20. Převod dat

Podobně jako stěhování domácnosti je změna knihovního softwaru příležitostí k „úklidu“ a k rozhodnutí, co s sebou chceme „stěhovat“ a kam přesně chceme či můžeme data v novém prostoru umístit. Proto je před přechodem na jiný software užitečné provést nejen revizi knihovního fondu, ale také analýzu stávajících dat a jejich kvality a případně je vyčistit. Důležitá je také analýza toho, jakým způsobem jsou na sebe ve stávajícím softwaru navázány jednotlivé údaje a zda jsou tyto návaznosti v souladu s návazností údajů v novém softwaru.

Díky přípravným krokům zahrnujícím revizi fondu a „úklid dat“ nebudete do nového softwaru přenášt nepotřebná data a informace. Tím se mimo jiné zmenšuje pravděpodobnost výskytu chyb při exportu a importu, protože nekvalitní záznamy, duplicitní záznamy a některé další chyby mohou při převodu (migraci) dat působit problémy automatického zpracování dat.

„Úklid dat“ může v některých případech také přispět ke znatelným úsporám nákladů vynaložených na jejich převod do nového softwaru.

20.1 Porovnání návaznosti činností

Převod některých dat nemusí být úplně jednoduchý, protože návaznosti dat a činností se v jednotlivých softwarech mohou lišit. Ať už se chystáte na převod dat vlastními silami nebo jej bude provádět dodavatel softwaru či jiný subjekt, nezapomeňte se zaměřit zejména na tyto oblasti:

- **oprávnění**, která se v jednotlivých softwarech mohou lišit jak mírou podrobnosti (granularitou), tak způsobem, jakým jsou navázána na jednotlivé činnosti nebo vztažena k jednotlivým organizačním složkám knihovny;
- **způsob, jakým jsou propojeny jednotky nebo čísla časopisů s bibliografickým záznamem** (např. údaje o číslech časopisů mohou být napojena přímo na bibliografický záznam nebo na informace o předplatném);
- **způsob, jakým software pracuje se signaturami, přívazky nebo přítisky, vícesvazkovými publikacemi** apod.;
- **spouštění automatických akcí a událostí** (v jednotlivých softwarech se může lišit způsob, jakým jsou oznámení o rezervacích, upomínkách, nabíhání zpozdného, vypršení platnosti registrace atd. navázána na jednotlivé komponenty a události systému);
- **některá data o čtenářích** (může se např. lišit způsob ukládání historie výpůjček nebo práce se seznamy knih, které si ukládají čtenáři);
- **funkce a možnosti, které v novém softwaru nejsou k dispozici**;
- **funkce a možnosti, které nejsou ve stávajícím softwaru, ale jsou k dispozici v softwaru novém**.



S přípravou podkladů pro převod dat a s jejich vlastním převodem si obvykle poradí dodavatel softwaru či jiný subjekt zajišťující převod dat. Potřebuje k tomu však

součinnost knihovny. Proto je důležité, aby knihovna měla zmapované používané postupy a probíhající procesy a věděla, která data k nim potřebuje a jakým způsobem s nimi pracuje.

20.2 Analýza a zpracování dat

Při přechodu na nový software hraje velkou roli kvalita dat. Proto je při změně knihovního softwaru důležitým krokem hodnocení stavu dat a jejich příprava (vyčištění), ať už jde o bibliografické a autoritní záznamy či např. údaje o jednotkách, čtenářích, výpůjčkách, rezervacích a akvizici.

20.2.1 Kontrola kvality bibliografických a autoritních záznamů

Bibliografické a autoritní záznamy jsou podstatné nejen z hlediska převodu do nového softwaru, ale také z hlediska zapojení do kooperativních projektů. Proto je nutné věnovat jejich kvalitě velkou pozornost. Tato data jsou obvykle exportována a importována ve formátu typu MARC (na rozdíl od dalších dat, především údajů o čtenářích, výpůjčkách, rezervacích, platbách a akvizici; ta jsou obvykle exportována a importována ve formátu CSV nebo TSV, případně v dalších formátech). Pokud nejsou bibliografické a autoritní záznamy dostatečně kvalitní (např. kvůli nedodržování katalogizačních standardů), může to při přechodu na nový software způsobit značné problémy.

Příprava bibliografických a autoritních záznamů by měla zahrnovat:

- kontrolu duplicit a následnou deduplikaci záznamů;
- odstranění bibliografických záznamů, které nemají připojeny žádné exempláře;
- kontrolu kvality záznamů z hlediska používaných standardů.

Pokud knihovna používá knihovní software, který splňuje všechny knihovní standardy (mj. formát MARC 21 a kódování UTF-8), a zároveň při katalogizaci dodržuje požadované standardy (RDA, ISBD), je pravděpodobné, že stav dat je vyhovující a při jejich převodu nedojde k závažnějším problémům.[1)]

20.2.2 Problémy způsobené omezením softwaru

V ČR se okrajově stále využívají některé starší knihovní softwary, které nesplňují některý z knihovních standardů, např. nepodporují strukturu záznamu ve formátu MARC nebo podporují pouze UNIMARC a nikoliv MARC 21, nepodporují kódování UTF-8 apod. V takových případech je nutné před převodem dat nebo při něm provést automatickou konverzi do jiného formátu či kódování, případně jinou formu úpravy záznamů.

