



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



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 2.0

2. 6. 2025

Digitalizační centrum UK
Oddělení repozitářů a digitalizace



UNIVERZITA KARLOVA
Ústřední knihovna

Obsah

Úvod	3
1. Příprava digitalizace	4
2. Hlášení do Registru digitalizace (RD)	6
3. Hlášení titulů k replikaci	7
4. Skenování	9
5. OCR zpracování	11
6. ProArc	14
7. Kramerius	15
8. ARCLib	16

Úvod

Tato metodika slouží pro interní účely Univerzity Karlovy (UK). Definuje standardy a popisuje procesy digitalizace tištěných dokumentů na univerzitě v rámci osmi dílčích kapitol. Standardy digitalizace přitom vycházejí ze standardů definovaných Národní digitální knihovnou (již zaštiťuje Národní knihovna ČR) a chtějí tak dosáhnout 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 dalším prohlubování a vylepšování nastavených procesů a digitalizačního workflow jako celku.

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 ([zde](#)) 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, 2. 6. 2025

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

Kontrolujeme také, zda vybraný dokument není třeba separátem nebo přílohou mateřského dokumentu, který již je digitalizovaný (informace o separátech či přílohách 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).

1.2. Dokument je digitalizovaný

Pokud však kvalitou neodpovídá standardům Národní digitální knihovny, lze jej opětovně digitalizovat. Nahlásíme jej do Registru digitalizace (viz kapitola 2.).

Pokud kvalitou odpovídá NDK standardům a fakulta či jiné pracoviště UK jej chce mít v univerzitní digitální knihovně (Kramerius), požádáme o jeho replikaci (viz kapitola 3.).²

Pokud jde o veřejně dostupné dílo (nepodléhá již autorsko-právní ochraně), nebudeme jej replikovat.³ Stejně tak nebudeme replikovat dokumenty dostupné v režimu DNNT (Díla

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/database.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.

² Replikovat můžeme jen ty tituly, které má UK ve svém knihovním fondu (nebo prokazatelně víme, že jej UK vlastní).

³ Čtenář se k takovému dokumentu dostane ideálně přes portál České digitální knihovny, <https://ceskadigitalniknihovna.cz/>.

nedostupná na trhu).⁴ Výjimku mohou tvořit dokumenty, pro které je pádný důvod k jejich duplikaci v univerzitní digitální knihovně (např. kvůli doplnění určité tematické sbírky nebo periodika atd.).

1.3. Dokument není digitalizovaný

Avšak jeho digitalizaci plánuje nebo už zpracovává jiná instituce (podle záznamu v RD), počkáme až jej dotyčná instituce digitalizuje.⁵

Pokud dokument je separátem nebo přílohou mateřského dokumentu, který není digitalizovaný, budeme digitalizovat mateřský dokument, pokud jej má UK ve svém knihovním fondu.⁶ Pokud mateřský dokument není ve fondu UK, lze separát digitalizovat.⁷ Separát či příloha se může digitalizovat také v případech, kdy se výrazně liší od exempláře v mateřském dokumentu nebo je autorsky unikátní (např. díky rukopisným poznámkám, podpisem, ex libris aj.). Unikátnost takového dokumentu posoudí dotyčná fakulta, která chce dokument digitalizovat. (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.)

Pokud vybraný dokument k digitalizaci není veden na seznamu DNNT (<https://sdnnt.nkp.cz/>), ale mohl by být, je vhodné jej na seznam nahlásit. Případně oznamte tuto informaci Ústřední knihovně UK, která dotyčný titul nahlásí. Nicméně, informace, zda je titul veden jako DNNT nebo ne, nijak nebrání jeho samotné digitalizaci.

Exkurz: Standardy NDK

Univerzita Karlova se při digitalizaci řídí standardy NDK (<https://standardy.ndk.cz/ndk/standardy-digitalizace/>), díky kterým je zajištěna interoperabilita mezi digitálními knihovnami na národní úrovni (například v podobě replikací). Díky těmto standardům také může fungovat Česká digitální knihovna (<https://ceskadigitalniknihovna.cz/>), do které se chce Univerzita Karlova zapojit. Standardy NDK řeší jak kvalitu naskenovaných dokumentů (rozlišení, barevnou hloubku, souborový formát atd.), tak jejich metadatový popis a strukturu digitálních objektů.

