

# Zkušenosti NK ČR s implementací EBSCO Discovery Service (EDS)

**Karolína Košťálová, Petra Šťastná**

Národní knihovna ČR

[karolina.kostalova@nkp.cz](mailto:karolina.kostalova@nkp.cz), [petra.stastna@nkp.cz](mailto:petra.stastna@nkp.cz)

INFORUM 2013: 19. ročník konference o profesionálních informačních zdrojích

Praha, 21. -22. 5. 2013

## **Abstrakt**

*NK ČR v roce 2012 zakoupila EDS v návaznosti na zkušenosti z předcházejícího cca půlročního zkušebního přístupu k EDS. Příspěvek přibližuje kroky, které byly nezbytné pro nastavení zmíněného zkušebního, respektive řádného přístupu k EDS. Zmíněna bude zejména problematika výběru lokálních zdrojů zapojených do EDS, požadavky na součinnost na straně těchto zdrojů, ladění prezentace informací o fondu NK ČR v prostředí EDS, definování licencovaných zdrojů, statistiky využití atd.*

## **Úvod**

Integrace více zdrojů do jednoho rozhraní není zcela novou myšlenkou. Se zastřešením různorodých zdrojů v jednom rozhraní při využití paralelního vyhledávání má NK ČR bohaté zkušenosti díky více než desetiletému provozu Jednotné informační brány založené na vyhledávání pomocí MetaLib. Na konci roku 2012 bylo spuštěno alternativní rozhraní JIB+, které kombinuje discovery rozhraní s vyhledáváním ve volně přístupném centrálním indexu a s paralelním hledáním pomocí MetaLib. V rámci JIB byla také pozornost věnována bližšímu zkoumání a srovnání vybraných discovery systémů. Pravému discovery systému založenému pouze na lokálních zdrojích byl pak blíže projekt ANL+, v jehož rámci byl vytvářen lokální index a ANL+ byla nabízena uživatelům v rozhraní Primo. Dalším krokem se nyní v oblasti integrace zdrojů pro NK ČR stal nákup EBSCO Discovery Service.

## **EBSCO Discovery Service (EDS)**

EDS patří do skupiny tzv. web-scale discovery systémů. Výhodou těchto systémů je jednoduché a uživatelsky přívětivé prostředí integrující v jednom rozhraní pokud možno všechny zdroje knihovny, počínaje knihovními katalogy a konče elektronickými zdroji. Základním kamenem web-scale discovery systémů je tzv. jednotný index (někdy je také označován jako centrální nebo velký), který obsahuje data z licencovaných i volně dostupných elektronických informačních zdrojů. Vyhledávání prostřednictvím indexu přináší uživatelům web-scale discovery systémů řadu výhod, např. rychlejší zobrazování nalezených výsledků z různých zapojených zdrojů či snadnější určení míry relevance vyhledaných dokumentů.

## **EDS v Národní knihovně ČR**

Národní knihovna ČR (NK ČR) měla k dispozici zkušební přístup do EDS zhruba od srpna 2011 do ledna 2012. Nastavení a testování probíhalo už od jara 2011. Šlo především o nastavení lokálních zdrojů. V případě zkušebního přístupu NK ČR byl do EDS integrován elektronický katalog – báze

NKC. Součástí centrálního indexu EDS se stala také digitální knihovna Manuscriptorium; smlouva o jejím zpřístupnění v rámci nabídky produktů EBSCO Publishing byla s Národní knihovnou ČR podepsána v květnu 2011. V roce 2012 pak NK ČR nákupem EDS pro období 2012-2015 navázala na tento půlroční zkušební přístup. Do ostrého provozu bylo EDS na základě této smlouvy uvedeno v květnu 2012.

## EDS a licencovaný obsah

Obsah EDS pro uživatele NK ČR je v současné době možné pomyslně rozdělit do dvou skupin. První tvoří centrální index EDS (uvádí se jako EDS Foundation Index nebo EDS Base Index, případně EDS Index), který obsahuje metadata od desítek tisíc poskytovatelů obsahu (cca 20.000 producentů dat a 70.000 knižních vydavatelů), repozitáře typu open access (DOAJ, OAISTER, arXiv.org). Druhou skupinu tvoří lokální zdroje knihovny.

