Výhody digitalizácie

Digitalizácia sa stala dôležitým nástrojom na zabezpečenie a prístup k informatizáciám, poznatkom a ku kultúrnym hodnotám.

Medzi hlavné výhody digitálnych služieb patria:

* Dostupnosť digitálnych dokumentov bez ohľadu na čas a miesto
* Jednoduchšie vyhľadávania vďaka navigačným pomôckam a plnému textu
* Ochrana originálov a uchovávanie a zabezpečenie digitálnych kópií dokumentov a predmetov

Táto metodika sa zaoberá jednotlivými projektovými krokmi a pracovnými postupmi digitalizácie od koncepčnej fázy cez spracovanie dokumentov až po prezentáciu a využitie elektronickej služby. Pritom často obsahuje konkrétne príklady z praxe.

Plánovanie

Riziká projektu, ako sú problémy s financovaním, nízka kvalita výsledkov, právne prekážky alebo neefektívne postupy, je možné minimalizovať pomocou starostlivého plánovania a konceptualizácie. Táto metodika vysvetľuje niektoré kľúčové body koncepcie projektu digitalizácie.

Rozsah projektu

Spolupráca

Faktory nákladov a forma financovania

Vlastná výroba alebo outsourcing

Právne aspekty

Rozsah projektu

Čiastočné alebo celé podiely?

Moderná technológia skenovania, klesajúce náklady na ukladanie a softvér, ktorý zjednodušuje pracovný tok, spustili zmenu konceptu digitalizácie: namiesto jednotlivých dokumentov sa digitalizujú celé zbierky. Staré staršie podiely v takzvanej verejnej doméne sú prioritou, pretože ich autorské práva vyčerpali.

Prvotnými príkladmi hromadnej digitalizácie je projekt digitalizácie spoločnosti Google (external link) alebo projekty ako e-rara.ch (external link) od ETH E-Pics (external link).

Výberové kritériá na digitalizáciu poľnohospodárskych podnikov môžu byť:

Relevantnosť výskumu

(Akademická) dopyt

Zachovanie hospodárstiev (často používané materiály, unikátne exempláre)

Virtuálna rekonštrukcia zbierok a hospodárstiev

Praktický príklad - rozsah projektu E-Periodica

Materiály pre digitalizačný projekt E-Periodica (externý odkaz) sa vyberajú podľa nasledujúcich kritérií:

Časopis: typické časopisy, ktoré sa zobrazujú častejšie ako raz za rok, majú prednosť; okrem toho, ročenky a iné série

Odkaz na Švajčiarsko na základe obsahu, autora a vydavateľa, jazyka, miesta uverejnenia atď.

Neobchodný charakter: neziskové vydavateľstvá (napríklad špecializované združenia) alebo malé, miestne vydavateľstvá

Podpora od vydavateľov: pre časopisy, ktoré sú stále publikované, digitalizácia sa posudzuje iba za aktívnej podpory vydavateľa

Pridanie k existujúcej kolekcii digitálnych časopisov

Rámec objemu

Podrobný objemový rámec ponúka dôkladný základ pre plánovanie projektu. Najmä v prípade heterogénnych podnikov môže byť užitočné zaznamenať počet, druh a povahu materiálov. Pri písomných dokumentoch, ako sú knihy a časopisy, je výhodou, ak je počet strán známy alebo sa dá odhadnúť čo najpresnejšie.

Pamäťový priestor

Pri rozsiahlych projektoch sa čoskoro dostanú dáta do terabajtového rozsahu. Preto je dôležité naplánovať miesto, kde a v akom formáte sa majú údaje uchovávať alebo archivovať (pozri tiež Následné náklady).

Sprievodca sa zameriava na digitalizáciu tlačového a obrazového materiálu, ako aj objektov, ale nie AV médií. V tomto zmysle nepredstavuje žiadne tvrdenia o úplnosti.

Spolupráca

V rámci spolupráce môžu byť aspekty súvisiace s obsahom koordinované, vyhnúť sa nadbytočným činnostiam a zdieľať náklady. Zníženie nákladov je možné dosiahnuť spoločným obstarávaním a / alebo používaním infraštruktúry skenera. Dôležité aspekty kooperatívnej dohody:

Ciele projektu a exempláre

Trvanie a organizácia projektu

Služby partnerov projektu

Udeľovanie práv

Financovanie

Záruky a zodpovednosť

Trvanie a ukončenie dohody

Náklady a zdieľanie nákladov

Včasná výmena informácií o návrhoch digitalizácie a používanie centrálnych referenčných nástrojov je vo všeobecnosti dôležitá.

