokumentu.

Výsledné úpravy lze také exportovat do 1bitové hloubky (Č/B). Takto je však možné exportovat, aniž by se tím znehodnotila kvalita digitalizovaného dokumentu, pouze čistě textové publikace (a vyjmout z takového exportu obálku, přední a zadní desky, případně jiné části dokumentu, které by tímto převodem ztratily na kvalitě a čitelnosti).

Návod k programu, zpracovaný Digitalizační centrem UK, je dostupný na intranetu knihoven UK (zde).

* * *

Jako poslední krok před přesouváním a dalším zpracováním skenů je nutné vygenerovat kontrolní součty MD5, které ověřují úplnost jednotlivých obrazových souborů a zda nedošlo k jejich poškození. Návod na vygenerování takovýchto kontrolních součtů pomocí programu Total Commander je dostupný na intranetu knihoven UK (zde).

Digitalizační centrum UK může zajistit, na přání konkrétního digitalizačního pracoviště, školení k procesům skenování.[9]

[9] Veškerá školení Digitalizačního centra v rámci UK jsou bez poplatku.

5. OCR zpracování

Naskenované dokumenty je zapotřebí opatřit také textovými soubory pomocí OCR zpracování (optické rozpoznávání znaků). Tyto soubory, s rozpoznaným textem dokumentu, umožňují čtenáři právě s tímto textem dále pracovat (např. text kopírovat, vyhledávat apod.). OCR zpracování se vyžaduje v podobě dvou souborů pro každou stránku (pro každý TIFF), a to ve formátu TXT (obsahuje text jednotlivých stránek) a ALTO XML (lokalizuje, kde se rozpoznaný text na stránce nachází).

OCR zpracování naskenovaného dokumentu lze zajistit buď lokálně, nebo pomocí celouniverzitního automatického serverového zpracování:

- **Lokální zpracování**:
    - Pokud bude fakulta či jiné pracoviště UK chtít naskenované dokumenty opatřit OCR zpracováním lokálně, informujte o tom Digitalizační centrum UK (jachym.senkyrik@ruk.cuni.cz), které vám na serveru vytvoří speciální adresář, kam budete takto zpracované dokumenty nahrávat, kvůli jejich dalšímu zpracování v programu ProArc.
    - Existuje několik programů, ať už komerčních nebo volně dostupných, které dokáží rozpoznat text v naskenovaných dokumentech a vygenerovat textové soubory pro jednotlivé stránky. Bohužel však ty nejběžnější a nejpoužívanější programy neumí generovat ALTO XML, které se při OCR zpracování dokumentů, jež mají být zveřejněny v digitální knihovně Krameriius, vyžaduje.
    - Poté, co program rozpozná text naskenovaných stránek, je zapotřebí rozpoznaný text překontrolovat, především ta problematická místa, která program sám označí.
    - Textové i ALTO XML soubory uložte pro každý TIFF soubor zvlášť (program by měl takovou možnost sám nabízet, ať už pomocí zaškrtávacích polí, rozbalovacího menu během ukládání souborů, nebo jinak).
    - Veškeré takto připravené soubory v jedné složce nahrajte do určeného adresáře na server, odkud Digitalizační centrum UK zajistí jejich přesun do vstupního adresáře pro program ProArc.
    - Návod na uložení digitalizace na server je dostupný na intranetu knihoven UK (zde).
    - Ve chvíli, kdy úspěšně nahrajete dokumenty do složky na serveru, informujte o tom Digitalizační centrum UK (jachym.senkyrik@ruk.cuni.cz).