V případě první skupiny byly v prostředí EDS dle možností centrálního indexu EDS pro NK ČR aktivovány licencované zdroje, které knihovna předplácí. Pro samotné vyhledávání jsou uživatelům NK ČR v EDS samozřejmě dostupné i zdroje, které NK ČR nepředplácí (pokud to licenční podmínky uzavřené mezi poskytovatelem databáze a EBSCO umožňují), přístup k plným textům je již limitován na zdroje dostupné pro NK ČR.

V prostředí EDS mohou uživatelé NK ČR využívat i SFX služby tak, jak jsou zvyklí např. z Jednotné informační brány.

## Provázání plných textů z různých zdrojů

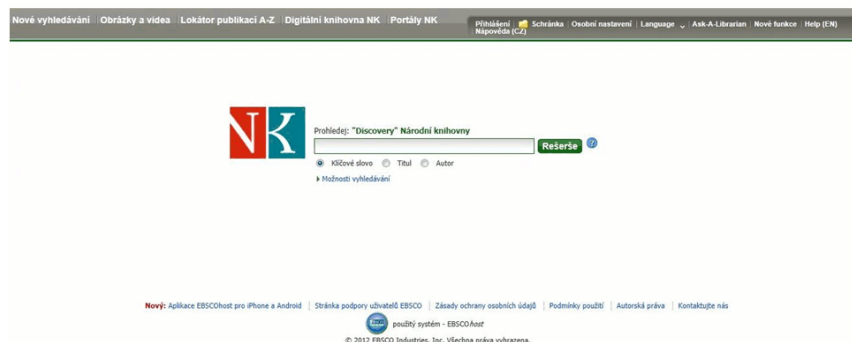
Dostupnost plných textů ze zdrojů, které nejsou součástí databází EBSCOhost, je v EDS možná po aktivaci propojení do konkrétních databází. Tímto způsobem byly pro NK ČR ošetřeny přístupy do databází ScienceDirect, Wiley Online Library, SpringerLink či Web of Science (viz obr. 1).

The screenshot displays the EDS search interface. At the top, there is a search bar with the text "Discovery: Národní knihovny" and "web scale discovery". Below the search bar, there are search results. The first result is "Review: Systems chemical biology and the Semantic Web: what they mean for the future of drug discovery research" by David J. Will, Ying Ding, Amel P. Sheth, Lee Harland, Eric H. Giffard, Michael S. Lapines. The second result is "LinkedWS: A novel Web services discovery model based on the Metaphor of 'social networks'" by Zakaria Maamar, Leandro Krug Wives, Youakim Badi, Saïd Elneffar, Khaled Boukad, Moura Fati. Below the search results, there is a section for "Simulation Modelling Practice and Theory" with a full-text link to a ScienceDirect article. The article title is "LinkedWS: A novel Web services discovery model based on the Metaphor of 'social networks'". The article is by Zakaria Maamar, Leandro Krug Wives, Youakim Badi, Saïd Elneffar, Khaled Boukad, Moura Fati. The article is published in the Journal of Logic and Algebraic Programming, Volume 61, Issue 1, January 2011, Pages 121-132. The article is available as a PDF (335 KB). The article abstract is: "Web services are increasingly becoming the de facto implementation for the service-oriented architecture paradigm for enterprises due to their ease of use. Nevertheless, discovering these Web services is still hindered by many challenges that are partially attributed to shortcomings found in the discovery registry models (e.g., UDDI) used nowadays. These registries do not, for example, capture the rich information resulting from the various types of interactions between Web services. To address these shortcomings, and inspired by the conventional human social networks on the net, via Facebook and Twitter, we develop LinkedWS, a social networks discovery model to capture the different interactions that occur between Web services. Based on these interactions, specialized relationships are spawned and discovered. Examples of these relationships are collaboration and substitution. The paper describes LinkedWS and its potential, and reports on its implementation status."