Praktický príklad - spolupráca e-rara.ch a e-manuscripta.chObe e-rara.ch (external link) a emanuscripta.ch (external link) sú koncipované a štruktúrované ako spoločné projekty medzi švajčiarskymi univerzitnými knižnicami. Medzitým sa však podieľajú aj mnohé ďalšie knižnice a inštitúcie a prispievajú k neustálemu rozširovaniu služieb s ich držbami

Faktory nákladov a forma financovania

Hlavnými nákladmi pre digitalizačné projekty sú personálne náklady a rozvoj technickej infraštruktúry a know-how.

Faktory nákladov vo všeobecnosti

Personálne náklady na skenovanie, kontrolu kvality, prípravu a prepracovanie materiálu, prípadne náklady na školenie na skeneri

Náklady na infraštruktúru (napr. Skener, softvérové ​​riešenia, databázy, IT podpora)

Náklady na licencovanie programov na spracovanie a indexovanie obrázkov, rozpoznávanie celého textu (OCR)

Povaha pôvodu (zviazané, voľné listy, špeciálne formáty, manipulácia)

Parametre skenovania (farba / stupnica šedi, rozlíšenie / dpi)

Online prezentácia (platforma IT infraštruktúry, podpora, nahrávanie metadát)

Prepravné a poistné

Faktory nákladov pre poskytovateľov externých služieb digitalizácie

Outsourcing externým poskytovateľom služieb sa najmä vyplatí pre individuálny projekt s objemom, ktorý by neodôvodňoval interné obstarávacie náklady. V závislosti od charakteru a rozsahu projektov je vhodné získať dve až tri úvodzovky.

Následné náklady

Následné náklady na údržbu softvéru, hardvéru a údajov závisia od konkrétneho prípadu a zvyčajne sa dajú odhadnúť len zhruba.

To isté platí pre náklady na ukladanie a dlhodobú archiváciu. Napriek tomu, že ceny skladovacích kapacít v posledných rokoch neustále klesajú, môžu naďalej rozširovať rozpočet projektu na veľké úsilie v oblasti digitalizácie. Je potrebné zvážiť skutočnosť, že ide o dlhodobé prevádzkové náklady nad rámec projektu. Okrem toho, čím väčšie je množstvo údajov, tým vyššie sú náklady na ochranu údajov a následné migrácie údajov.

## Formy financovania

V závislosti od projektu a jeho rámcových podmienok existujú rôzne možnosti financovania:

samofinancovania

spolupráca

Financovanie tretích strán (napríklad podporné programy)

Partnerstvo verejného a súkromného sektora: kultúrne inštitúcie spolupracujú so súkromnými spoločnosťami a spoločne vykonávajú verejné služby, napr. spoluprácu medzi Bavorskou štátnou knižnicou (externý odkaz) a Google (external link).

sponzoring

Praktický príklad - financovanie E-Periodica

Špecializované združenia a vydavatelia zvyčajne majú k dispozícii iba obmedzené prostriedky na financovanie projektov digitalizácie. V prvých dňoch bola E-Periodica (externý odkaz) spolufinancovaná z projektových peňazí z e-lib.ch a prostredníctvom konzorcia švajčiarskych akademických knižníc (external link) a vlastného financovania ETH-Bibliothek.

V súčasnosti model nákladov vyzerá takto:

Jednorazové náklady na spracovanie a aktiváciu retro holdingov pre časopis: financované partnermi projektu (vydavateľ, odborné združenie atď.) A ETH-Bibliothek

Ročné náklady na pridanie nových emisií a pamäte a prevádzkové náklady pre celý časopis sa v plnej miere prenášajú na partnerov projektu.

Švajčiarska organizácia Švajčiarska akadémia humanitných a spoločenských vied (external link) prispieva k financovaniu ročných nákladov v časopisoch humanitných a spoločenských vied.

Právne aspekty

Právnu situáciu je potrebné overiť a práva na používanie digitálnych dokumentov by sa mali získať čo najskôr, podľa možnosti už vo fáze plánovania.

## Autorské práva (Copyrights)

Podľa zákona o autorských právach na Slovensku sú diela chránené až 70 rokov po úmrtí autora, po ktorom sú voľne dostupné ako takzvané "verejné domény".