Automaticky lze řešit i některé opakovaně se vyskytující katalogizační chyby, např. problémy s interpunkcí podle ISBD. Dodavatelé nebo provozovatelé knihovních softwarů obvykle mají k dispozici konverzní programy pro převod dat z formátu UNIMARC do formátu MARC 21. Jsou také schopni vytvořit konverzní programy na míru dané knihovně. Ta však musí počítat se součinnosti při analýze a testování těchto konverzí.

Z hlediska možnosti provedení automatické konverze dat mohou být problematické některé starší softwary, v nichž je omezen počet znaků, takže některé údaje jsou zkrácené. V takových případech je obvykle třeba provést manuální úpravu záznamů.

20.2.3 Problémy způsobené chybným zpracováním

Důvodů, proč jsou v některých knihovnách méně kvalitní záznamy, může být více. Např.:

- záznamy pocházejí z retrokonverzí, tudíž nesplňují požadavky na minimální záznam do Souborného katalogu ČR, případě vůbec nejsou ve struktuře formátu MARC;
- v knihovně neprobíhá (nebo v určité době neprobíhala) kontrola bibliografických záznamů a nejsou sledovány nové metodiky a postupy;
- knihovna nemá (nebo v určité době neměla) k dispozici dostatečně kvalifikované pracovníky znalé katalogizačních pravidel a postupů.



Některé chyby v kvalitě záznamů je možné vyřešit pomocí konverzních programů na míru. Je to však možné pouze v těch případech, kdy jsou chyby „konzistentní“, tj. opakují se u více záznamů.

Manuální opravy záznamů je optimální provést před přechodem na nový software. Novou katalogizaci je naopak obvykle vhodné provést až v novém softwaru.



Pokud má knihovna nekvalitní bibliografické záznamy a přitom velikost fondu je maximálně dva až tři tisíce knihovních jednotek a zároveň jde o běžnou knižní produkci, bývá efektivnější knihovní jednotky zkatalogizovat znovu. Stažení záznamu prostřednictvím protokolu Z39.50 s připojením exemplářů je obvykle rychlejší než výběrová ruční oprava záznamů. Náklady na stažení záznamů tedy mohou být mnohem nižší než náklady na manuální opravy záznamů a/nebo vytvoření programu pro jejich hromadnou opravu.

20.3 Proces přípravy a převod dat

Jak už bylo řečeno, převodu dat by měla předcházet fyzická příprava fondu (revize fondu a případné hromadné změny a přesuny fyzických jednotek a údajů o nich) i revize dalších oblastí činnosti a s tím související čištění dat (neaktivní čtenáři, staré objednávky apod.).

Samotný převod dat sestává z několika fází. Obvykle zahrnuje:

- export dat ze stávajícího knihovního softwaru;
- konverzi dat z jednoho formátu do jiného (např. z formátu UNIMARC do formátu MARC 21 nebo z formátu MARC 21 do formátu MARCXML) a/nebo změnu kódování na UTF-8; tyto konverze

mohou být součástí exportu dat ze stávajícího softwaru (pokud to tento software umožňuje);
- hromadnou (dávkovou) úpravu dat, např. odebrání nestandardních polí, úpravu nekvalitních záznamů, hromadné změny fondu či uživatelských dat/nastavení, které chce knihovna spojit s přechodem na nový software;
- případnou manuální opravu záznamů, které nelze opravit dávkově (oprava může probíhat jak před prvním testovacím importem dat, tak v průběhu všech testovacích importů nebo po dokončení ostrého importu);
- import dat do nového softwaru;
- testování softwaru;
- opravy chyb zjištěných při importu.

20.4 Spolupráce s externím subjektem

Na převodu dat ze stávajícího softwaru do nového se obvykle podílí externí subjekt. Měla by mu proto předcházet jednoznačná dohoda o tom, kdo má jakou roli a odpovědnost, jaké konkrétní úkoly je třeba vykonat, případně jaká rozhodnutí učinit a kdo je učiní[2)]. Specifikace těchto úkolů a rolí by měla být (přinejmenším v základních rysech) součástí poptávkového dokumentu a měla by být detailně popsána ve smlouvě o pořízení/zavedení knihovního softwaru.

Subjekty, které dodávají knihovní softwary nebo zajišťují přechod na ně a jejich provoz, obvykle mají znalosti nutné pro převod dat z různých softwarů. Pokud navíc knihovna přechází na komerční knihovní software, který je následníkem stávajícího softwaru, může být celý proces jednodušší, protože software obvykle mívá (alespoň v některých rysech) podobnou architekturu a pracovníci dodavatelské firmy mají znalosti potřebné pro převod dat. V takovém případě je tedy nutné pouze řešit specifické záležitosti, které se týkají postupů konkrétní knihovny a kvality jejích dat (např. využívání některých funkcí stávajícího softwaru nestandardním způsobem, nestandardní využívání polí MARC apod.).