Obr. 1. Odkaz na plný text v databázi ScienceDirect

## EDS a lokální zdroje

Druhou skupinu v EDS představují lokální zdroje Národní knihovny ČR. Jak již bylo zmíněno, součástí indexu EDS je od května 2011 digitální knihovna Manuscriptorium. Pro zkušební přístup byly do EDS dále importovány záznamy z elektronického katalogu NK ČR - báze NKC. Na základě uzavřené tříleté smlouvy se pak dalším lokálním zdrojem NK ČR v EDS stal Souborný katalog ČR (SKC) a v současné době je pro import do EDS připravována i Digitální knihovna Kramerius NK ČR.



Obr. 2. Vstupní obrazovka EDS pro NK ČR

Velká část činností, které předcházely uvedení EDS v NK ČR do ostrého provozu v květnu 2012, plynule navázala na předchozí spolupráci z roku 2011 související se zkušebním přístupem NK ČR k EDS.

V roce 2012 byla řešena zejména integrace Souborného katalogu ČR a dále pak ve spolupráci s technikou EBSCO zejména detaily, které souvisely s ostrou integrací NKC do EDS (zobrazování dostupnosti jednotky, překlady názvů dílčích knihoven atd.).

## Integrace elektronického katalogu – báze NKC a SKC

Samotné integraci elektronického katalogu NK ČR (NKC) a posléze i Souborného katalogu ČR (SKC), předcházelo vyplnění dotazníků o technických aspektech. Dotazník zjišťoval, jaký systém knihovna využívá (Aleph verze 20), kolik obsahuje katalog bibliografických záznamů (NKC - 1.950.000 záznamů, Souborný katalog ČR – 5 000 000 záznamů, stav k 1.7.2011) a jak často budou do EDS zasílány aktualizace. U obouází jsou v současné době jednou týdně na FTP server EBSCO předávány informace o smazaných, nových a změněných záznamech. Aktualizace dat probíhá vždy v neděli v noci, pokud v NKC nebo SKC neproběhly v uplynulém týdnu velké změny (např. dávkové opravy záznamů), předává se FTP server EBSCO cca 5-10 000 záznamů z báze NKC a 30-60 000 záznamů z báze SKC.

## Metadata a propojení na záznamy

Další otázky se v dotazníku týkaly metadat (jednoznačný identifikátor záznamu, jazyk, formát dat) a propojení na záznamy (tj. syntax URL). Preferovaným formátem pro EDS je MARC21, importovat však lze i záznamy ve formátech UNIMARC, KORMARC či MARCXML. Z těchto možností byl pro export z NKC a SKC zvolen formát MARCXML.

Do EDS je předáván pouze standardní MARC záznam, tj. neexportují se např. národní pole 9XX, jedinou výjimkou je v NKC pole 996 pro lokální údaje. Drobné úpravy byly také provedeny v některých případech u polí pro věcný popis dokumentu.

Pro vytvoření přímého odkazu z EDS na konkrétní záznam v NKC, respektive SKC se používá předdefinovaná URL. Ta se v EDS doplňuje o údaj získaný ze záznamu, který daný záznam jednoznačně v bázi identifikuje a umožní tak vytvořit plně funkční odkaz. URL pro odkazy do NKC jsou v EDS doplňovány o údaj z pole 001. V Souborném katalogu ČR dochází k velkému množství změn, záznamy jsou slučovány, přepisovány, mazány atd. Odkaz z konkrétního záznamu v EDS do SKC tedy nebylo možné, na rozdíl od NKC, konstruovat s využitím pole 001 jako jednoznačného

identifikátoru záznamu. Při importu záznamů ze SKC do EDS se nahrazuje pole 001 tzv. systémovým číslem (pole SYS). Formát předdefinované URL je tak z pohledu EDS podobný pro NKC i SKC, jediným rozdílem mezi oběma bázemi je pouze údaj použitý pro identifikaci záznamu, tj. v bázi NKC je to pole 001, v SKC pak pole SYS.

V dotazníku byla věnována pozornost i identifikaci elektronických a zvukových knih a dostupnosti plného textu (pole pro jednoznačnou identifikaci, limity pro vyhledávání podle typu dokumentu, propojovací pole na plné texty). Elektronické knihy v NKC jsou slovně specifikovány v poli 655 \$a - elektronické knihy, zvukové knihy specifikovány nejsou, plný text je označen v poli 856 \$y - Plný text.