(§ 32 *Trvanie majetkových práv*) "(1) Majetkové práva trvajú od okamihu vytvorenia diela počas autorovho života a 70 rokov po jeho smrti. Pri spoluautorskom diele majetkové práva trvajú počas života posledného zo spoluautorov a 70 rokov po jeho smrti. Ak je audiovizuálne dielo vytvorené ako spoluautorské dielo, majetkové práva trvajú počas života poslednej osoby spomedzi režiséra, autora scenára, autora dialógov a autora hudby, ktorá bola vytvorená osobitne pre toto dielo a 70 rokov po jeho smrti." http://www.zakonypreludi.sk/zz/2015-185#cast2

Napríklad práca zverejnená v roku 1905, ktorého autor zomrel v roku 1955, je preto autorským právom do roku 2025.

## Vydavateľské a licenčné zmluvy

Je potrebné overiť, či *vydavateľ* diela vlastní práva na jeho (dodatočné) zverejnenie na *internete.* (1990, t. j. koniec systému ARPANET a prvé nasadenie WWW v Cerne vo Švajčiarsku).

Znamená to, že do roku 1990 neexistovali vydavateľské zmluvy, v ktorých by si vydavatelia dohodli s autormi práva zverejnenia na internete. Podľa našich skúseností sa na Slovensku podmienky zverejňovania dohodnutého diela na internete v autorských zmluvách upravovali a upravujú len zriedka. Hoci prax v tomto ohľade sa postupne mení. V prípade starších publikácií to možno vylúčiť hneď od začiatku, pretože internet neexistoval pri dohode medzi autorom a vydavateľom.

## Praktický príklad

Okrem niekoľkých výnimiek sa staré a vzácne tlače, ktoré boli publikované pred rokom 1900 a na ktoré sa už nevzťahujú žiadne obmedzenia týkajúce sa autorských práv, *môžu voľne digitalizovať a sprístupňovať* organizáciami, ktoré ich vlastnia alebo spravujú vo svojich zbierkach.

Praktický príklad – seriálové publikácie, noviny, časopisy, periodické zborníky

Predpokladom na zaradenie týchto dokumentov na digitalizáciu je súhlas držiteľov autorských práv. Výnimku tvoria časopisy až do 19. storočia, kedy už nie je možné zistiť držiteľov autorských práv.

Skúsenosti ukázali, že iba vydavatelia novších obchodných časopisov uzatvárajú zmluvy s ich autormi. Pritom *autori prenášajú plné práva na používanie daného článku* na vydavateľa (dohoda o autorských právach, licencia). Pokiaľ ide o historický obsah, takéto jasné dohody zvyčajne neexistujú a získanie spätného povolenia od autorov a ich potomkov prostredníctvom správnych právnych kanálov by znamenalo neúmerné úsilie.

## Dispenzácia s formálnym povolením

Pre digitalizáciu novín a časopisov je možné prijať pragmatický postup (ako napr. vo Švajčiarsku): autorský súhlas nie je získaný, ale autor môže následne požadovať zablokovanie alebo vymazanie jeho práce z online služby.

Tento prístup predpokladá, že držitelia autorských práv sú akademicky aktívni ľudia alebo autori na čiastočný úväzok, ktorí podporujú ďalšie šírenie svojej práce v záujme slobodnej vedy.

Príprava materiálu

S cieľom zabezpečiť čo najplynulejšie fungovanie projektu by sa otázky týkajúce sa prípravy a zaobchádzania s dokumentami mali vyriešiť čo najdôkladnejšie.

## Indexovanie

Ak nie je k dispozícii záznam v katalógu, databáza alebo podobne, je potrebné, aby boli dokumenty indexované pomocou základných metadát.