Přesto je důležité, aby si knihovna vyjasnila některé otázky, které souvisejí s převodem dat[3)]. Může jít např. o otázky související s pracovníky, daty nebo stávajícím softwarem a jeho funkcí.

Pokud jde o **pracovníky**, jedná se o to:

- kdo z nich má (bude mít) odpovědnost za převod dat;
- jaké mají v této oblasti zkušenosti;
- jakým způsobem bude zajištěna priorita převodu dat před případnými dalšími úkoly.

V případě **dat** je vhodné mít jasno v tom:

- jaká data knihovna používá;
- kde a jakým způsobem jsou tato data uložena (pouze v knihovním softwaru nebo i jinde);
- jaký je objem dat;
- jaké typy souborů nebo programů knihovna používá pro správu těchto dat;
- jaké standardy je nutné dodržovat;
- která data se mají importovat do nového softwaru;
- jaké je kvalita jednotlivých souborů dat (bibliografické záznamy, údaje o exemplářích, čtenářích atd.).

Z hlediska **stávajícího softwaru a jeho funkce** zjišťujeme:

- jakým způsobem je z něj možné exportovat data;
- jak jsou v něm nastavena a spravována uživatelská oprávnění a jak jsou přenositelná do nového softwaru;
- jak je zabezpečen;
- z jakých dalších softwarů nebo systémů bude nutné převádět data a jak (v jakých formátech) je možné tato data exportovat.

20.5 Nástroje pro kontrolu a čištění dat

Pro kontrolu a čištění dat lze použít jak běžně dostupné, tak specializované nástroje. Může jít o tabulkové procesory, speciální nástroje nebo nástroje, které jsou součástí některých knihovních softwarů.

20.5.1 Tabulkové procesory

Tabulkový procesor (např. MS Excel, LibreOffice nebo Apache OpenOffice) lze využít zejména pro data vyexportovaná z databáze knihovního softwaru ve formátu CSV, TSV apod. Zobrazení dat v tabulce díky automatické i vizuální kontrole dat umožňuje jednoduše a rychle odhalit některé druhy chyb (např. nesprávné typy či délky polí, duplicity či chybějící údaje).



Možnost zobrazení alespoň některých dat formou tabulky, které umožní vizuální kontrolu dat, nabízejí i některé knihovní softwary.

20.5.2 Speciální nástroje

MarcEdit je volně dostupný nástroj původně určený pro editaci záznamů ve formátu MARC. V editoru záznamů je možné provádět kontrolu polí, deduplikaci záznamů, hromadné mazání, editaci nebo doplnění polí, podpolí, indikátorů apod. MarcEdit se však neomezuje pouze na formát MARC, nýbrž nabízí nástroje pro práci s metadaty, umožňuje stahování záznamů pomocí Z39.50 nebo SRU/SRW, sklízení pomocí OAI-PMH, konverzi do různých formátů nebo znakových sad, dávkovou úpravu záznamů aj.

Existuje i řada dalších nástrojů, které je možné využít pro práci s daty. Některé z nich jsou však určeny spíše pro programátory. Těm, kdo by případně chtěli získat další informace o těchto softwarech, doporučujeme knihu *Migrating library data: a practical manual* (viz Doporučené zdroje níže).

20.5.3 Nástroje v knihovních softwarech

Podpora pro hromadnou úpravu nebo kontrolu dat a údajů bývá součástí knihovních softwarů. Může jít např. o následující funkce[4]:

- hromadná úprava bibliografických dat;
- sady pravidel a nástrojů pro přepsání méně kvalitních záznamů při importu/exportu;
- deduplikace bibliografických záznamů;
- hromadná úprava údajů o jednotkách;
- hromadná úprava dat o uživatelích;
- kontrola odkazů (URL) u elektronických zdrojů.

Doporučené zdroje

Pro ty, kdo připravují převod dat vlastními silami, doporučujeme knihu: BANERJEE, Kyle a Bonnie PARKS (eds.). *Migrating library data: a practical manual*. Chicago: Neal-Schuman, 2017. xix, 251 s. ISBN 978-0-8389-1503-5.

1) Problémy kvality údajů o knihovních jednotkách do značné míry souvisejí s aktuálním stavem fondu. Část problémů s kvalitou dat může vyřešit revize fondu, která předchází převodu dat.

2) BANERJEE, Kyle a Bonnie PARKS (eds.). *Migrating library data: a practical manual*. Chicago: Neal-Schuman, 2017, s. 219. ISBN 978-0-8389-1503-5.

3) Viz též kapitolu Analýza situace a potřeb knihovny, popř. publikaci BANERJEE, Kyle a Bonnie PARKS (eds.). *Migrating library data: a practical manual*. Chicago: Neal-Schuman, 2017, s. 230. ISBN 978-0-8389-1503-5.

4) Nabídka funkcí a jejich možnosti se ovšem v jednotlivých knihovních softwarech liší.

From:
https://prirucky.ipk.nkp.cz/ - **Metodické příručky pro knihovny**

Permanent link:
https://prirucky.ipk.nkp.cz/aks/prevod_dat

Last update: **2020/09/24 12:49**