

# DOZP 1.1 – export obálek a obsahov

---

Prepojenie je na základe systémového čísla, ktoré je vyhľadané cez čiarový kód spracovávanej knihovnej jednotky. Zo záznamu sú získané aj ďalšie údaje pre webovú aplikáciu (autor, názov, rok), údaje pre import na server ObalkyKnih.cz (ISBN, ISNN, EAN, NBN, OCLC), ale tiež pre import do centrálneho katalogu (ISBN pre viacväzkové publikácie).

Obálky nebudú predávané priamo do ALEPHu, ale budú po naskenovaní okamžite odoslané na server ObalkyKnih.cz. Obsahy následne prejdú na OCR spracovanie, kde výstupom bude dvojvrstvý PDF súbor a zformátovaný text obsahu.

Export spracovaných obsahov bude prebiehať vždy raz za den v požadovaný čas (napr. v noci 23:00) a výsledok bude uložený na FTP do odpovedajúcej složky katalogu (napr. CKS01) na serveri FFAS04. Pre každý export bude vytvorená samostatná podsložka pomenovaná podľa časového údaju v dobe exportu tj. RRRRMMDD, ktorá bude obsahovať jednotlivé súbory vrátane log súboru obsahujúceho popis exportovaných súborov. Po naimportovaní do centrálneho katalogu bude zapísaná informácia do log súboru, že import do ALEPHu úspešne prebehol.

```
C:\FTP\  
    CKS01\  
        20150201\  
        20150202\  
        20150203\  
            000000642_8070669683.jpg  
            000000642_8070669683.pdf  
            000000642_8070669683.txt  
            000126355.pdf  
            000126355.txt  
            ...  
            Export.log  
...
```

Pomenovanie súborov je vždy na základe systémového čísla (SysNo) a odpovedajúceho záznamu z CKIS a ISBN (nemusi obsahovať), ktoré budú oddelené podtržítkom. Tento spôsob označenia je dôležitý hlavne pre záznamy, ktoré obsahujú viac dielov. Do adresára je vždy pridaný Export.log súbor, ktorý obsahuje nasledovné položky:

- **SysNo** – systémové číslo záznamu (pole 001/SYS) načítané z centrálneho katalogu (CKIS),
- **Názvové údaje** – obsahuje popis publikácie ako sú autor (pole 100\$c), názov (pole 245\$a) a rok vydania (pole 260\$c), tieto údaje slúžia len pre lepšiu kontrolu a identifikáciu záznamu,
- **Díl, svazek** – text zadany užívateľom, ak sa jedná o viacväzkovú publikáciu, tento text sa bude automaticky prebrať a uvádzať v popise záznamu pre pomenovanie obsahu/digitálneho objektu ALEPHu (obr.),
- **Súbor obálky** – názov súboru naskenovanej obálky, uvádza sa len v prípade, že bola obálka importovaná na server ObalkyKnih.cz (zároveň by slúžil pre zalohovanie mimo server),
- **Súbor obsahu** – názov súboru spracovaného obsahu (može byť aj bez textovej vrstvy),
- **Súbor s OCR textom** – názov textového súboru, ak obsah nebol spracovaný pomocou OCR, ale len naskenovaný, tak sa položka neuvádza (text bude vkladany do pola 505\$a).

jednotlivé položky sú v textovom súbore oddelené znakom „|“ zvislá čára (ASCII kód 124).

Postup pri importe do ALEPHu by bol taký, že by bola vždy dávkovo pomocou skriptu importovaná celá složka spracovaná v daný deň, najprv by bol načítaný Export.log súbor a podľa jeho obsahu by boli importované odpovedajúce súbory a pridané texty obsahov do záznamov v CKIS.

## Príklad:

Obsah log súboru: Export.log

000624936 | Art brut, Sbírka abcd, 2006 | | 000624936\_8023972405.jpg | 000624936\_8023972405.pdf |  
000624936\_8023972405.txt  
001842310 | Gebhard, A. C.: Kinderboeken, 1959 | 1.díl | 001842310\_9029070608.jpg | 001842310\_9029070608.pdf |  
001842310\_9029070608.txt  
001842310 | Gebhard, A. C.: Kinderboeken, 1959 | 2.díl | | 001842310\_9029070609.pdf | 001842310\_9029070609.txt  
001404447 | Ikona v ruském myšlení 20. století, 2011 | | | 001404447.pdf |  
...