⁴ DNNT (<https://dnnt.cz/>) jsou díla, která jsou chráněna autorským právem, která však nejsou dostupná na trhu a byla zařazena do seznamu DNNT (<https://sdnnt.nkp.cz/>) 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).

⁵ Může však nastat situace, kdy pro pedagogické, studijní či vědecké účely je důležité digitalizovat vybraný dokument v co nejrychlejší možné lhůtě. V takové chvíli kontaktujte tu instituci, která dokument digitalizuje, abyste se domluvili, jak dlouho jim bude digitalizace trvat, případně zda by bylo možné, aby digitalizaci provedla UK.

⁶ Pokud se titul nachází na jiné fakultě či pracovišti, než které chce titul digitalizovat, je zapotřebí se domluvit s tímto pracovištěm na možném vypůjčení titulu za účelem jeho digitalizace.

⁷ 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 začíná až po kontrole tohoto záměru ze strany RD. Digitalizační centrum UK informuje příslušnou fakultu či pracoviště o výsledcích této kontroly.

Vybrané tituly k digitalizaci nahlaste přes univerzitní servicedesk (<https://service-desk.cuni.cz>) pomocí formuláře [Digitalizace a tisk pro fakulty a součásti UK](#). Stačí přiložit nebo vypsát seznam MMS ID vybraných titulů (pole 001).

Ve formuláři můžete případně doplnit informace o titulech, jako např. proč je nutné dokument digitalizovat i přes to, že už jako digitalizovaný existuje, zda plánujete digitalizaci celého časopisu nebo jen jeho části 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 hlášením záměru digitalizace.

Následné opětovné hlášení do RD o hotové digitalizaci vybraných titulů (tedy ve chvíli, kdy jsou tituly již nahrané v univerzitní digitální knihovně Kramerius) 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 (také prostřednictvím servicedesku – [Digitalizace a tisk pro fakulty a součásti UK](#)), 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 prostřednictvím vzdáleného přístupu. Mohou se však vyskytnout výjimky, a tak lze i takové dokumenty replikovat, pokud pro to existuje pádný důvod (např. kvůli doplnění určité tematické sbírky nebo periodika atd.).

Tituly určené k replikaci se hlásí skrze univerzitní servicedesk (<https://service-desk.cuni.cz>) pomocí formuláře [Digitalizace a tisk pro fakulty a součásti UK](#). K tomuto hlášení je nutné přiložit tabulku s vypsáním titulu k replikaci. Šablona tabulky je dostupná na intranetu knihoven UK ([zde](#)). Jméno šablony je ve formátu „šablona-replikace--interni_RRRRMMDD_ZkrFakulty“, kde „RRRRMMDD“ označuje datum vytvoření tabulky 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-zzzzzzzzzzz“, 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://kramerius5.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.).

Digitalizační centrum 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 fakulta 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).⁸ 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:

- Formát jednotlivých souborů (jednotlivých naskenovaných stran): **TIFF** (bez komprese nebo s bezztrátovou kompresí LZW).
- Barevná hloubka: **24bitová hloubka** (plně barevný sken)
- Rozlišení: minimálně **300 DPI**. Lze zvážit i vyšší rozlišení (400 DPI) především ve chvílích, pokud dokument obsahuje ilustrace, fotografie, detailní grafy, pokud jde o rukopis apod.⁹

⁸ V případě potřeby zvažte i možnost restaurování či konzervaci dokumentů.

⁹ V případě nejistoty, zda je rozlišení 300 DPI dostatečné, vyzkoušejte různé varianty (alespoň 300 a 400 DPI) a zhodnoťte, jaké rozlišení je pro dobrou čitelnost skenů vyhovující.

Obálku knihy skenujeme také celou samostatně, viz první a poslední obrázky této knihy <https://ndk.cz/uuid/uuid:d1df9c80-d254-11ea-9c41-005056827e52>.

Grafiky či mapy doporučujeme skenovat ve vyšším 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.¹⁰

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 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, v 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 použijte 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. Návod k programu, zpracovaný Digitalizačním centrem UK, je dostupný na intranetu knihoven UK ([zde](#)).

* * *

Digitalizační centrum UK může zajistit, na přání konkrétního digitalizačního pracoviště, školení k procesům skenování.¹¹

¹⁰ 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_on-line_zpristupnovani_starých_map_1.pdf [16. 9. 2022].

