

# **CORE**

## **C**onnecting **RE**positories

The Open University, Jisc  
<https://core.ac.uk/>

## Obsah

CORE.....	1
COnnecting REpositories.....	1
Současná mise CORE.....	3
Principy fungování CORE služby.....	3
1. Agregace dat.....	3
1a. Agregace dat: získání metadat.....	4
1b. Agregace dat: harmonizace a validace metadat.....	4
1c. Agregace dat: získání plných textů.....	4
2. Předzpracování dat.....	4
3. Obohacování (meta)dat.....	4
4. Deduplikace a indexování.....	5
5. Zpřístupnění dat.....	5
5a. Zpřístupnění dat: transaction information access.....	5
5b. Zpřístupnění dat: analytical access.....	6
5c. Zpřístupnění dat: raw data access.....	6
Aktuální stav.....	7
Zhodnocení projektu.....	7
Seznam literatury a zdrojů.....	8
Metadata v Dublin Core.....	10

Systém CORE, COnnecting REpositories, je největší světová kolekce a zároveň vyhledávač plně přístupných (open access, OA) vědeckých textů z mnoha akademických oblastí. CORE sdružuje výlučně recenzované akademické články, nerecenzované předtisky a odkazy na takové texty do jiných repozitářů. Vyjma kolekce a vyhledávání CORE rovněž poskytuje technologické zázemí a užitečné služby nad daty pro výzkumníky, běžné uživatele i knihovnické systémy a green OA repozitáře.

Tento projekt vznikl roku 2010 na The Open University ve Spojeném království a původně si kladl za cíl vytvořit kolekci metadat v Linked Data formátu, jenž by popisoval sémantické vztahy mezi články v open access repozitářích napříč Spojeným královstvím (1). Současně měl vyvinout Webové služby a doplňky pro repozitáře, které by usnadnily vyhledávání těchto zdrojů. Malá anglická kolekce se však časem vyvinula ve světově používaný systém a dnes běží jako služba stále zdokonalovaná a vyvíjená Media Knowledge institutem na The Open University a organizací Jisc.

## Současná mise CORE

Stálý rozvoj projektu podněcuje dosud nenaplněná vize, že vědomosti by měly být vytvářeny a používány pro dobro veřejnosti, a proto by každý měl mít právo se k nim snadno a bezplatně dostat (2). Nelze než souhlasit s názory (3), dle nichž společnost s volným přístupem k vědě může být společnost vzdělanější a vyspělejší. Uvažme navíc, že veřejnost pomáhá financovat vědu prostřednictvím daní. Otevřenému přístupu bohužel stojí v cestě finanční zájmy autorů a distributorů článků.

CORE už od svých počátků ctí principy Budapest Open Access Initiative (BOAI). Proto z repozitářů po celém světě agreguje jen ty vědecké publikace, které jsou veřejně přístupné ve smyslu definice ‚open access‘ z deklarace BOAI z roku 2001:

*By 'open access' to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited. (3)*

CORE dává tuto agregaci plně k dispozici a poskytuje k tomu nejen nezbytnou technickou infrastrukturu, ale i další užitečnou funkcionalitu. To vše dělá bezplatně, aby širší otevřený přístup k vědě.

Nejsou to však jen ušlechtilé vize a možnost bezplatně stahovat akademické články, co svým uživatelům CORE nabízí. Nad veřejně přístupnými daty lze snáze zdokonalovat a aplikovat moderní algoritmy, které jsou typicky založené na metodách strojového učení a vyžadují tedy určité množství dat pro natrénování. Díky CORE a podobným iniciativám máme dokonalejší algoritmy pro vyhledávání, ověřování faktů a detekci dezinformací, analýzu výzkumných trendů, detekci plagiátů, analýzu vlivu výzkumů, kontrolu kvality publikací aj. (4). Rovněž, pokud uživatelům internetu umožníme nakládat s něčím zcela libovolně, má to typicky mnohem větší šanci dosáhnout širšího publika a být dostupné z více zdrojů pro případ, že by jeden přestal být k dispozici.

## Principy fungování CORE služby

### 1. Agregace dat

K tomu, aby (ta kolekce/služba) CORE umožnila výše zmíněné, potřebuje nejprve získat tzv. raw data. Raw data jsou libovolná data o publikacích jako plné texty článků, metadata či abstrakty, která CORE získala přímo z repozitářů nebo odvodila během procesu získávání (neboli agregace) a kromě začlenění do kolekce je nijak dále neměnila (5, s. 1).

