

## Zkušenosti z implementace AKS KP-WIN v Národní pedagogické knihovně Komenského (NPKK)

Zuzana Švastová\*  
svastova@npkk.cz

**Abstrakt:** Poznatky a zkušenosti NPKK z přechodu na AKS KP-WIN. NPKK do konce roku 2004 provozovala čtyři různé AKS (jednotlivá oddělení používala různé systémy, tato situace byla dána postupnou automatizací knihovny a nedostatky některých AKS); od poloviny r. 2004 dochází v rámci přechodu na jeden AKS k postupnému sjednocování dat i pracovních postupů.

Zkušenosti NPKK s AKS KP-WIN z hlediska: speciálních importů dat, nových funkcí v různých oblastech zpracování i zpřístupňování dat, specifických požadavků NPKK na budování Českého pedagogického tezauru.

**Klíčová slova:** NPKK, automatizace knihovny, KP-WIN, databáze, věcný popis, tezaurus, analytické zpracování

### 1 Úvod

Národní pedagogická knihovna Komenského (dále jen NPKK) je podle zákona č. 257/2001 Sb., o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb (knihovní zákon), **specializovanou knihovnou zaměřenou na oblast výchovy, vzdělávání a školství s celostátní působností**. Je součástí (jednou z divizí) **Ústavu pro informace ve vzdělávání**, který je přímo řízenou organizací Ministerstva školství, mládeže a tělovýchovy ČR.

NPKK poskytuje knihovnické a informační služby právníkům a fyzickým osobám, především ve vysokém školství, výzkumu a vzdělávání, a dalším pracovníkům školství. Současně plní funkce veřejně přístupné vědecké knihovny, která vytváří a spravuje státem dotovaný fond domácí a zahraniční literatury, informačních pramenů a zdrojů z oblasti výchovy, vzdělávání, školství a příbuzných společenských věd. Svůj fond zpřístupňuje prostřednictvím výpůjčních, reprografických, elektronických, rešeršních, bibliografických, informačních a publikačních služeb – při uplatnění zásad jejich diferenciací – především pedagogické a studující veřejnosti.

NPKK pro automatizaci svých procesů a činností dosud využívala dva automatizované knihovnicko-informační systémy (přičemž tyto systémy byly využity v celkem čtyřech dílčích instalacích): KP-SYS a ISIS. Oba systémy byly provozovány v prostředí operačního systému MS-DOS, který je v současnosti zastaralý. Funkce, jimiž oba používané systémy disponovaly, vycházely z možností prostředí MS-DOS a již zcela neodpovídaly potřebám knihovny, a to jak z hlediska zajištění jejích procesů a funkcí, tak z hlediska uživatelů, kteří jsou nevyhovujícími systémy do značné míry limitováni. Systém ISIS rovněž nenabízel funkce nezbytné pro zpřístupňování informací v prostředí world wide webu. Proto se NPKK rozhodla přistoupit k nezbytnému kroku v procesu komplexní automatizace – k přechodu na AKS

---

\* Ústav pro informace ve vzdělávání – Národní pedagogická knihovna Komenského, Mikulandská 5, 116 74 Praha 1

jednotný pro všechny činnosti a procesy, který zároveň umožňuje standardní moderní zpřístupňování informací a služeb v prostředí www.<sup>1</sup>

Na základě řady hledisek a analýzy dostupných AKS na českém trhu se NPKK rozhodla pro **implementaci AKS KP-WIN** firmy KP-SYS zejména z těchto důvodů:

- v NPKK je provozován KP-SYS (operační systém MS-DOS), tj. předchůdce KP-WIN (OS MS WINDOWS), to znamená kompatibilitu dat a menší náročnost při převodech dat z jednoho systému do druhého;
- funkce, jimiž KP-WIN disponuje, odpovídají (vyhovují) potřebám vnitřních i vnějších procesů NPKK;
- KP-WIN disponuje rovněž funkcemi nezbytnými pro standardizované zapojení NPKK do kooperačních systémů knihoven zabývajících se koordinovanou tvorbou dat – tj. katalogizačních záznamů i bází/souborů autorit;
- jde o systém prověřený řadou instalací ve větších knihovnách.