### **Lokace a informace o aktuální dostupnosti**

Dále dotazník zjišťoval, v jakých polích jsou uvedeny lokační údaje. V případě NKC byly pro potřeby EDS tyto údaje převedeny do pole 996 (\$h signatura, \$l umístění).

V nativním rozhraní NKC se uživatelé setkávají s řadou knihovních jednotek, které není možné z různých důvodů prostřednictvím NKC objednat. Při exportu do EDS byl proto nastaven filtr, který zajišťuje, že se do EDS nedostanou „neobjednatelné“ jednotky. V EDS tak uživatelé nenaleznou záznamy, respektive jednotky, které jsou v NKC označeny jako ztráty, jsou součástí služebních příručních knihoven nebo tzv. rezervních fondů, procházejí linkou zpracování atd.

Součástí záznamů z NKC je v EDS i informace o aktuální dostupnosti konkrétní jednotky. Pro zjištění dostupnosti v reálném čase EDS navazuje vždy spojení s NKC prostřednictvím protokolu Z39.50 na základě autentizačních parametrů (adresa serveru, port a jméno databáze).

Během testování zobrazení aktuální dostupnosti jednotek se při ostré integraci NKC do EDS objevovaly z počátku určité nesrovnalosti. U části dokumentů nebyly v EDS aktuální údaje dostupné a k dispozici zde byla pouze informace o signatuře, umístění dokumentu a dostupnosti svazku k datu, k němuž byl záznam předán do EDS. Z analýzy záznamů vyplynulo, že problémy způsobuje nastavení příliš nízkého limitu pro objem dat předávaných při komunikaci přes Z39.50 mezi NKC a EDS. V případě většího počtu jednotek se díky této blokadě nepodařilo prostřednictvím Z39.50 získat k záznamu informace o aktuální dostupnosti. Po zvětšení zmíněného limitu již nejsou se zobrazování aktuálních dat problémy.

V případě SKC není zatím součástí předávaných údajů pole 9XX s informací o lokaci. Tyto údaje EBSCO v současné době získává ze SKC dotazováním webového serveru, nicméně možné je dodávat lokace ze SKC do EDS i prostřednictvím vlastního profilu přes Z39.50.

### **Odlíšnosti záznamů vyhledaných v prostředí Aleph a EDS**

Výhodou vyhledávání v EDS je nalezení informace v kontextu s dalšími zdroji, které jsou v indexu EDS obsaženy, např. vyhledávání recenzí nalezeného titulu, dalších vydání nebo jiných forem dokumentu (např. videa). Pro dotazy, které nemohou být při hledání v NKC nebo SKC zodpovězeny – knihovna nemá titul, který by odpovídal požadavku uživatele, uživatel zadává např. klíčová slova v podobě, která se ve standardizovaném věcném popisu nenacházejí atd. – může EDS z dalších zapojených zdrojů najít relevantní dokumenty.

RUR. : Rosum's Universal Robots : Kollektiivdraama kolmes vaatuses sissejuhatava komöödiaga / [Autor:] Karel Čapek ; Tšehhi keelest tõlkinud [do eston.] Lembit Remmelgas, [kt. také naps. doslov] Karel Čapeki antiutoopia

Jazyk: Estonian

Autoři: Čapek, Karel, 1890-1938

Informace o publikaci: Tallinn : Perioodika, 1985 (EKP Keskkomitee Kirjastuse trükkkoda)

Datum publikování: 1985

Fyzický popis: 78, [2] s. ; 8°

Seriály: Loomingu; No 3 (1415) 1985

Typ publikace: Book

Typ dokumentu: Book

Poznámky: Alguparand: R.U.R.