### **1a. Agregace dat: získání metadat**

Jednotné formáty metadat propojují rozmanité druhy repozitářů a typy textů, tudíž extrakce metadat představuje základní krok agregace. CORE k extrakci používá zejména OAI-PMH (6), ale i jiné protokoly. Také podporuje dávkové importy raw dat přímo ze souborových systémů.

OAI-PMH (Open Access Initiative Protocol for Metadata Harvesting) je protokol speciálně navržený pro zajištění interoperability repozitářů během sklizně (harvesting) metadat (7). Mezi jeho přednosti můžeme zařadit aplikační nezávislost a možnost snadného zjištění změn v obsahu repozitářů, které jej využívají. OAI-PMH tak činí proces synchronizace mezi CORE kolekcí a repozitáři jednodušší. Stačí vyslat na daný repozitář OAI-PMH požadavek. Úspěšné požadavky vrátí XML dokument s informací o dostupných článcích (6).

Pro správu samotných repozitářů a vyhledávání dalších zdrojů článků CORE využívá globální adresář akademických open access repozitářů OpenDOAR a jeho API (6).

### **1b. Agregace dat: harmonizace a validace metadat**

Aby byla (meta)data pocházející z různých zdrojů snadno strojově zpracovatelná, je potřeba je zkombinovat do jednoduchého společného formátu. To je úkolem harmonizace. V současné době CORE podporuje dva formáty metadat: jeden definovaný protokolem OAI-PMH se slovníkem Dublin Core a druhý definovaný dle instrukcí aplikačního profilu RIOXX (8).

Během převodu do jednotného formátu zároveň probíhá kontrola (validace) metadat. Pokud kontrola narazí na nějaké nesrovnalosti (například záznam neodpovídá známému standardu), poskytuje CORE repozitářům informace o tomto problému prostřednictvím služby CORE Repository Dashboard, o které bude řeč později.

### **1c. Agregace dat: získání plných textů**

Pokud se mezi metadaty nachází také odkaz na plný text dokumentu, použije se k jeho stažení. CORE získává text samostatně za pomoci crawlerů. Crawlery nazýváme programy, které pomocí http[s] protokolu následují odkazy a za pochodu z Webových stránek extrahují požadované informace a další odkazy, které mohou později následovat. Následování odkazů se využívá z toho důvodu, že metadata mohou obsahovat odkazy na různé cíle, a ne vždy přímo na soubor s plným textem. CORE crawlery, stejně jako ty od Google, sledují instrukce uvedené v ‚robots.txt‘ souboru daného webového serveru (9). Tento specifikuje, který obsah repozitáře [ne]smí být procházen a případně indexován určitou automatizovanou entitou, stanovuje jednoduchá indexovací pravidla apod. Díky němu mají servery možnost regulovat chování agregátorů a alespoň vyvinout snahu chránit práva a zájmy autorů zveřejněných dat.

Je-li výsledný soubor získán, provede se extrakce prostého textu a jeho jednotlivých sekcí. V současné době CORE umí získat prostý text pouze z PDF souborů, které ovšem pokrývají většinu online dostupných akademických publikací.

## **2. Předzpracování dat**

Pro zpracování textového obsahu položek metadat i plných textů CORE používá tzv. content-based analýzu. Tento proces začíná tím, že data jsou podrobena tokenizaci, různému filtrování, stemmingu (odsekávání předpon a přípon) a indexaci pro snadnější vyhledávání obsahu na základě klíčových slov (6).

## **3. Obohacování (meta)dat**

Někdy nebývají získaná data úplná. Často z nich ale CORE může vyvodit další informace, jež mohou zkvalitnit výsledky vyhledávacích i jiných algoritmů. Proto takto předzpracovaná data CORE dále obohacuje pomocí metod strojového učení a zpracování přirozeného jazyka. Z prostého textu lze například algoritmicky odvodit kategorii obsahu, jazyk textu, abstrakt nebo sémantickou podobnost s jiným textem (6). Články

bývají zakončeny soupisem citací. Tyto lze také extrahovat a s využitím služby CrossRef nalézt citované objekty. Následováním URL, DOI či ORCID identifikátorů a za pomoci fulltextového vyhledávání nad vlastní databází je CORE schopna sloučit informace z více různých zdrojů a doplnit chybějící údaje (10).