Textový OCR súbor: 000624936\_8023972405.txt

Obsah: -- Kapitola 1: Paže -- Kapitola 2: Hrudník -- Kapitola 3: Ramena -- Kapitola 4:  
Břicho a bedra -- Kapitola 5: Záda -- Kapitola 6: Nohy -- Kapitola 7: Srdce -- Kapitola  
8: Strečink

textový súbor je formátovaný tým spôsobom, že znaky konca riadkov (CR/LF) sú vždy nahradené „--“ dvojicou pomlčiek.  
Odstránené bude stránkovanie a pokiaľ to bude možné aj číslovanie kapitol.

Varianta č.2 by mohol byť XML formát, kde jeho súčasťou by bol aj celý OCR text, podobne poskytuje texty obsahov  
server ObalkyKnih.cz

Názov súboru: Export.xml

```
<?xml version="1.0" encoding="utf-8" ?>
<Book>
  <SysNo>000624936</SysNo>
  <ISBN>978-80-239-7240-5</ISBN>
  <BibInfo>Art brut, Sbírka abcd, 2006</ BibInfo >
  <Volume> </Volume>
  <Cover>000624936_8023972405.jpg</Cover>
  <TOC>000624936_8023972405.pdf</TOC>
  <OCR> Obsah: -- Kapitola 1: Paže -- Kapitola 2: Hrudník -- Kapitola 3: Ramena -- Kapitola 4: Břicho a bedra -- Kapitola 5: Záda
-- Kapitola 6: Nohy -- Kapitola 7: Srdce -- Kapitola 8: Strečink
  </OCR>
</Book>
<Book>
...
</Book>
```

---

Ďalšia navrhovaná varianta je úplne vypustiť predávanie dát cez FTP a realizovať celú komunikáciu medzi serverom  
s obsahmi **FFAS04** a **ALEPH**om priamo pomocou HTTP/RESTful webovej služby (ASP.NET Web API) vo formáte XML alebo  
JSON.

# Úplné zobrazení záznamu

**Zvolte formát:** [Standardní formát](#) [Katalogizační lístek](#) [Zkrácený záznam](#) [S návěštími polí](#) [S kódy polí MARC](#)

**Záznam 1 z 2**

[◀ Předchozí záznam](#) [▶ Další záznam](#)

Číslo záz.	001590712
ISBN	978-80-85005-83-7 (1. díl : váz.) 978-80-7404-011-5 (2. díl : váz.) 978-80-7404-044-3 (3. díl : váz.)
Název	● <a href="#">O češtině / [Aleš Cibulka ... et al. ; editorka Jitka Saturková]</a>
Vydání	Vyd. 1.
Naklad. údaje	Praha : Česká televize, [2007]-
Popis	sv. : barev. il., portréty ; 22 cm
Edice	● <a href="#">Edice České televize</a>
VedlZ-osobní jm.	● <a href="#">Cibulka, Aleš, 1977-</a> ● <a href="#">Saturková, Jitka</a> ● <a href="#">Pravdová, Markéta, 1974-</a>
Poznámka	Editorkou 2. a 3. dílu též Markéta Pravdová
Bibliogr.	Obsahuje bibliografie a bibliografické odkazy
Neúplný obsah	[sv. 1.] 2007. 198 s. -- [sv.] 2. 2008. 195 s. -- [sv.] 3. 2010. 185 s.
Jednotky-vše	Všechny jednotky
Jednotky-DK	FFUK Knihovna Jana Palacha 
El.umístění	 <a href="#">Obsah dokumentu sv. 2</a>  <a href="#">Obsah dokumentu sv. 3</a>
O názvu	● <a href="#">O češtině (televizní pořad)</a>
Výraz tez.-dle \$2	● <a href="#">televizní pořady -- Česko</a> ● <a href="#">čeština</a>
Žánr/forma	● <a href="#">rozhovory</a> ● <a href="#">populárně-naučné publikace</a>
Skupina Konspektu	811.162.3 - Čeština



[http://ckis.cuni.cz/F/?func=direct&doc\\_number=001590712&local\\_base=CKS01&format=999](http://ckis.cuni.cz/F/?func=direct&doc_number=001590712&local_base=CKS01&format=999)