Jiné tituly: RUR. Estonsky

Přirůstkové číslo: nkcr.bknzdr00519

Databáze: Katalog Národní knihovny ČR

Lokace	Signatura	Dostupnost	Vypůjčeno do
NÁRODNÍ KONZERVACÍ FOND	I 281240	Dostupné	-

Recenze tohoto titulu

1. **RUR or RU Ain't A Person?**  
By: Madigan, Tim. *Philosophy Now*, Mar/Apr2009, Issue 72, p48, 1p

Obr. 3. záznam z báze NKC v kombinaci s recenzí z indexu EDS

Oproti NKC a SKC se v EDS pro zobrazování obálek nepoužívá tzv. obálkový server, ale zahraniční zdroje. V případě českých dokumentů může být méně záznamů doplněno o tuto informaci. Užitečnou službou, kterou uživatelům NKC a potažmo i SKC nabízejí, je obohacování záznamů o naskenované obsahy. Odkazy na obsahy naleznou uživatelé také v prostředí EDS. Část naskenovaných obsahů prošla i OCR. V prostředí Aleph lze tedy hledat nejen v bibliografickém záznamu, ale i v textu, který vznikl na základě OCR. Tuto možnost EDS nemá vzhledem k tomu, že zmíněný text není přímou součástí MARC záznamu dodávaného do EDS.

Obsah bází NKC a SKC se do určité míry shoduje, tj. záznam publikace nalezne uživatel jak v katalogu NK ČR, tak je dohledatelný v Souborném katalogu ČR. V současné době nejsou v prostředí EDS záznamy z NKC a SKC deduplikovány, EDS je však schopno tuto funkci nabídnout.

### EDS nejen pro registrované uživatele NK ČR

Výhodou discovery systémů je kromě velkého objemu dat integrovaného do jednoho indexu i přívětivé rozhraní, v němž se uživatel může pohybovat jednoduše a intuitivně, bez nutnosti jeho ovládání předem dlouho zkoumat. NK ČR již řadu let předplácí pro své uživatele databáze na platformě EBSCOhost. Vzhledem k tomu, že design a ovládání EDS vychází právě z této platformy, může se jeho ovládání stát pro řadu uživatelů naprostou samozřejmostí.

Obdobně jako další discovery systémy se i EDS otevírá všem zájemcům o vyhledávání. Režim typu Guest/Host je otevřený pro vyhledávání komukoli, při požadavku na zobrazení záznamu z licencovaného zdroje je však uživatel již vyzván, aby se identifikoval prostřednictvím Centrální autentizační služby NK ČR. Nevýhodou přístupu typu Guest/Host tak zůstává pouze fakt, že EDS po identifikaci uživateli nenabídne možnost pokračovat v předchozích vyhledávaních a dotaz je tak nutné znovu zadat. Druhý režim EDS již od počátku vyžaduje, aby uživatel k EDS přistupoval buď z IP adres NK ČR, nebo se přihlásil prostřednictvím zmíněné Centrální autentizační služby NK ČR.