Aplikace metod strojového učení a zpracování přirozeného jazyka nad tak velkým datasetem, jaký má CORE k dispozici, není možná bez využití sofistikovaných heuristik a nejmodernějších nástrojů. Z hlediska obohacování dat je pro CORE klíčové partnerství s firmou Microsoft (11), díky němuž CORE může propojovat své agregované články s dokumenty v Microsoft Academic Graph. MAG je entitní graf, který obsahuje šest typů entit: obor studia, autor, instituce, článek, místo konání a událost. Nové entity a propojení mezi nimi vznikají automaticky zpracováním dat z mnoha různých zdrojů. Mezi hlavní zdroje se řadí kanály vydavatelů článků, výsledky těžby dat z webového indexu nebo třeba logové záznamy vyhledávače Bing, které obsahují například informace o každém kliknutí na nějaký odkaz. Automatizované algoritmy data předzpracovávají, filtrují a odvozují chybějící údaje. K odvozování se využívají analýza kliků z Bingu, již existující propojení v grafu, extrakce z klíčových slov a podobně (12).

Výsledkem propojení vztahů z MAG s CORE datasetem je největší světový graf akademických článků, který obsahuje jak bibliografické informace, tak plné texty článků. Mapování z CORE datasetu do MAG bylo taktéž letos poprvé zveřejněno ke stažení na stránkách CORE.

#### **4. Deduplikace a indexování**

Na závěr jsou po fázi obohacování dat odstraněny duplicity (deduplikace) a záznamy jsou indexovány pro potřeby fulltextového prohledávání (13). Takto zušlechťená data tvoří obsah volně přístupných datasetů.

#### **5. Zpřístupnění dat**

CORE umožňuje uživatelům přistupovat k datům třemi způsoby, jež nazývá raw data access, transaction information access a analytical access (5).

##### **5a. Zpřístupnění dat: transaction information access**

Základním způsobem využití CORE služeb, typickým například pro výzkumníka nebo studenta, je tzv. transaction information access. Při něm uživatel může vyhledávat v článcích a stahovat je. K tomu CORE přímo na svých webových stránkách poskytuje vyhledávač CORE Search. Vyhledávač CORE je dostupný i v mobilní verzi pro iOS, ale ta není tolik využívána a již ani udržovaná.

Nyní také CORE nabízí doplněk CORE Discovery, a to ve dvou verzích. Zaprvé je CORE Discovery dostupný jako prohlížečový plugin pro Google Chrome, Operu nebo Mozillu. Funguje tak, že na stránkách vydavatelů s placeným přístupem k článkům zobrazuje i odkazy na volně přístupné kopie, pokud takové existují. K tomu využívá dat z MAG, CrossRef, Unpaywall a CORE datasetu (14). Zadruhé jej mohou využívat samotné repozitáře pro usnadnění vyhledávání volně přístupných kopií.

Co však činí vyhledávání prostřednictvím CORE skutečně jedinečným, je její generátor doporučení na další zajímavé články. Knoth a spol. (15, s. 1-2) tvrdí, že integrací nástrojů generujících taková doporučení do repozitářů můžeme podstatně zvýšit dostupnost open access zdrojů a zlepšit uživatelský komfort při hledání informací. Proto CORE vyvinula plugin CORE Recommender, který po instalaci do repozitáře, vydavatelského systému nebo webového rozhraní generuje doporučení na relevantní články podobné tomu, který je v daném systému aktuálně zobrazen.

Primárně se CORE Recommender od jiných generátorů liší tím, že generuje doporučení pouze na open access články. Navíc nezatežuje svými výpočetními nároky systémy, do kterých je integrován, ale posílá systému CORE informace o navštívené položce a CORE mu odpovídá seznamem doporučených článků. Ukázal se rovněž jako velice úspěšný. Některé instituce hlásají, že díky jeho implementování zaznamenaly i více než 50% prodloužení času, který uživatelé stráví procházením článků (16).

Unikátní je i samotná implementace jeho algoritmu, protože narozdíl od spousty soudobých generátorů doporučení CORE Recommender těží i ze samotných textů článků. Následující popis principů jeho fungování vychází z (15, s. 2-5).