### 1.1 Základní cíle přechodu na jednotný systém

- modernizace hardwaru pracovníků NPKK,
- implementace AKS KP-WIN v NPKK.

**Z hlediska uživatelů:** transparentní databáze v jednotném databázovém prostředí KP-WIN s možností on-line objednávky výpůjčky, včetně on-line rezervací knih a řady navazujících služeb.

**Z hlediska pracovníků:**

- jednotný knihovní systém a databázové prostředí,
- odstranění duplicitních činností,
- zapojení NPKK do kooperačních systémů – možnost přebírání dat (katalogizačních záznamů i souborů autorit),
- uživatelsky příjemné prostředí s řadou standardních funkcí.

### 1.2 Postup implementace AKS KP-WIN

1. přechod Českého pedagogického tezauru (DB ISIS 1),
2. přechod analytické Pedagogické bibliografické databáze – PBD (DB ISIS 1),
  - části: Články z periodik, Stati ze sborníků, Recenze knih,
3. přechod analytické Komeniologické databáze – KOM (DB ISIS 1),
  - části: Články z periodik, Stati ze sborníků, Recenze knih, Monografie
4. tvorba databáze periodik,

---

<sup>1</sup> NPKK současně řešila další problém, a to modernizaci stávajícího hardwaru. Většina dosud používaných počítačů byla konfigurace Pentium I s procesorem 100 MHz (PC zakoupena v letech 1995–1997) a jsou de facto za hranic své životnosti. V roce 2003 se podařilo díky dotaci MŠMT v rámci rekonstrukce studovny pořídit tři nové počítače pro pracovníky služeb (dva do výpůjčního protokolu a jeden do studovny), další počítače byly určeny výhradně pro uživatele NPKK – pro vyhledávání v databázích NPKK a pro práci na internetu. V roce 2004 bylo v rámci investičního záměru pořízeno dalších 12 počítačů a koncem roku ještě z úspor ÚIV čtyři PC. Stav roku 2005 je však stále nevyhovující, pořád chybí cca 10 až 15 kvalitních počítačů.

5. implementace modulu KP-WIN OPAC,
6. databáze Digitální knihovna,
7. přechod monografií (DB KPSYS 1),
8. přechod periodik (DB KPSYS 1),
9. přechod audiovizuálních dokumentů (DB KPSYS 1),
10. přechod příručky komeniologických monografií (DB KPSYS 2),
11. přechod databáze Sukovy studijní knihovny literatury pro mládež (DB ISIS 2),
12. přechod výpůjčního modulu včetně databáze čtenářů a výpůjček (DB KPSYS 1),
13. implementace modulu KP-WIN REVIZE.

## 2 Dílčí kroky implementace AKS KP-WIN

Přechod na AKS KP-WIN se ukázal jako poměrně složitý a následující části příspěvku se snaží v zásadních bodech postihnout jeho nejdůležitější části a roli jednotlivých standardních nebo speciálních funkcí AKS KP-WIN, které přechod dat umožnily.

### 2.1 Český pedagogický tezaurus (dále jen tezaurus)

Aby mohlo dojít k převodu první části dat, ukázalo se jako nezbytné importovat do KP-WIN nejprve tezaurus, na který bylo třeba navázat jednotlivé záznamy bází PBD a KOM. **Český pedagogický tezaurus** vznikl na začátku 90. let minulého století na základě překladu Evropského pedagogického tezauru, frekvenční analýzy deskriptorů tezauru pro oblast výchovy a vzdělávání vytvořeného v letech 1976–1980 a dalších termínů vzešlých z potřeb indexační praxe.<sup>2</sup>

V databázi ISIS byly na stejné „úrovni“ uloženy záznamy deskriptorů i nedeskriptorů. To se ukázalo jako problém při přechodu na KP-WIN. Tato situace si vyžádala mimořádné úpravy importovacího programu, který při importu záznamů rozhodoval podle speciálního algoritmu, zda daný záznam do báze uloží, nebo neuloží. Kromě toho bylo nutné ignorovat části záznamů tak, aby nevznikly zvláštní duplicity dané opakováním nadřazených či podřazených hesel u jednotlivých deskriptorů či tzv. asociovaných termínů.