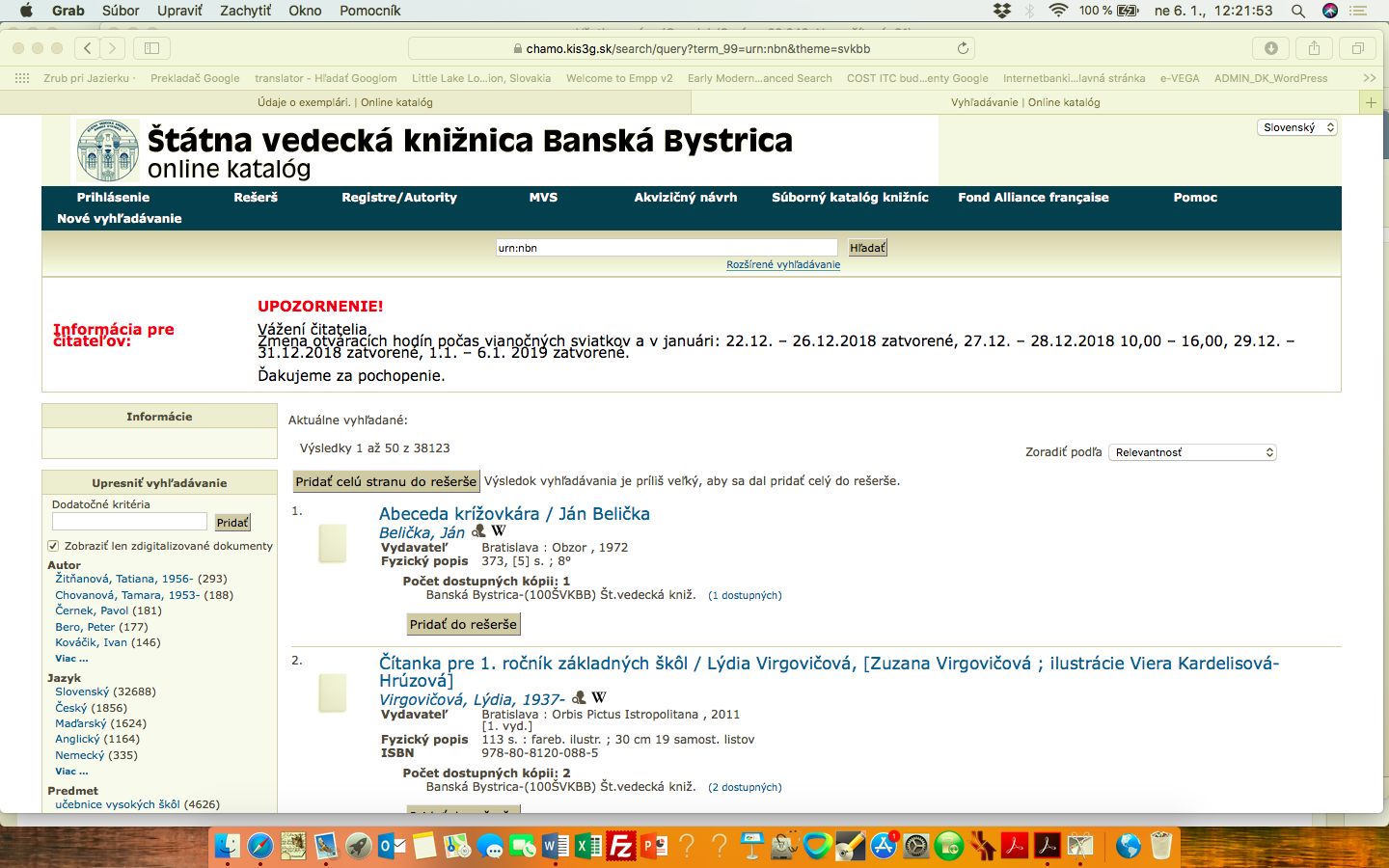
Každý dokument vyžaduje identifikátor pre jednoznačnú identifikáciu diela, na základe ktorého sa skeny môžu priradiť k príslušnému súboru údajov.

Môže ísť napríklad o individuálne číslo systému v katalógu knižníc.