Algoritmus CORE Recommenderu je založený na metodě content-based filtering, tedy na porovnávání podobnosti článků podle atributů. Na vstup dostává referenční dokument, pro nějž chceme nalézt další doporučené dokumenty. Z dostupných informací si vyhledá nebo vypočítá hodnoty zmíněných atributů, kterými jsou například jazyk textu, počet citací, počet stažení, počet uživatelů, kteří mají dokument mezi svými oblíbenými položkami aj. Poté si reprezentuje referenční dokument jako množinu vektorů, jejichž souřadnice jsou určeny těmito hodnotami. Podobně jsou reprezentovány dokumenty, se kterými bude algoritmus porovnávat.

Hodnoty souřadnic mají různou váhu, tedy různé atributy jsou různě důležité. Například rok vydání je upraven klesající funkcí tak, aby se mezi doporučeními ve výsledku objevily spíše novější články. Dvojice článků, které jsou podle MAG často citovány spolu, by si měly být ve vektorovém prostoru blíže. Podobně pro další souřadnice. Následně algoritmus pomocí kosinové podobnosti vyhledá nejbližší vektory a vygeneruje požadovaný počet záznamů. Ze seznamu jsou ještě odfiltrovány články, které nejsou plně přístupné, chybí jim základní metadata, jsou duplicitami jiných doporučení apod.

Takto vytvářená doporučení jsou nepersonalizovaná. Uživatelé mohou jenom podávat negativní feedback na špatná doporučení, a tak pomoci zlepšovat parametry popsaného algoritmu. Zvažme, zdali je vůbec možné si představit objektivní a užitečný personalizovaný generátor doporučení na vědecké články.

## 5b. Zpřístupnění dat: analytical access

Sponzoři, vládní a jiné organizace mohou chtít přistupovat ke statistickým informacím (6) nad daty a repozitáři. Těmito mohou být statistiky obsahu (počty článků v jednotlivých kolekcích, ...), statistiky z metadat, trendy z různých oblastí výzkumu, počty citací napříč repozitáři a další (5, s. 2). CORE tyto statistické údaje bezplatně vytváří a poskytuje.

Pro analýzu statistik také existuje balíček „rcoreoa“ poskytující rozhraní mezi R a CORE API.

## 5c. Zpřístupnění dat: raw data access

CORE je výjimečná v tom, že zveřejňuje i nashromážděná raw data. Jednou ročně na své stránky přidává aktuální obsah datasetu ve verzi s plnými texty nebo bez. Vývojáři, akademičtí pracovníci, digitální knihovny a další subjekty (5) jej mohou využívat pro nekomerční účely a dále zpracovávat.

Pro snadné vyhledávání a získávání dat z repozitářů a metadat článků má CORE k dispozici své API. Pro využití této služby je však nutné se bezplatně zaregistrovat. CORE též nabízí ke stažení službu FastSync umožňující inkrementální aktualizace dat z CORE systému (10). To znamená, že uživatelský server může udržovat svoji kopii dat aktuální například s nově přidanými publikacemi. K tomuto účelu FastSync využívá ResourceSync protokol, který je zaměřený na synchronizaci v reálném čase. ResourceSync specifikuje, jak vytvářet seznamy zdrojů, které mají být synchronizovány, indexy nad těmito seznamy, dokumenty popisující změny, patche, balíčky, notifikace, zrcadla, jak synchronizovat metadata apod. (17).

Jakkoli jsou vybudovaná infrastruktura a objem CORE datasetu hodny obdivu, ne vždy si autoři či vlastníci dat přejí být jejich součástí. CORE ovšem klade důraz i na souhlas zdroje s volným přístupem k dílu. Již během samotné agregace se snaží respektovat přání poskytovatelů vyjádřených v „robots.txt“ souboru, ale jediným souborem nelze zamezit šíření dat internetem, i kdyby jej všichni roboti respektovali. CORE situaci řeší tak, že umožňuje poskytovatelům repozitářů zasahovat do procesu agregace nástrojem CORE Repository Dashboard. Jedná se o on-line rozhraní udělující majitelům repozitářů přístup k informacím o tom, který obsah a kdy byl z repozitáře sklizen, jaké chyby se při agregaci vyskytly atp. Kromě toho se z něj majitelé repozitářů mohou dozvědět různé statistiky o repozitáři, třeba jeho srovnání s jinými z dané oblasti. Hlavně

ale mají příležitost odstranit nebo přidat plné texty článků z/do CORE datasetu, upravit metadata, znovu spustit proces sklizení nad upravenými (meta)daty, a další (18). Touto cestou má poskytovatel možnost kontrolovat a zasahovat do procesu harvestingu nad jeho repozitáři a spolupracovat s CORE na šíření open access vize.