V době, kdy vznikal a rozvíjel se Evropský pedagogický tezaurus, měly jednotlivé státy, které na jeho tvorbě spolupracovaly, možnost doplnit své „národní“ údaje. Do tezauru se tak dostaly názvy institucí, lokálních/regionálních systémů pro zpracování dat a mimo jiné i řada geografických termínů.

Při přechodu na KP-WIN se všechny tyto termíny ukázaly jako nadbytečné pro Pedagogický tezaurus a tuto situaci bylo třeba řešit: z tezauru byly zcela odstraněny mikrotezaury 38–42, které obsahovaly výhradně geografické termíny, a byly včetně

---

<sup>2</sup> Vybrané lexikální jednotky byly rozděleny na deskriptory a nedeskriptory a uspořádány do deskriptorových odstavců. Nedeskriptory byly doplněny odkazy a deskriptory byly podle potřeby upřesněny vysvětlivkami a zařazeny do mikrotezaurů. Tezaurus byl vypracován v souladu s normou ČSN 01 0193 *Dokumentace. Pokyny pro vypracování a rozvíjení jednoznačných tezaurů* a mezinárodní normou ISO 5964 – 1985 *Pokyny pro vypracování a rozvíjení vícejazyčných tezaurů* a po oponentuře na MŠMT ČR byl doporučen k užívání v dokumentační praxi resortu školství. V roce 1994 vyšlo 1. tištěné vydání tohoto tezauru, které kromě úvodu s vysvětlením struktury tezauru a návodem na jeho používání obsahovalo abecední strukturovaný rejstřík, rotovaný abecední rejstřík a graficky zpracovaný přehled jednotlivých mikrotezaurů. Mikrotezaury (v současnosti celkem 38 tematických oblastí) jsou abecedně uspořádané deskriptory s graficky znázorněnými vzájemnými vztahy. Do databázové podoby byl tezaurus převeden v polovině 90. let. Všechny deskriptory mají anglické ekvivalenty.

hierarchických úrovní převedeny do připravovaného souboru geografických autorit. Rovněž došlo k značné redukci mikrotezauru 37, který obsahoval konkrétní regionální, národní i mezinárodní instituce – tyto záznamy byly převedeny do připravovaného souboru jmenných autorit korporací. Odstranění hesel z mikrotezaurů se provádělo novou speciální funkcí v KP-WIN, která „odvázala“ zrušené záznamy deskriptorů či nedeskriptorů od tzv. asociovaných termínů (vazby dané odkazem typu „viz též“). Po těchto „očistách“ bylo možné importovat tezaurus v konečné verzi tak, aby splňoval standardní kritéria pro budování selekčních jazyků tohoto typu. Po importování záznamů a kontrole byly z názvů mikrotezaurů vytvořeny nejvyšší hierarchické úrovně tak, aby celý tezaurus fungoval jako strom. Tyto nejvyšší úrovně v ISIS nebyly potřeba, protože nebylo možné zobrazit graficky stromovou strukturu.

Potřeba usnadnit uživatelům vyhledávání v databázích vedla NPKK k vytvoření projektu s názvem *Český pedagogický tezaurus – 3. verze. Doplnění chybějících hierarchických úrovní tezauru a úprava webové aplikace pro zobrazení stromové struktury tezauru a pro možnost vícejazyčného vyhledávání*, jež knihovna v roce 2005 řeší. Cílem projektu je:

- vytvoření 3. verze Českého pedagogického tezauru a doplnění chybějících hierarchických úrovní tezauru;
- vytvoření webového rozhraní pro možnost vkládání návrhů na doplňky, opravy, resp. aktualizace deskriptorů i nedeskriptorů;
- zpřístupnění 3. verze tezauru ve stromové a zároveň vícejazyčné podobě na webu NPKK, včetně vazeb na záznamy dokumentů (knih, článků z časopisů, statí ze sborníků a recenzí), které NPKK spravuje ve svém knihovním fondu.