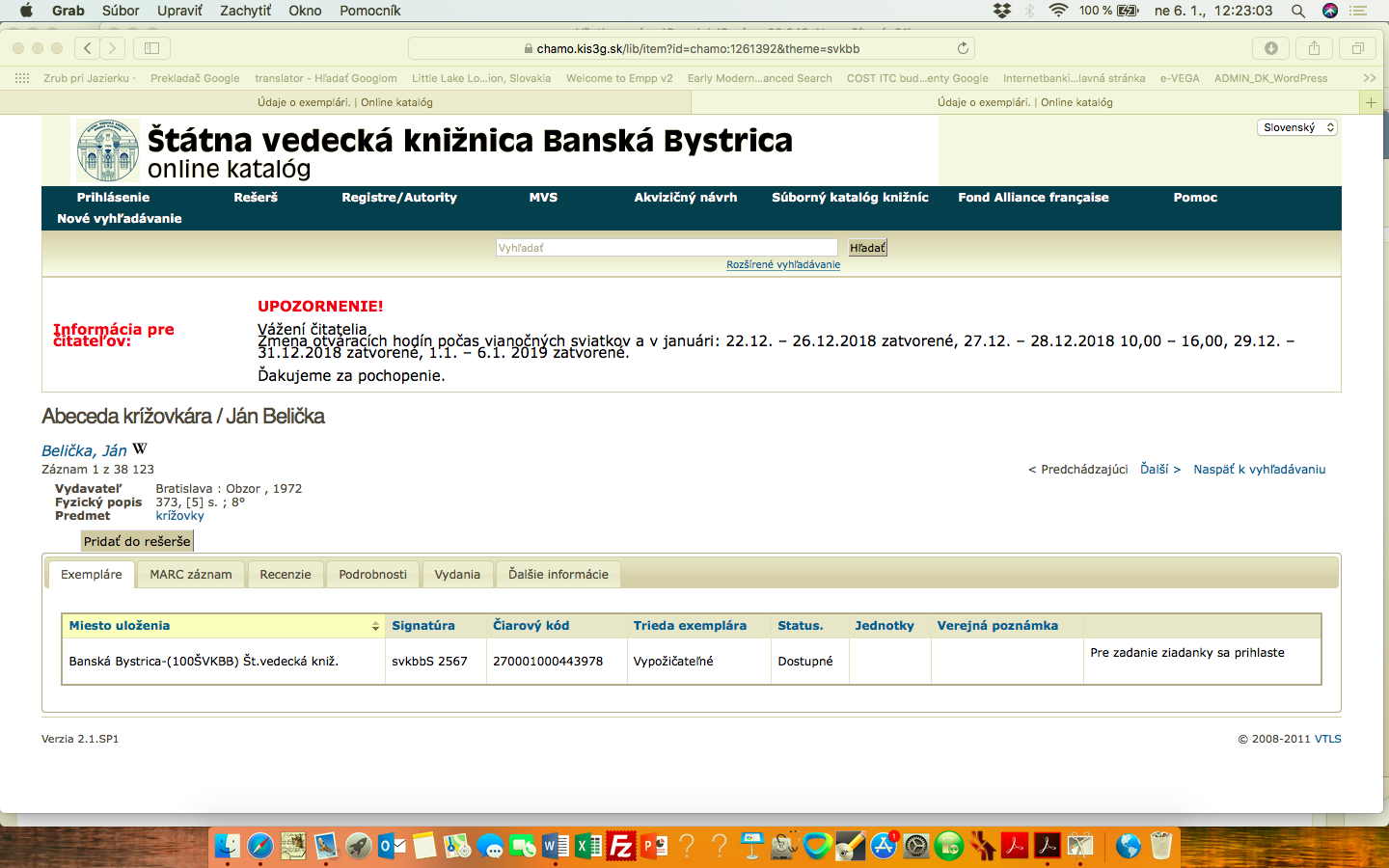
## Praktický príklad prebratia záznamu a identifikátorov zo súborného katalógu

Knižnica chce sprístupniť exemplár dokumentu, ktorý je už digitalizovaný v SNK.

## Skrátený bibliografický záznam o dokumente v súbornom katalógu Slovenská knižnica



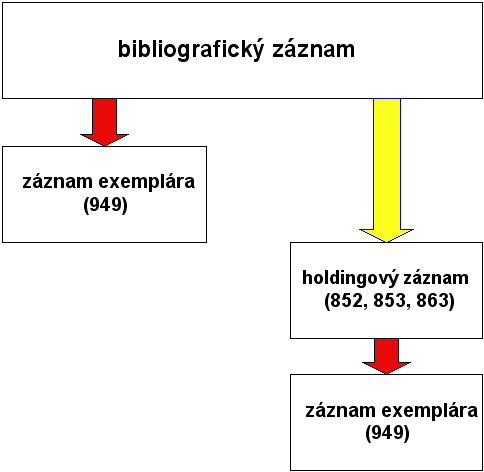
## Záznam o exemplári, ktorý chceme digitalizovať alebo je už digitalizovaný