¹¹ 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 rozpoznáním 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 rozpoznáný 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 skrze univerzitní servicedesk UK (<https://servicedesk.cuni.cz>) pomocí formuláře [Digitalizace a tisk pro fakulty a součásti UK](#). Digitalizační centrum vám vytvoří na serveru 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 Kramerijs, vyžaduje.¹²
- Poté, co program rozpozná text naskenovaných stránek, je dobré rozpoznáný text přezkontrolovat, 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 zaškrtačkových 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 opět přes servicedesk UK.

¹² Jaké jsou možnosti lokálního OCR zpracování, lze konzultovat s Digitalizačním centrem UK.

- **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 skrze servicedesk UK (<https://servicedesk.cuni.cz>) pomocí formuláře [Digitalizace a tisk pro fakulty a součásti UK](#).
- 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-Fine-Reader-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 latinou (např. arabština, farsí, japonština atd.) není schopné serverové OCR zpracování na UK zajistit (ale třeba hebrejské písmo nebo cyrilici zvládne).¹³ Většinu ze seznamu jazyků však 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 hlášení 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](#)).

¹³ 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.

5.1. Projekt PERO¹⁴

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/>).¹⁵ 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 nezkruskuje text na obrázku);
3. layout analysis (analýza rozložení textu na stranách) – výsledek je možné ručně upravit;
4. OCR zpracování – výsledek je možné ručně opravit.¹⁶

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 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.¹⁷

¹⁴ Hlavním řešitelem projektu, financovaného z NAKI II, 2018–2022, byla FIT VUT v Brně a MZK jako spoluřešitel.

¹⁵ Program však umí dobře rozpoznávat i text psaný latinkou, kurentem, švabachem či jinými typy písma.

¹⁶ Abyste zjistili stav posledních dvou kroků (zda je už krok dokončen či ne), je u každého zapotřebí opětovně načíst hlavní stránku programu (např. pomocí klávesy F5).

¹⁷ 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 v systému ARCLib.

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://pro-arc.cuni.cz/proarc>), kde je zapotřebí mít založený vlastní uživatelský účet. Pro jeho založení kontaktujte Digitalizační centrum UK skrze servicedesk UK (<https://servicedesk.cuni.cz>) pomocí formuláře [Digitalizace a tisk pro fakulty a součásti UK](#). K požadavku přidejte informaci 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 jednotlivá pracoviště sama (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>). Návod, 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 zajistí Digitalizační centrum UK jejich nahrání do digitální knihovny UK Kramerius, jejich archivaci v programu ARCLib a nahlásí hotovou digitalizaci do Registru digitalizace.

* * *

Digitalizační centrum UK může zajistit, na přání konkrétního digitalizačního pracoviště, školení k programu ProArc.¹⁸

¹⁸ Veškerá školení Digitalizačního centra v rámci UK jsou bez poplatku.

7. Kramerius

Kramerius je program pro zpřístupnění digitálních dokumentů, který v ČR využívá velké množství knihoven i jiných institucí (např. Národní knihovna ČR, Moravská zemská knihovna, Knihovna AV ČR aj.). 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.¹⁹ Pokud máte zájem o zprovoznění takového terminálu v prostorách Vaší knihovny, obraťte se na Digitalizační centrum UK skrze servicedesk UK (<https://servicedesk.cuni.cz>) pomocí formuláře [Digitální repozitáře a knihovny](#).

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.

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 opět skrze servicedesk UK.

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](#).

¹⁹ V současné chvíli nefunguje v Krameriovi UK licence pro zpřístupnění DNNT. Její zprovoznění však patří k prioritám ÚK UK, až to bude technicky možné.

8. ARCLib

ARCLib je program určený pro dlouhodobou archivaci digitálních dokumentů umožňující bitovou i logickou ochranu dat, na základě modelu OAIS. Program kontroluje integritu uložených dokumentů, které máme v současné době uložené na dvou místech v republice. Digitalizační centrum UK v ARCLibu archivuje jak dokumenty, které se nově digitalizovaly a nahrály do digitální knihovny Kramerius, tak postupně archivuje i ty, jejichž digitalizace proběhla dříve, před zavedením ARCLibu do digitalizačního workflow UK (podzim 2024).