- **Serverové zpracování**:
    - UK disponuje serverovým automatickým OCR zpracováním (od společnosti ABBYY). Pro využití této možnosti musí mít digitalizační pracoviště zřízený vlastní uživatelský účet pro přístup k vlastní importní složce na serveru, kam naskenované tituly bude nahrávat. Pro zřízení takového účtu kontaktujte Digitalizační centrum UK (jachym.senkyrik@ruk.cuni.cz).
    - Pro každé digitalizační pracoviště lze po domluvě nastavit, s jakými jazykovými slovníky má program při OCR zpracování dokumentů pracovat, tedy jej připravit na to, v jakých jazycích budou nahrávané dokumenty.
        - Specifikace a seznam jazyků, které program umí rozpoznat, viz: https://support.abbyy.com/hc/en-us/articles/360007070419-FineReader-Server-14-Specifications. Přesto ne všechny uvedené jazyky je schopná licence, kterou UK disponuje, zajistit. Především staré typy písma, jakými jsou švabach, fraktura, gotické písmo, nebo ručně psané dokumenty či některé jazyky psané jiným typem písma než latinkou (např. arabština, fársí, japonština atd.) není schopné serverové OCR zpracování na UK zajistit (ale třeba hebrejské písmo nebo cyrilici zvládne).[10] Většinu ze seznamu jazyků umí licence na UK zpracovat, konkrétní nastavení již diskutujte s Digitalizačním centrem UK.
    - Licence pro serverové zpracování, kterou UK disponuje, je omezena jednak časem (pořizuje se vždy na rok), jednak počtem zpracovaných stránek za rok. To znamená, že je zapotřebí ze strany každého pracoviště důsledně kontrolovat, aby se do importních složek nahrávaly opravdu dobře připravené dokumenty, aby nedocházelo třeba k duplicitám, a tak nesmyslnému čerpání omezeného množství OCR zpracovaných stránek. Digitalizační centrum UK bude kvůli omezenému počtu stránek průběžně kontrolovat, co jednotlivá pracoviště plánují digitalizovat (pomocí tabulek s hlášením digitalizace do RD, viz kapitola 2), a bude s fakultami komunikovat, pokud by bylo zapotřebí práci na digitalizaci rozvrhnout do delšího časového období. Taková situace může nastat např. ve chvíli, kdyby fakulta, či jiné pracoviště UK, chtěla digitalizovat seriál obsahující velký počet ročníků a čísel.
    - Návod na uložení digitalizace na server je dostupný na intranetu knihoven UK (zde).

[10] Pokud byste potřebovali digitalizovat právě dokumenty s nějakým takto „speciálním“ typem písma, lze např. zakoupit speciální licenci, díky které OCR server zvládne i takové dokumenty zpracovat. Existují však i jiná řešení. Jaké jsou možnosti digitalizace jiných typů písma, lze diskutovat s Digitalizačním centrem UK. Rádi se s Vámi domluvíme, jak v jednotlivých případech nastavit OCR zpracování tak, aby co nejlépe vyhovovalo konkrétnímu digitalizačnímu projektu.

- Ve chvíli, kdy úspěšně nahrajete dokumenty do importní složky na serveru, informujte o tom Digitalizační centrum UK (jachym.senkyrik@ruk.cuni.cz).

5.1. Projekt PERO[11]

Pro OCR zpracování ručně psaných dokumentů doporučuje Digitalizační centrum UK použít program PERO OCR, dostupný přes webový prohlížeč na stránkách Vysokého učení technického v Brně (https://pero.fit.vutbr.cz/).[12] K používání programu je zapotřebí si vytvořit jen vlastní účet.

Každé jednotlivé zpracování se pak skládá z několika kroků:

1. pojmenování dokumentu;
2. nahrání skenů (je potřeba nahrát JPEG soubory nepřevyšující velikost 8 MB, vhodná je i nižší velikost, dokud výrazně nezkresluje text na obrázku);
3. layout analysis (analýza rozložení textu na stranách) – výsledek je možné ručně upravit;
4. OCR zpracování – výsledek je možné ručně opravit.[13]

V sekci „Help" lze nalézt podrobnější video návody, jak program používat. Lehkou nevýhodou programu je pak nemožnost jednoduchého dávkového stažení ALTO XML souborů. Abyste nemuseli stahovat tyto soubory po jednom, po každé stránce, je možné soubory stáhnout i dávkově, a to pomocí skriptu od vývojářů (https://github.com/DCGM/pero_ocr_web/tree/master/user_scripts). Pokud přidáte Digitalizační centrum UK (jachym.senkyrik@ruk.cuni.cz) jako spolupracovníka k Vašemu dokumentu, může Digitalizační centrum již zajistit stažení všech potřebných textových souborů.

* * *

Digitalizační centrum UK může zajistit, na přání konkrétního digitalizačního pracoviště, školení k OCR zpracování.[14]

[11] Vývoj projektu vede Michal Hradiš z FIT VUT v Brně a Petr Žabička z MZK.
[12] Program však umí dobře rozpoznávat i text psaný latinkou, kurentem, švabachem či jinými typy písma.
[13] Abyste zjistili stav posledních dvou kroků (zda je už krok dokončen či ne), je u každého zapotřebí opětovně načíst hlavní stránku programu (pomocí F5).
[14] Veškerá školení Digitalizačního centra v rámci UK jsou bez poplatku.

6. ProArc

ProArc je program umožňující metadatový popis digitálních dokumentů, jejich správu a přípravu jak pro nahrání do digitální knihovny UK Kramerius, tak pro jejich archivaci dle standardů Národní digitální knihovny.

Digitalizační centrum UK zajistí po proběhlém OCR zpracování přesun veškerých potřebných souborů na serveru do vstupního adresáře, odkud lze soubory natáhnout do programu ProArc. Do tohoto programu se vstupuje přes webové rozhraní (https://proarc.cuni.cz/proarc), kde je zapotřebí mít založený vlastní uživatelský účet. Pro jeho založení kontaktujte Digitalizační centrum UK (jonas.kvet@ruk.cuni.cz) s informací o technických specifikacích skeneru nebo skenerů, které jsou a budou konkrétním digitalizačním pracovištěm používány pro skenování dokumentů.