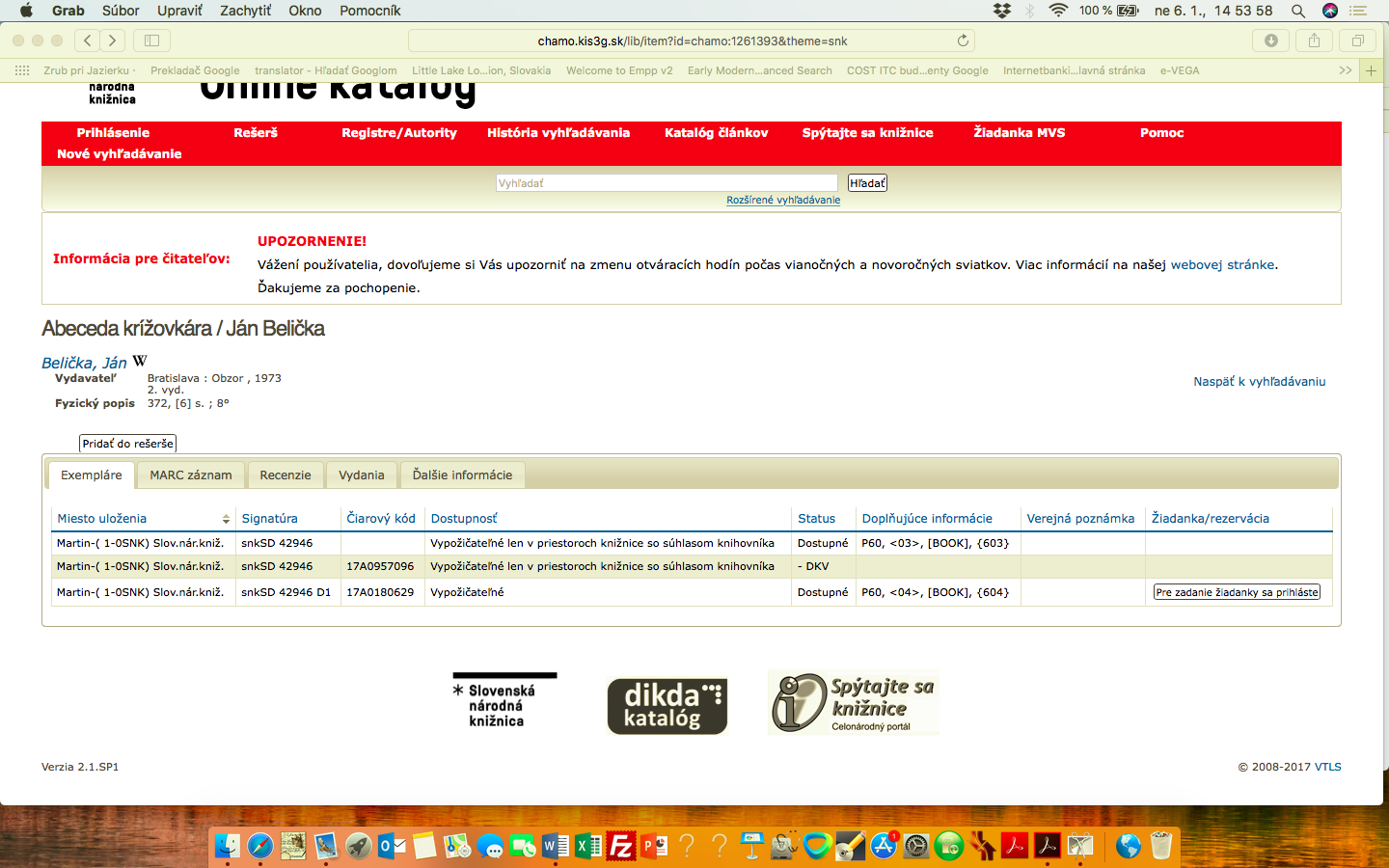


Ak digitalizujeme alebo preberáme údaje o exemplári, ktorý máme v SVKBB získavame údaje zo záznamu o exemplári, ktorý je v lokácii SVKBB. Konkrétne preberáme údaje:

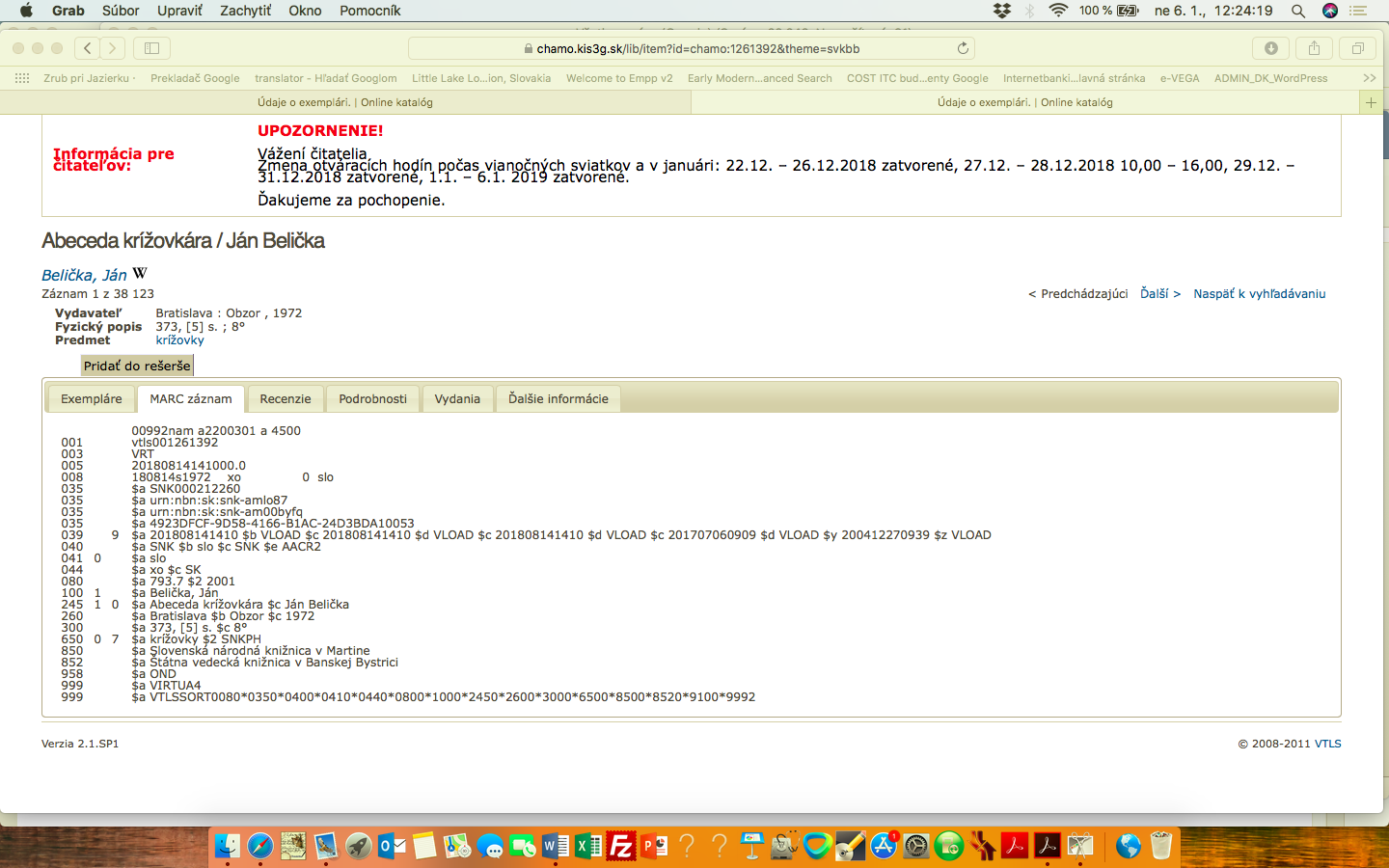
Signatúru v SVKBB: svkbbS 2567

Čiarový kód SVKBB: 270001000443978





1. Prebrať bibliografické údaje zo súborného katalógu



Príprava materiálu

V závislosti od typu dokumentu môže byť príprava materiálu viac alebo menej komplikovaná.

Prípadné prípravné práce môžu byť:

Zrušenie obalov, ktoré už nie sú potrebné (napr. Časopisy) knihármi

Označovanie dokumentov

Odstránenie sponiek

Predtriedenie, vytváranie jednotiek digitalizácie

Zostavovanie dokumentácie pre proces digitalizácie

Podrobné informácie o dokumentoch uľahčujú plánovanie personálnych a finančných zdrojov. Potrebné informácie môžu byť uvedené v rámci objemu.

Rôzne materiály majú rôzne požiadavky na digitalizáciu. Medzi dôležité parametre patrí druh papiera, stránky v špeciálnych formátoch, rozložené časti, farebné stránky, obrázky, kresby atď.

Pre cenné dokumenty môže byť užitočná aj dokumentácia o stave prác na zaznamenanie akýchkoľvek existujúcich škôd.

DIGIREP SVK BB

Digitalizácia a budovanie inštitucionálnych digitálnych repozitov, ako je napríklad DIGIREP SVKBB by sa mali opierať o poznatky o najlepšej praxi. Najlepšia prax okrem iného znamená, že sa v najvyššej možnej miere *automatizujú* procesy, operácie a úkony v digitalizácii, spracovaní, uchovávaní a prezentácii digitálneho obsahu používateľom.

Odborná práca a erudícia knihovníkov, bibliografov sa však musí uplatniť v procese rozhodovania o *výbere* dokumentov, v rozhodovaní o *spôsobe* skenovania, v inštrukciách o šetrnom *zaobchádzaní* s dokumentami, v procesoch kontrole *kvality* skenovania, v procesoch kontroly *identifikácie* objektov, v kontrole kvality deskriptívnych (bibliografických) *metadát*, v určovaní pravidiel sprístupnenia digitálneho materiálu v súlade s *copyrightom* a pod.

