

Finanční analýza prostředků v oblasti přístupu k vědeckým informacím v období 2017-2022

Souhrnná zpráva navazující na “Zprávu o plnění Opatření 8 Akčního plánu pro implementaci Národní strategie otevřeného přístupu České republiky k vědeckým informacím na léta 2017–2020”¹. Cílem je analýza monitoringu tzv. “APC in the wild” a předplatných časopisů realizovaných mimo smlouvy Národního licenčního centra CzechELib² v období let 2019 až 2023.

Národní technická knihovna

DOI <https://doi.org/10.48813/sb5p-vj27>

Jiří Jiráť,  <https://orcid.org/0000-0002-9562-3053>

Eva Stopková,  <https://orcid.org/0000-0003-2928-6914>

Jana Bystroňová,  <https://orcid.org/0009-0007-8042-2308>

Martin Svoboda,  <https://orcid.org/0000-0002-8255-348X>

Praha, listopad 2023



Užití tohoto díla se řídí mezinárodní licencí [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/), která umožňuje dílo kopírovat a šířit pouze v nezměněné podobě a pro nekomerční účely, za podmínky řádného uvedení původních autorů a zdroje.

Obsah

1	Preambule	3
2	Úvod	3
2.1	Pilotní fáze	3
2.2	Sběry v následujících letech	4
2.3	Detailní historie sběrů dat	4
2.4	Mezinárodní kontext.....	5
3	Návratnost dotazníku a kvalita dat	6
3.1	Subjektivní hodnocení kvality získaných dodávaných dat	6
3.2	Předplatné časopisů mimo smlouvy NTK/CzechELib	7
3.3	Poplatky APC.....	7
4	Analýza získaných dat	8
4.1	Celkový objem vynaložených prostředků – vývoj v letech 2018-2022	8
4.2	Režim předplatného	9
4.3	APC	9
5	Závěr	12
6	Přílohy	13
6.1	Příloha 1: Zpracování dat.....	13
6.1.1	Unifikace a čištění primárních dat	13
6.1.2	Unifikace finančních údajů pro následné analýzy	13
6.2	Příloha 2: Seznam zkratk a termínů	14
7	Odkazy	15

1 Preambule

Nákup elektronických informačních zdrojů (EIZ) pro výzkum, vývoj a vzdělávání je v ČR dlouhodobě podporován z prostředků Státního rozpočtu na podporu VaVal. V letech 2017-2022 se podařilo distribuci podpory výrazně zefektivnit centralizací díky projektu CzechELib, který podpořil zřízení národního centra CzechELib² (jako součásti Národní technické knihovny) a ověřil možnost podpory nákupu ze zdrojů Operačního programu OP VVV. CzechELib v současnosti organizuje odhadem zhruba 85 až 90 % výdajů veřejných prostředků na nákup všech typů EIZ v celostátním měřítku. Pokud jde o elektronické časopisy, v roce 2019 zajišťoval CzechELib plných 87,5 % nákupu v celostátním měřítku, v r. 2022 to bylo již téměř 90 %. Objem výdajů komerční sféry nebyl dosud analyzován, podle kusých informací české firmy v této oblasti výrazně zaostávají za rozvinutými zeměmi.

2 Úvod

2.1 Pilotní fáze

V r. 2019, na základě Opatření 8 Akčního plánu pro implementaci Národní strategie otevřeného přístupu České republiky k vědeckým informacím na léta 2017–2020 vypracoval tým NTK/CzechELib první podrobnou specifikaci pokynů pro sběr dat potřebných pro finanční analýzu prostředků v oblasti přístupu k vědeckým informacím.

Účelem sběru dat bylo zjistit přibližné roční výdaje České republiky na zajištění přístupu výzkumné komunity k vědeckým informacím,

1. a to jak formou předplatného odborných vědeckých časopisů zprostředkovaného přes CzechELib i mimo něj,
2. tak formou poplatků APC ("APCs in the wild"), placených decentralizovaným způsobem.

Vzhledem k tomu, že paradigmatická změna modelu zajišťování přístupu k vědeckým informacím od současného předplatného na model otevřeného přístupu bude mít v nadcházejících letech finanční dopad pouze na EIZ typu publikačních výstupů, zejména článků v časopisech, byl sběr dat zaměřen právě na elektronické časopisy, nikoliv na faktografické, citační aj. databáze.

Na základě zjištění a zkušeností pak navrhnout metodiku každoročního sledování objemu výdajů na obě tyto komponenty.