Zpracování digitálních dokumentů v programu ProArc si zajišťují fakulty samy (dá se případně požádat Digitalizační centrum UK o toto zpracování, ale služba je zpoplatněna dle aktuálního ceníku, viz https://digitalizace.cuni.cz/DKU-99.html). Návody, jak postupovat při zpracování digitálního dokumentu v programu ProArc, jsou dostupné na intranetu knihoven UK (viz návod k monografiím a periodikům). Po úspěšném exportu digitálních dokumentů v programu ProArc informujte Digitalizační centrum UK (jonas.kvet@ruk.cuni.cz), které tituly jste exportovali, včetně jejich UUID a URN:NBN. Digitalizační centrum již zajistí jejich nahrání do digitální knihovny UK Kramerius a nahlásí hotovou digitalizaci do Registru digitalizace.

Digitalizační centrum UK může zajistit, na přání konkrétního digitalizačního pracoviště, školení k programu ProArc.[15]

[15] Veškerá školení Digitalizačního centra v rámci UK jsou bez poplatku.

7. Kramerius

Kramerius je program pro zpřístupnění digitálních dokumentů, jehož vývoj ve spolupráci s dalšími knihovnami a firmami koordinuje Knihovna Akademie věd ČR. UK využívá tento program pro svou digitální knihovnu (viz https://kramerius.cuni.cz), pro zprostředkování digitalizovaných dokumentů z fondů knihoven UK. Zpřístupňování dokumentů v Krameriovi se řídí Autorským zákonem, proto novější tituly nelze zobrazit skrze vzdálený přístup, ale jsou uzamčené a dostupné jen z konkrétních terminálů v knihovnách UK.[16] Pokud máte zájem o zprovoznění takového terminálu v prostorách Vaší knihovny, obraťte se na Digitalizační centrum UK (jonas.kvet@ruk.cuni.cz).

Kompletně zpracované digitální balíčky z předchozích kroků, obsahující tedy jednotlivé skeny (TIFF), kontrolní součty (MD5), OCR zpracování (TXT a ALTO XML) a metadatový popis (skrze ProArc), jsou připravené jednak pro archivaci, jednak pro nahrání do digitální knihovny Kramerius. Nahrání zajišťuje Digitalizační centrum UK, po předchozím upozornění od fakulty či jiného pracoviště UK. Digitalizační centrum následně informuje příslušné fakulty o úspěšném nahrání a také o nahlášení hotové digitalizace do Registru digitalizace.

Informace o dostupnosti digitalizovaného dokumentu v Krameriovi se již automaticky propíše do univerzitního katalogu, Almy.

V Krameriovi si fakulty a jiná pracoviště UK mohou nechat vytvořit tematické sbírky, tedy prostor, kde budou shromážděny jen určité dokumenty nahrané v Krameriovi. Pro zřízení takové sbírky kontaktujte Digitalizační centrum UK (jonas.kvet@ruk.cuni.cz).

V zájmu UK je pak ve své digitální knihovně zveřejňovat především historické dokumenty, DNNT a nedostatkovou studijní literaturu. V současné chvíli tvoří obsah Krameria UK monografie, periodika, grafiky, mapy a další dokumenty v rámci určitých tematických sbírek. Více ke koncepci digitální knihovny Kramerius UK lze najít v samostatném dokumentu zde.

[16] V současné chvíli nefunguje v Krameriovi UK licence pro zpřístupnění DNNT. Její zprovoznění však patří k prioritám ÚK UK, až to bude technicky možné.

Průběh digitalizace

1.
Dokument k digitalizaci
- kontrola pomocí **RD** aj., zda dokument již není digitalizovaný

2.
Hlášení záměru digitalizace do RD
- zaslání vyplněné **tabulky** do Digitalizačního centra UK

3.
Replikace
- zaslání vyplněné **tabulky** do Digitalizačního centra UK

již digitalizovaný

nová digitalizace

4.
Skenování
- lokálně
- v Digitalizačním centru UK
  A. vyslat pracovníka do centra
  B. digitalizuje centrum (**ceník**)
- TIFF pro každou stránku
- 300/400 DPI
- 24bit barva / 8bit stupně šedi

4.1.
Postprocesy
- kontrola skenů - narovnání, ořez aj.
- **pojmenování souborů**
- **vygenerovat md5**

5.
OCR
- lokálně - např. pomocí **PERO OCR**
- **serverově** - doporučené

6.
ProArc
- **návod - monografie**
- **návod - periodikum**

7. Kramerius

UNIVERZITA KARLOVA
Digitalizační centrum