Maximálna automatizácia ostatných procesov šetrí *náklady* a zvyšuje *produktivitu*.

V DIGIREP SVKBB by sme mali smerovať k flexibilnému systému pracovných postupov (workflow) na riadenie procesov.

Súčasťou procesu musí byť automatické získavanie metadáta z existujúceho knižničného systému (napr. Virtua) a spárovanie deskriptívnych metadát s digitálnym objektom v celom procese digitalizácie, až po sprístupnenie digitálneho objektu na Webe.

Považujeme za dôležité, aby sa dokumenty pre používateľov na Webe zobrazovali tak, ako boli vytlačené. Používateľ by teda mal mať možnosť *vidieť* jednotlivé *strany knihy* tak, ako keby si prezeral analógovú papierovú knihu. Mal by mať možnosť *vidieť* *novinovú stranu* tak ako bola vytlačená aj so všetkými článkami na strane, nie len jednotlivý *vybratý* článok.

Samozrejme, že na účely rešeršovania a vyhľadávania musí byť po automatickom procese optického rozlíšenia (OCR) a po automatickej segmentácii strany dostupné a vyhľadateľné všetko čo sa nachádza na celej strany knihy alebo novín (text, obrázky, nadpisy, hranice článkov, reklamy, ilustrácie a pod. Digitalizácia umožňuje omnoho efektívnejšie rešeršovanie ako tzv. prekonaná „*analytika“*.

Tzv. „*analytiku*“, ktorá je známa hlavne ako *bibliografický rozpis novín a časopisov*, nahrádza v digitalizácii *automatická segmentácia*, ktorá je súčasťou automatického procesu OCR po skenovaní.

Tradičné manuálne rozoberanie dokumentu (t. j. napr. zdrojov na pokračovanie, teda seriálov či periodík) na články, názvy, titulky, obrázky a text je na jednej strane *málo efektívne* a pre väčšinu softvérových aplikácií nadbytočné a pre inštitúcie mimoriadne *finančne náročné* a *informačne nevýznamné*.

„Analytika“ ako *výberové* rozoberanie dokumentov na články, teda tzv. rozčlánkovanie, analytický rozpis či excerpovanie bolo na Slovensku rozvíjané od konca 19. st. a čias geniálneho V. Riznera v čase tzv. modernej, „ručnej“ bibliografie. Uplatňovaním informačných a komunikačných technológií tzv. *analytika* stratila opodstatnenie pri rozpise *novín* a bežných neodborných a nevedeckých časopisov.

*„Analytika“*, ktorá sa v súčasnosti robí automaticky, však má zmysel pri *rozpise vedeckých časopisov a zborníkov*, pričom každý článok má v digitálnych repozitoch svoju štandardnú jednoznačnú identifikáciu SICI a URN (DOI), podobne ako majú jednoznačnú identifikáciu knižničné jednotky vo fonde alebo v zbierke knižnice alebo inej informačnej inštitúcie.

Články z novín, samozrejme, vzhľadom na svoju efemérnu hodnotu, žiadnu jednoznačnú identifikáciu nemajú a nie sú samostatnými objektami digitalizácie.

Rozvoj technológií spôsobil, že digitalizácia je pri seriáloch, teda novinách a bežných neodborných a nevedeckých časopisoch orientovaná na ***stranu* novín** (page oriented) a **nie** na ***článok z novín***(article oriented).

Prostredníctvom softvéru *Mediainfo* je možné preletieť cez stránky, priblížiť a oddialiť bez ohľadu na ich rozlíšenie a veľkosť súboru. Používateľ vyhľadávať v obsahu dokumentu, strany, čísla, ročníka, zbierky. Môže vždy vidieť celú stránu, rolovať stranu, plynule môže stranu zväčšovať a zmenšovať, vidí, kde článok začína a končí, identifikovať jeho názov a autora, pokračovanie na ďalšej stránke, vidí rozlíšenie obrazu od textu, reklamy z článku, noticky, správa, ktoré by sa inak do „analytiky“ nikdy nedostali, bez toho, aby bolo potrebné, aby niekto určil zónovanie a riadil proces rozpisu. Rýchle odkazy na titul (iný typ písma, tučné, verzály, kurzíva...) sú plne automaticky vytvorené a môžu byť použité aj ako rýchle odkazy na obsah knihy kapitoly alebo články v novinách.