## **2.2 Pedagogická bibliografická databáze – PBD a Komeniologická databáze – KOM**

Jakmile byl v KP-WIN připravený tezaurus, přikročilo se k importu dvou analytických bází – PBD a KOM. Databáze PBD vznikla na přelomu let 1989/90, báze KOM v roce 1993 a jejich datová struktura se příliš nepodobala formátům používaným během 90. let (Výměnný formát, Unimarc atd.). I přes omezení daná datovou strukturou byly záznamy díky speciálnímu exportu do struktury Unimarc předávány v rámci spolupráce do báze ANL NK ČR.

Při vodu dat do KP-WIN se ukázala nutnost zásahy a úpravy těchto dat, aby v novém systému data odpovídala současným bibliografickým formátům. Proto byly pro potřebu importu dat NPKK připraveny tři speciální importy (pro články a statí, pro recenze a pro monografie), které v závislosti na typu importovaného dokumentu prováděly „opravy a úpravy“ dat, a to zejména u dat tvořících selekční prvky.

V případě importu údajů věcného popisu byl importovací program upraven tak, aby kontroloval data z konkrétního pole pro tezaurus a v případě nalezené shody uložil údaje do pole 606 (Unimarc) pro tezaurus a ostatní data pak do pole 610 pro tzv. volně tvořené předmětové termíny. Při analýze prvního zkušební vzorku se však ukázalo, že do pole 610 jsou ukládány údaje všech věcných kategorií – od jmen osob i korporací přes geografické či chronologické údaje, názvy projektů, názvy děl J. A. Komenského až po nezařaditelné volně tvořené termíny.

Proto bylo po zvážení situace rozhodnuto o kroku, se kterým se původně nepočítalo a jehož řešení zcela záviselo na schopnostech a možnostech jak na straně knihovny, tak na straně KP-WIN. Toto rozhodnutí znělo:

1. Vyexportovat veškeré údaje z pole 26 (z DB ISIS), ve kterém byl veškerý věcný popis obou bází, a tyto údaje roztřídit a vytvořit z nich jakési „předchůdce“ připravovaných souborů autorit;
2. poté upravit importovací program tak, aby importem a kontrolou přes jednotlivé autoritní

rejstříky program roztřídil a uložil data věcného popisu podle jednotlivých kategorií.

Výše popsaným exportem pole 26 s veškerým věcným popisem vznikl seznam obsahující cca 600 000 použitých hesel věcného popisu. Speciálními skripty (skriptovací jazyk AWK) byly odstraněny duplicity a vznikl seznam cca 14 000 unikátních termínů, mezi nimiž byla i „oficiální“ hesla pedagogického tezauru.

Dalšími skripty byla ze seznamu odstraněna hesla tezauru a poté nastal proces nejprve automatizovaného a posléze i ručního třídění použitých hesel do jednotlivých kategorií. Z cca 11 500 hesel bylo automatizovaně vytříděno cca 3 500 osobních jmen, stovky, resp. tisíce korporací, chronologických i geografických údajů. Zbylá neroztříděná hesla byla předtříděna podle odhadovaných kategorií a v podobě seznamů v excelových souborech rozdána mezi pracovníky oddělení bibliografie. Ti podle přesného zadání označili každé jednotlivé heslo k zařazení do příslušné kategorie. Takto bylo zpracováno několik tisíc hesel, která byla nakonec podle kategorií setříděna, sloučena do výsledných seznamů, a ty byly importovány jako „předchůdci“ připravovaných souborů autorit.

Následně byl importovací program KP-WIN doplněn o speciální algoritmus, který v případě, že se dostal k importu údajů z pole 26, tedy věcného popisu, ověřoval hesla postupně v jednotlivých autoritních rejstřících. Ověřená hesla ukládal do konkrétních polí a pouze ta hesla, která „propadla sítí“ všech autoritních rejstříků, byla uložena do pole 610 jako volně tvořené předmětové termíny.