Tehdejší první místopředseda Rady pro výzkum, vývoj a inovace prof. Dvořák zaslal výše uvedené pokyny e-mailem v červenci 2019 všem statutárním zástupcům výzkumných organizací (VO), které jsou členskými institucemi projektu CzechELib, výsledek sběru dat byl shrnut detailně ve zprávě.¹

2.2 Sběry v následujících letech

Po pilotní fázi k žádné zásadní systematické změně v metodice sběru zatím nedošlo, pouze k drobné úpravě, kdy všechna finanční data pro APC jsou požadována bez DPH (aby došlo k očištění od tohoto “vnějšího” vlivu a aby byla uváděna shodně jako údaje pro předplatná).

Aktuální nastavení sběru dat je následující:

Odeslání žádosti o dodání dat:	RVVI / Ministryně pro vědu, výzkum a inovace
Zpracování dat:	NTK/CzechELib
Dodavatelé dat:	členské instituce zapojené v CzechELib + některé afiliované ústavy AV ČR (zapojené do CzechELib nepřímým způsobem)
Požadovaná data:	APC předplatná časopisů mimo smlouvy NTK/CzechELib

2.3 Detailní historie sběrů dat

Sběr dat za období	Datum odeslání	Žádost posílal(a)	Pozn.
2016-2018*	2019	úřad vlády / 1. místopředseda RVVI	CzechELib požádal i o uvedení nákladů za “diamond” časopisy (nebylo součástí oficiální žádosti)
2019	2020		-
2020	2021		-
-	2022 - neodesláno	(obměna RVVI)	-
2021-2022	konec 2022, termín odevzdání jaro 2023	ministryně pro vědu, výzkum a inovace	-

* předmětem minulé zprávy¹

2.4 Mezinárodní kontext

Projekt OpenAPC

Projekt OpenAPC byl založen r. 2014 v Bielefeld University Library, s cílem monitorovat transparentním a efektivním způsobem výši APC placených u jednotlivých vydavatelů. Z původně německého projektu se rychle stal projekt mezinárodní – do datasetu, který obsahuje téměř 200 000 APC, již přispělo v období více než 400 institucí (vč. některých z ČR). Od r. 2021 je součástí OpenAIRE Nexus.³

Doporučení European University Association (EUA) z r. 2021

Dokument EUA *The new university Open Access checklist (A practical guide on implementation)* jednoznačně doporučuje systematický monitoring APC na úrovni konsorcia. (v části *BUILD CAPACITY, Libraries and consortia*, bod č. 1: *Monitor APC costs. Centralise and streamline APC reporting*).⁴

3 návratnost dotazníku a kvalita dat

V průběhu zkoumaného období došlo ke zvýšení počtu institucí, které byly v rámci dotazníkového šetření osloveny. Návratnost dotazníků s postupem let narůstala, v roce 2023 – zřejmě po jednorozhodčí přestávce – došlo k jejímu nepatrnému poklesu. Obecně ale návratnost dotazníků oslovenými institucemi v posledních třech letech překračuje 90 %.

Rok sběru	Dodali	Nedodali	Návratnost dotazníků	Počet oslovených institucí
2019	98	19	84 %	117
2020	112	9	93 %	121
2021	120	5	96 %	125
2022	-	-	-	-
2023	123	9	93 %	132

3.1 Subjektivní hodnocení kvality získaných dodávaných dat

V průběhu jednotlivých kol sběru dat bylo možné pozorovat jednoznačně tento trend:

- První kolo sběru (r. 2019) bylo ve znamení vysvětlování, co je vlastně požadováno a proč, jak směrem k pověřeným osobám institucí, tak i k jejich vedení. Na mnoha institucích byly pověřené osoby patrně ve složité situaci vůči vedení institucí (nutnost taktéž vysvětlovat, proč sběr provádět). Typické problémy a nedostatky v datech byly detailně popsány v Příloze 4 zprávy z pilotního sběru.¹
- Ovšem po finalizaci prvního sběru dat došlo na mnoha institucích k uvědomění si, o jak velkou částku se v souhrnu může jednat a u většiny institucí bylo možné sledovat změnu přístupu k následujícím sběrům dat.
- Došlo i k pozorovatelnému zlepšení kvality sběru dat (např. komunitní setkání a výměna zkušeností mezi VŠ), standardizace a zavedení různých identifikátorů v účetních systémech atp.
- Obecně lze říci, že kvalita dat dodávaných oslovenými institucemi s pokračujícími sběry narůstá. I přes tento fakt bylo následně nutné data dále rozsáhle zpracovat a jednotlivé datasey vyčistit (viz Příloha 1: Zpracování dat).