## **Aktuální stav**

CORE je první kolekce svého druhu, která byla od svého vzniku zaměřena na open access vědecké články. Doposud zdarma zpřístupnila necelých 25 milionů plných textů, z nichž přes 14 milionů je uloženo přímo na jejích serverech. Celkově však umožňuje vyhledávat ve více než 135 milionech článků. Globální dosah tohoto projektu podtrhuje, že CORE agreguje data od poskytovatelů ze 144 zemí, kterými jsou zejména různé open access repozitáře a vydavatelé periodik. V roce 2018 se dokonce stala největším agregátorem veřejně přístupných článků (19).

Její služby využívá přes 25 milionů uživatelů z celého světa (20) a v současné době se řadí mezi top 5000 webových stránek z hlediska tzv. user engagement – kombinace návštěvnosti dané domény a počtů shlédnutí jejich jednotlivých stránek za poslední tři měsíce. Dle Alexy (21) byla k 23. 11. 2019, 11:13 na 4894. pozici.

CORE vzájemně propojila své služby s mnoha dalšími nástroji a organizacemi. Zmiňme zejména Tournitin (pro usnadnění hledání plagiátů a tím pádem zajištění ochrany veřejně přístupného výzkumu před zneužitím), ProQuest (pro rozšíření možností vyhledávání v knihovnických systémech), OpenMinTeD, Microsoft Academic, PubMedCentral, ... (9) (22) (23)

A na závěr v říjnu letošního roku získal tým CORE ocenění za výjimečný přínos jejich výzkumu pro společnost a prosperitu na The Open University Research Excellence Awards 2019 (24).

## **Zhodnocení projektu**

Tento projekt hodnotím velmi příznivě. Jako student jsem ráda, že mohu skrze OA zdroje anebo předplacené zdroje Masarykovy univerzity stahovat a číst odborné práce a nedokážu si představit studovat bez těchto služeb. Sdílím vize CORE a BOAI, že open access urychlí pokrok digitálních technologií, usnadní vědeckou spolupráci a zvýší vzdělanost populace. Ocenění Research Excellence Awards svědčí o tom, že CORE na tomto poli značně přispívá k uskutečnění těchto ideálů všemi inovacemi, které přinesla, a infrastrukturou, kterou pomohla vybudovat.

Sama přes CORE Search občas vyhledávám a musím podotknout, že CORE Recommender skutečně funguje ve většině případech tak, jak byl navržen. Doporučuje obvykle související články novějšího data a přístup k plnému textu článku je mnohem snadnější než například přes discovery.muni.cz – a to nejen vzhledem k samotné dostupnosti, ale i uživatelské přívětivosti.

Stále sice přes CORE nenaleznu většinu textů, které hledám, není to ovšem její chyba. Dle loňských dat bylo pouze 28 % odborných článků veřejně přístupných a z těchto pouhých 17 % pocházelo z veřejných repozitářů (25). Jejich rostoucí podíl a stále se zvyšující uživatelská základna služeb jako CORE ale jasně ukazují, jakým směrem se společnost ubírá. Demonstrují, že CORE úspěšně plní svoji misi a že tento systém může fungovat.