Ještě před finálním importem dat byly doplněny vytvořené rejstříky o nepreferované formy/nedeskriptory autoritních záznamů. Tato úprava byla potřebná zejména v případě tezauru, který byl doplněn o termíny používané ve starších bibliografických záznamech. Speciální algoritmus importovacího programu byl mj. nastaven tak, že kromě vyhledávání v deskriptorech hledal i ve všech nepreferovaných formách; tak bylo možno uskutečnit importem i věcné opravy, které v databázovém prostředí ISIS nebyly proveditelné. Opravy v ISIS musely totiž být (stejně jako v obdobných systémech) prováděny přímo v jednotlivých záznamech a nebylo reálné takto postupovat (vzhledem k četnosti výskytů) v případě zastaralých či již nepreferovaných termínů.

Následkem všech výše popsaných speciálních úprav bylo, že import jednoho záznamu trval cca 12–13 sekund. To by při cca 100 000 záznamech bází PBD a KOM znamenalo takřka 14 dní nepřetržitého importu. Proto došlo k další úpravě importovacího programu, který byl upraven (optimalizován) tak, že import jednoho záznamu trval cca 1,5 sekundy a cca 86 000 záznamů článků z PBD bylo importováno za necelý víkend. Se stejnou rychlostí pak byly importovány i další typy dokumentů – stati a recenze z bází PDB a KOM.

Následně po importu se uskutečnilo finální doladění pracovních postupů a na základě požadavku NPKK i vývoj nové funkce v KP-WIN. Funkce při svém zadání dostala pracovní název „časopisecká autorita“ a jejími základními cíli bylo:

- zajistit propojení záznamů titulů periodik s jednotlivými analytickými záznamy;
- umožnit přebírání údajů tzv. zdrojového dokumentu do analytických záznamů.

Aby tato funkce mohla pracovat podle zadání, bylo nutné k jednotlivým titulům časopisů doplnit údaj (na bázi identifikátoru), jehož prostřednictvím by bylo možné záznamy propojovat. Bylo posouzeno několik možností – čárové kódy, identifikační čísla – a nakonec byla zvolena varianta tzv. evidenčního čísla. Toto číslo je unikátní pro každé evidované číslo časopisu bez ohledu na duplicitu výtisků. Číslo se skládá ze dvou částí: část před lomítkem tvoří evidenční číslo titulu časopisu a část za lomítkem generuje systém jako pořadové číslo bez jakéhokoli dalšího významu nebo vztahu k jednotlivým odebíraným ročníkům.

Evidenční čísla jsou napsána na obálku všech evidovaných časopisů a využita při analytickém zpracování titulů excerpovaných pro báze PBD a KOM. Při založení každého nového záznamu PBD a KOM je prostřednictvím funkční klávesy F12 vyvoláno okno pro zadání evidenčního čísla a po jeho potvrzení jsou do záznamu přeneseny veškeré dostupné údaje konkrétního zdrojového dokumentu – s výjimkou údaje o lokaci článku, který je doplněn v rámci analytického zpracování.

Tato funkce rovněž umožnila provázat tituly a analytické záznamy v prostředí OPAC. Po vyhledání konkrétního titulu si uživatel může kliknutím na odkaz „články“ zobrazit záznamy článků zpracovaných z daného čísla časopisu, které byly z obsahu časopisu vybrány jako relevantní pro báze PBD a KOM.

### 2.3 Tvorba databáze periodik

Současně s přípravami importu PBD a KOM bylo nutné vytvořit databázi časopisů pro evidenci docházejících čísel a pro možnost propojení titulů s analytickými záznamy. Databázi bylo třeba vybudovat od základů, protože NPKK dosud automatizovaně zpracovávala pouze vázané svazky časopisů a evidence odebíraných titulů a došlých čísel byla vedena na klasických kartách.

NPKK se proto rozhodla mezi dvěma řešeními: buď zpracovat záznamy minimálně všech v současnosti odebíraných titulů časopisů, nebo převzít záznamy z externího zdroje. Proto byla prověřena možnost převzít záznamy ze Souborného katalogu časopisů (dále jen SK ČR), odkud bylo možné vybrat záznamy podle sigly knihovny a vyexportovat v datové struktuře Unimarc. Při posuzování obou možností byla nakonec zvolena možnost převzetí záznamů ze SK ČR. Celkem bylo převzato 921 záznamů časopisů.