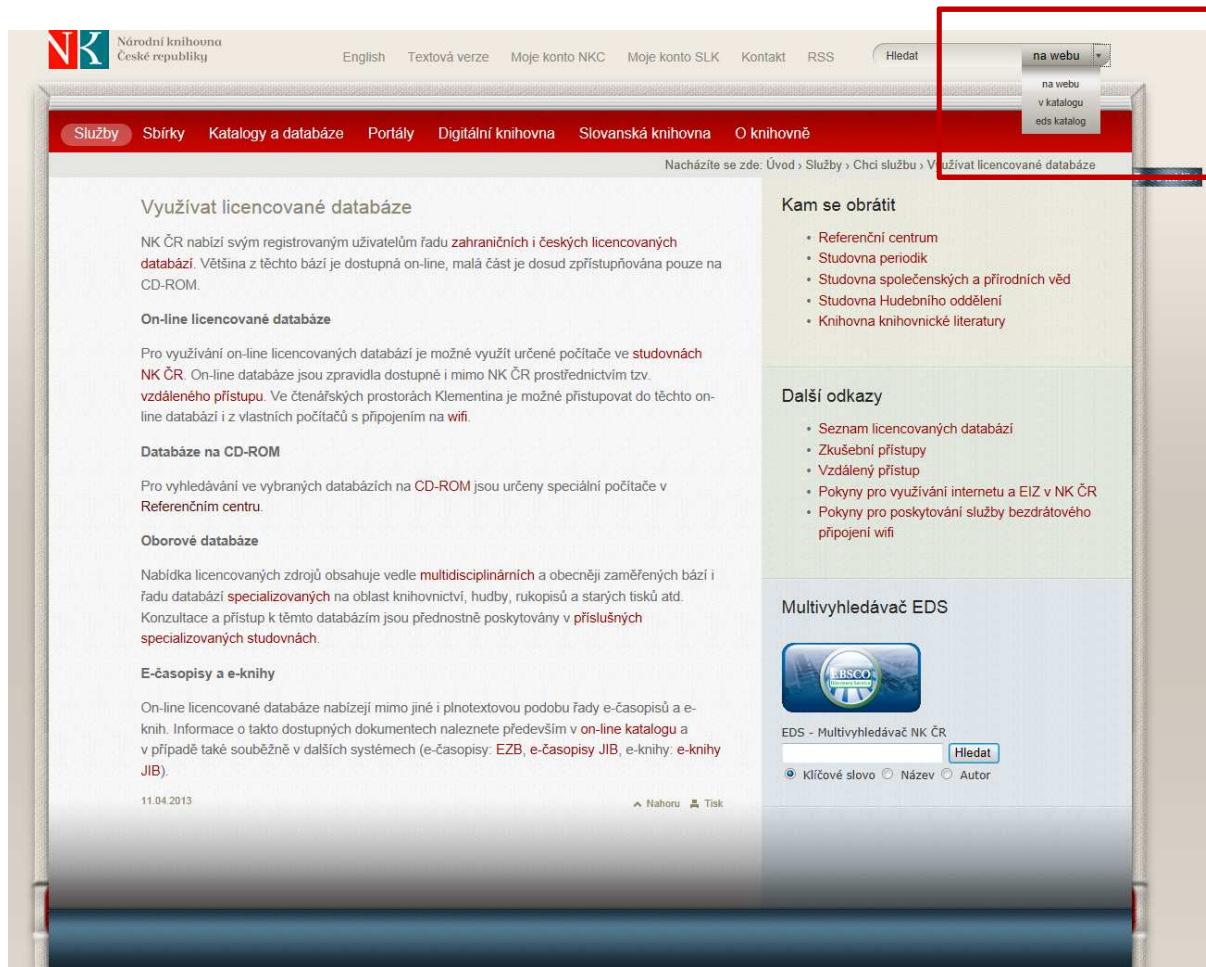
Pro větší pohodlí uživatelů je možné vytvořit i tzv. search boxy, které mohou být vnořené do běžných stránek knihovny. Uživatel tak zadá dotaz velmi pohodlně na stránce knihovny a v dalším kroku je již přenesen do prostředí EDS. Ve spolupráci s EBSCOhost jsou pro stránky NK ČR vytvořeny search boxy, které respektují design webu NK ČR, respektive rozložení informací na stránce.

The image shows a screenshot of the National Library of the Czech Republic (NK ČR) website. The page is titled "Využívat licencované databáze" (Using licensed databases). The main content area contains several sections: "On-line licencované databáze", "Databáze na CD-ROM", "Oborové databáze", and "E-časopisy a e-knihy". On the right side, there are sections for "Kam se obrátit" (Where to turn) and "Další odkazy" (Further links). A search box for EDS is highlighted with a red box, containing the text "Multivyhledávač EDS" and a search button labeled "Hledat". The search box also includes radio buttons for "Klíčové slovo", "Název", and "Autor".

Obr. 4. search-box pro vyhledávání v EDS umístěný na stránkách NK ČR

Využití zmíněných search-boxů je ideální pro propagaci a upoutání pozornosti uživatelů. V současné době jsou všechny search-boxy umístěné na webových stránkách NK ČR vytvořeny v režimu Guest/Host, tj. pro vyhledávání je může využít skutečně kdokoliv.

EDS je integrováno i ve vyhledávacím okně v záhlaví každé stránky webu NK ČR. Uživatel si zde pomocí rolovacího menu může vybrat mezi hledáním na webu NK ČR, v katalogu a v EDS (režim Guest/Host).



Obr. 5. integrace EDS do vyhledávacího okna na webu NK ČR

### Další možnosti nastavení

EDS nabízí svým zákazníkům řadu možností jak uzpůsobit prostředí systému jejich požadavkům. K dispozici je API rozhraní. Některé z účastnických knihoven využily možnosti a grafickou podobu EDS podřídily designu užívanému na svých stránkách. Kromě úprav designu lze také pracovat s nabídkou tzv. widgetů. Z té lze vybrat například aktuální zpravodajství agentury AP, vztahující se k zadanému dotazu, obrazové materiály atd. Je možné vytvořit i vlastní widget. V případě přístupu NK ČR je zde umístěno logo Souborného katalogu NK ČR. V plánu jsou i další úpravy, které by mohly být realizovány v blízké budoucnosti.