## Seznam literatury a zdrojů

1. KNOTH, Petr. Project Plan. In: *blog.core.ac.uk* [online]. 24. 3. 2011 [cit. 22. 11. 2019]. Dostupné z: <https://blog.core.ac.uk/2011/03/24/project-plan/>
2. About CORE, 2019. *CORE* [online]. Milton Keynes (Spojené království): The Open University. [Cit. 22. 11. 2019]. Dostupné z: <https://core.ac.uk/about/#mission>
3. BUDAPEST OPEN ACCESS INITIATIVE. Read the Budapest Open Access Initiative. In: *www.budapestopenaccessinitiative.org* [online]. 14. 2. 2002 [cit. 23. 11. 2019]. Dostupné z: <https://www.budapestopenaccessinitiative.org/read>
4. What others say about us, 2019. *CORE* [online]. Milton Keynes (Spojené království): The Open University. [Cit. 22. 11. 2019]. Dostupné z: <https://core.ac.uk/about/endorsements/>
5. KNOTH, Petr a ZDRAHAL, Zdenek. CORE: aggregation use cases for open access. In: *2nd International Workshop on Mining Scientific Publications (WOSP 2013)* [online]. Indianapolis: 26. 7. 2013 [cit. 22. 11. 2019]. Dostupné z: <http://oro.open.ac.uk/37823/>
6. KNOTH, Petr a ZDRAHAL, Zdenek. CORE: Three Access Levels to Underpin Open Access. *D-Lib Magazine* [online]. November/December 2012, 18 [cit. 22. 11. 2019]. ISSN 1082-9873. DOI: 10.1045/november2012-knoth
7. LAGOZE, Carl a VAN DE SOMPEL, Herbert a NELSON, Michael a WARNER, Simeon. The Open Archives Initiative Protocol for Metadata Harvesting. *Open Archives Initiative* [online]. OAI. Poslední změna 8. 1. 2015. [Cit. 22. 11. 2019]. Dostupné z: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
8. [DRAFT] CORE support for UK HEIs in the REF 2021 Open Access Audit, 2019. *CORE* [online]. Milton Keynes (Spojené království): The Open University. [Cit. 24. 11. 2019]. Dostupné z: <https://core.ac.uk/ref-audit/>
9. STEPHENS, Owen. Best of both worlds. In: *blog.core.ac.uk* [online]. 18. 4. 2012 [cit. 23. 11. 2019]. Dostupné z: <https://blog.core.ac.uk/2012/04/18/best-of-both-worlds/>
10. KNOTH, Petr. Better together: building services for public good on top of content from the global network of open repositories [elektronická prezentace]. In: *OAI 11 – The CERN-UNIGE Workshop on Innovations in Scholarly Communication*. Ženeva (Švýcarsko), 21. 6. 2019. [Cit. 24. 11. 2019]. Dostupné z: <https://www.slideshare.net/petrknoth/better-together-building-services-for-public-good-on-top-of-content-from-the-global-network-of-open-repositories>
11. KNOTH, Petr. Data discovery opportunities: Lessons learned from CORE [elektronická prezentace]. In: *Open Science Fair 2019*. Porto (Portugalsko), 16. 9. 2019. [Cit. 24. 11. 2019]. Dostupné z: <https://www.slideshare.net/petrknoth/data-discovery-opportunities-lessons-learned-from-core>
12. SINHA, Arnab a SHEN, Zhihong a SONG, Yang a MA, Hao a EIDE, Darrin a WANG, Kuansan. An Overview of Microsoft Academic Service (MAS) and Applications. In: *WWW 2015 Conference Proceedings and Companion* [online]. Florencie (Itálie): WWW 2015 Companion, May 2015, 243-246 [cit. 23. 11. 2019]. ISBN 978-1-4503-3473-0. DOI: 10.1145/2740908.2742839
13. PONTIKA, Nancy a KNOTH, Petr a CANCELLIERI, Matteo a PEARCE, Samuel. Developing Infrastructure to Support Closer Collaboration of Aggregators with Open Repositories. *LIBER Quarterly* [online]. 2016, 25(4), 172–188. E-ISSN 2213-056X. DOI: 10.18352/lq.10138
14. KNOTH, Petr a CANCELLIERI, Matteo. Analysing the performance of open access papers discovery tools [elektronická prezentace]. In: *Proceedings of Open Repositories 2019*. Hamburk (Německo), 13. 6. 2019. [Cit. 23. 11. 2019]. Dostupné z:



- <https://www.slideshare.net/petrknoth/analysing-the-performance-of-open-access-papers-discovery-tools>
15. KNOTH, Petr a LUCAS, Anastasiou a CHARALAMPOUS, Aristotelis a CANCELLIERI, Matteo a PEARCE, Samuel a PONTIKA, Nancy a BAYER, Václav. Towards effective research recommender systems for repositories. In: *Proceedings of Open Repositories 2017*. Brisbane (Austrálie): Open Repositories, Jun 2017, 26-30 [cit. 22. 11. 2019]. Dostupné z: <http://oro.open.ac.uk/49366/>
  16. PONTIKA, Nancy a MACGREGOR, George. Implementing the CORE Recommender in Strathprints: a “whitehat” improvement to promote user interaction. In: *blog.core.ac.uk* [online]. 31. 10. 2017 [cit. 23. 11. 2019]. Dostupné z: <https://blog.core.ac.uk/2017/10/31/implementing-the-core-recommender-in-strathprints-a-whitehat-improvement-to-promote-user-interaction/>
  17. VAN DE SOMPEL, Herbert a LAGOZE, Carl a WARNER, Simeon a NELSON, Michael a HASLHOFER, Bernhard a SANDERSON, Robert a KLEIN, Martin a KLYNE, Graham. ResourceSync Framework Specification (ANSI/NISO Z39.99-2017). *Open Archives Initiative* [online]. OAI. Poslední změna 2. 2. 2017. [Cit. 23. 11. 2019]. Dostupné z: <http://www.openarchives.org/rs/1.1/resourcesync>
  18. PONTIKA, Nancy. CORE Repositories Dashboard: An infrastructure to increase collaboration of Aggregators with Open Repositories. In: *blog.core.ac.uk* [online]. 8. 9. 2015 [cit. 23. 11. 2019]. Dostupné z: <https://blog.core.ac.uk/2015/09/08/core-repositories-dashboard/>
  19. CANCELLIERI, Matteo. CORE becomes the world’s largest open access aggregator (or how about them stats 2018 edition). In: *blog.core.ac.uk* [online]. 18. 12. 2018 [cit. 23. 11. 2019]. Dostupné z: <https://blog.core.ac.uk/2018/12/18/core-becomes-the-worlds-largest-open-access-aggregator-or-how-about-them-stats-2018-edition/>
  20. Data, data, data, 2019. *CORE* [online]. Milton Keynes (Spojené království): The Open University. [Cit. 22. 11. 2019]. Dostupné z: <https://core.ac.uk/data/>
  21. *Alexa* [online]. Alexa Internet, Inc.: © 1996-2019 [cit. 23. 11. 2019]. Dostupné z: <https://www.alexa.com/>
  22. FAHMY, Sarah. Enhancing library discovery services with CORE content. In: *scholarlycommunications.jiscinvolve.org* [online]. Jisc scholarly communications, 14. 12. 2017 [cit. 24. 11. 2019]. Dostupné z: <https://scholarlycommunications.jiscinvolve.org/wp/2017/12/14/enhancing-library-discovery-services-with-core-content/>
  23. HARRISON, Jennifer a HAND, Amanda Z. Turnitin Partners with CORE, the World’s Leading Aggregator of Open Access Research Articles. In: *www.businesswire.com* [online]. 31. 5. 2019 [cit. 24. 11. 2019]. Dostupné z: <https://www.businesswire.com/news/home/20190531005083/en/Turnitin-Partners-CORE-World%E2%80%99s-Leading-Aggregator-Open>
  24. NOTAY, Balviar. CORE wins an Outstanding Impact of Research on Society and Prosperity Award. In: *scholarlycommunications.jiscinvolve.org* [online]. Jisc scholarly communications, 4. 11. 2019 [cit. 24. 11. 2019]. Dostupné z: <https://scholarlycommunications.jiscinvolve.org/wp/2019/11/04/core-wins-an-outstanding-impact-of-research-on-society-and-prosperity-award/>
  25. PIWOWAR, Heather a PRIEM Jason. The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ* [online]. 13. 2. 2018, 6 [cit. 24. 11. 2019]. ISSN 2167-8359. DOI: 10.7717/peerj.4375

## Metadata v Dublin Core

```
<?xml version="1.0" encoding="UTF-8"?>
<metadata xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>CORE</dc:title>
  <dc:creator>Pavlına Klimeřov</dc:creator>
  <dc:subject>Kolekce COnnecting REpositories</dc:subject>
  <dc:date>2019-12-01</dc:date>
  <dc:description>CORE je největř svetov kolekce a vyhledava OA vedeckch text poskytu
  technologick zazem a sluby nad daty. Esej popisuje misi, principy fungov a souasn stav tohoto
  projektu.</dc:description>
  <dc:language>cs</dc:language>
  <dc:type>text</dc:type>
  <dc:format>application/pdf</dc:format>
</metadata>
```