Při podrobné analýze se však ukázalo, že bude nutno převzaté záznamy před importem do KP-WIN upravit. Pomocí skriptů AWK byla odstraněna řada polí, především z oblasti poznámek, protože údaje v nich uvedené popisovaly spíše stav titulu v konkrétních knihovnách než obecné údaje, které patří do báze typu souborný katalog.

Po skončení všech úprav došlo k importu dat a následně bylo možné zahájit provoz této části databáze včetně evidence došlých čísel časopisů. Na data pak byly navázány doplňující funkce umožňující např. propojení více typů záznamů.

### 2.4 Implementace modulu KP-WIN OPAC

Následně po importu dat a doladění linky zpracování se uskutečnila implementace modulu OPAC.

V době psaní tohoto příspěvku jsou prostřednictvím webového OPAC dostupné tyto základní báze:

**A. Pedagogická bibliografická databáze (PBD)** obsahuje anotované dokumentační záznamy české i zahraniční článkové literatury. Databáze obsahuje více než 90 000 záznamů (články, stati, recenze) od roku 1990 do současnosti, excerpční základnu tvoří zhruba 230 titulů periodik.

**B. Databáze komeniologické literatury (KOM)** obsahuje dokumentační záznamy knih a časopiseckých článků o životě a díle J. A. Komenského, záznamy o různých vydáních jeho prací. Databáze je vytvářena od roku 1993 a v současné době obsahuje přibližně 6000 záznamů.

**C. Databáze časopisů** (záznamy časopisů odebíraných NPKK): u aktuálně odebíraných titulů časopisů je k dispozici přehled došlých čísel a u titulů, které jsou excerpovány pro

PBD a KOM, je možné po kliknutí na odkaz „Články“ (uvedený u každého došlého čísla) zobrazit záznamy článků zpracovaných z daného čísla časopisu.

## 2.5 Databáze Digitální knihovna

Na tři základní báze navazuje zcela nová databáze **Digitální knihovna**. Tato databáze obsahuje ve zkušebním provozu 280 z cca 700 digitalizovaných historických učebnic dějepisu, čítanek a slabikářů (formát PDF) z fondu ÚIV – NPKK, které již nejsou chráněny autorským zákonem.

## 2.6 Další procesy

- přechod monografií (DB KPSYS 1),
- přechod periodik (DB KPSYS 1),
- přechod audiovizuálních dokumentů (DB KPSYS 1),
- přechod příručky komeniologických monografií (DB KPSYS 2),
- přechod databáze Sukovy studijní knihovny literatury pro mládež (DB ISIS 2),
- přechod výpůjčního modulu včetně databáze čtenářů a výpůjček (DB KPSYS 1),
- implementace modulu KP-WIN REVIZE.

Tyto dílčí procesy jsou v době psaní příspěvku v různých stadiích řešení a jejich popis bude k dispozici během semináře AKP 2005 a na webu NPKK.

## 3 Závěr

Přechod knihovny do jednotného databázového prostředí je přínosem jak pro pracovníky knihovny, tak pro její uživatele.

**Pracovníkům** tento krok umožňuje:

- prostřednictvím moderních informačních technologií zpřístupnit fondy/činnosti knihovny, jež byly dosud „utajeny“ kvůli softwarovým omezením daným nepropojenými a nekompatibilními databázemi;
- odstranit duplicitní vkládání údajů na různých pracovištích knihovny;
- možnost zapojit se do systémů spolupráce knihoven na budování národních souborů autorit a souborných katalogů.

**Uživatelům** tento krok umožňuje:

- v prostředí jedné on-line databáze na základě rešeršního dotazu získat komplexní informace o primárních i sekundárních dokumentech uložených v knihovním fondu NPKK, včetně možnosti vyhledávání zadaného tématu prostřednictvím internetového vyhledávače;
- on-line objednávat výpůjčky literatury z fondu NPKK i další služby z prostředí OPAC.

Realizace jednotného databázového prostředí v NPKK je prvním krokem k plnohodnotnému zpřístupnění celé knihovny uživatelské veřejnosti. Umožní zpřístupnění kvalitních a zároveň jedinečných dat, která dokumentují vývoj české pedagogiky, výchovy, vzdělávání a školství od konce 19. století až do současnosti, a na ně navazujících služeb.