V dohledné době bude v přístupu NK ČR do EDS aktivován také nástroj pro správu časopisů A-to-Z, který nabídne uživatelům přehlednější informace o periodikách v elektronické i tištěné podobě dostupných nejen v NK ČR, ale prostřednictvím SKC také v dalších českých knihovnách.

Dalším plánované obohacení nabídky EDS pro uživatele NK ČR se týká meziknihovní výpůjční služby. Její integrace může mít různou podobu v závislosti na požadavcích a možnostech dané knihovny. V případě NK ČR přichází do úvahy především varianta linkování na informace o tomto typu služby u dokumentů, které neobsahují plný text či lokační údaje NK ČR.

### Statistiky

Využívání EDS je možné sledovat podle různých hledisek. Generování statistik umožňuje administrativní modul.

**Interface Usage Report** Site: NARODNI KNIHOVNA CR Detail Level: Interface Period: January 2012 - December 2012

InterfaceID/Version	Year	Month	Sessions	Searches	Requests					
					PDF Full Text	HTML Full Text	Image/Video	Abstract	Smart Link To	Custom Link
eds / live	2012	December	397	78924	238	29	0	407	14	75
eds / live	2012	November	574	116951	473	75	5	679	14	117
eds / live	2012	October	680	112269	296	105	0	695	13	224
eds / live	2012	September	594	95231	331	50	0	619	4	132
eds / live	2012	August	638	116028	859	48	2	466	0	134
eds / live	2012	July	397	48243	347	56	1	430	7	84
eds / live	2012	June	325	53937	339	53	0	367	9	74
eds / live	2012	May	515	52216	288	97	0	489	15	97
eds / live	2012	February	86	13856	31	26	0	112	2	28
eds / live	2012	January	64	5326	17	6	0	50	0	27
<b>Grand Total</b>			4270	692981	3219	545	8	4314	78	992

*Obr. 6. statistika přístupů do EDS v NK ČR za rok 2012*

V roce 2012 bylo zadáno téměř 693.000 dotazů. EDS v NK ČR si od svého uvedení do ostrého provozu v květnu 2012 udržuje stabilně dobrou úroveň.

**Interface Usage Report** Site: NARODNI KNIHOVNA CR Detail Level: Interface Period: January 2012 - December 2012

InterfaceID/Version	Year	Month	Sessions	Searches	Requests					
					PDF Full Text	HTML Full Text	Image/Video	Abstract	Smart Link To	Custom Link
ehost / live	2012	December	131	2770	251	52	0	246	13	69
ehost / live	2012	November	197	2960	98	15	0	271	3	56
ehost / live	2012	October	198	3975	147	88	0	838	30	33
ehost / live	2012	September	111	2013	111	25	0	150	10	8
ehost / live	2012	August	151	3193	94	47	0	501	6	20
ehost / live	2012	July	157	3226	99	54	0	133	3	31
ehost / live	2012	June	146	4251	130	51	0	348	10	9
ehost / live	2012	May	297	4626	275	77	0	457	3	36
ehost / live	2012	April	775	11044	914	174	0	1063	38	106
ehost / live	2012	March	865	12065	1196	675	0	1802	30	133
ehost / live	2012	February	658	7974	487	141	0	709	18	90
ehost / live	2012	January	541	8999	469	164	0	662	33	27
<b>Grand Total</b>			4227	67096	4271	1563	0	7180	197	618

*Obr. 7 statistika přístupů do databází v rozhraní EBSCOhost za rok 2012*

Zajímavé je porovnání s využíváním databází EBSCO v rozhraní EBSCOhost. Od května 2012, kdy bylo EDS uvedeno do provozu, zde došlo ke snížení počtu přístupů a dotazů. Lze z toho tedy usuzovat, že poměrně značná část uživatelů přešla k novému vyhledávacímu rozhraní.