3.2 Předplatné časopisů mimo smlouvy NTK/CzechELib

Počet institucí majících jakékoli předplatné časopisů mimo smlouvy NTK/CzechELib je víceméně konstantní, změny v absolutních počtech nebo podílu institucí s předplatným jsou způsobeny spíše tím, že některé instituce neodpověděly v některých letech, a/nebo došlo také ke změnám ve složení celého konsorcia (jiný počet oslovených institucí, změněná skladba).

Rok sběru	Rok předplatného	Předplatné: ANO	Předplatné: NE	Bez odpovědi	Předplatné: ANO (podíl z oslovených)
2019	2019	77	21	19	66 %
2020	2020	80	32	9	66 %
2021	2021	81	39	5	65 %
2022	-	-	-	-	-
2023	2022*	82	41	9	55 %
	2023	79	44	9	61 %

* data údaje o předplatném na r. 2022 byla sbírána v rámci sběru v r. 2023.

Mezi instituce, které předplatné mimo smlouvy NTK/CzechELib neodebírají, patří především ústavy Akademie věd ČR, které mají svá předplatná zprostředkována Knihovnou Akademie věd ČR.

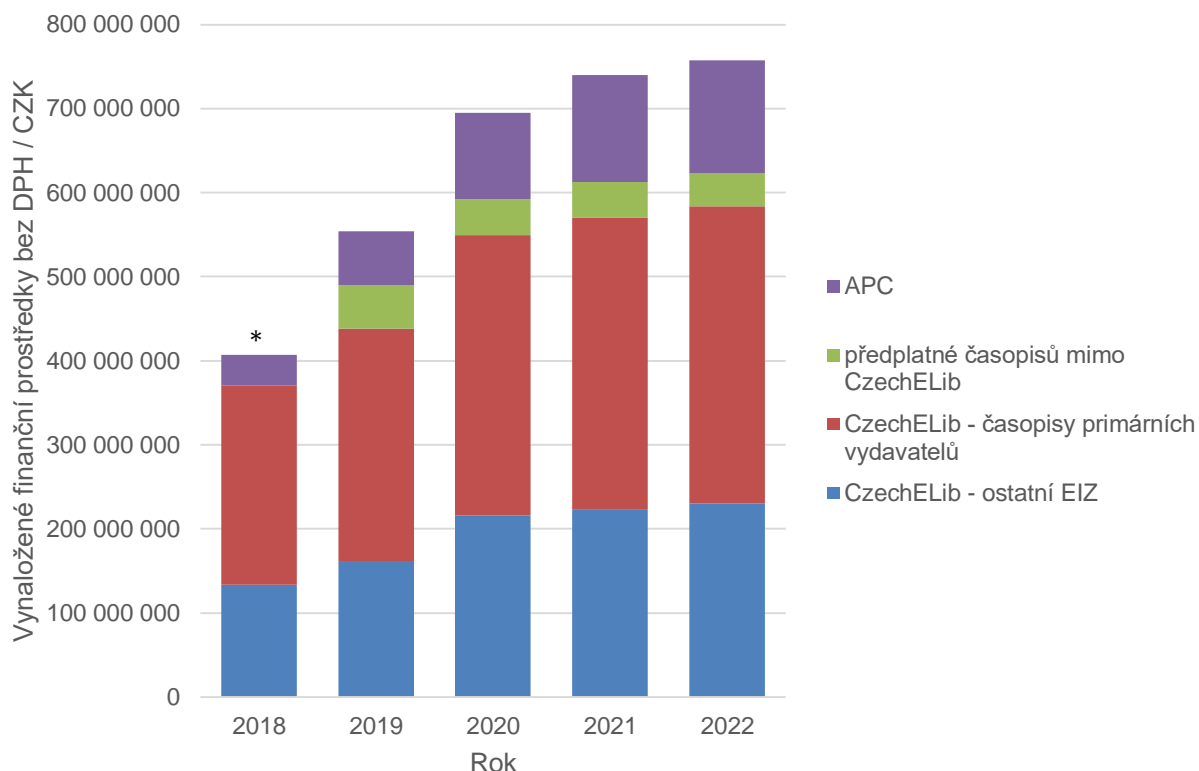
3.3 Poplatky APC

Od začátku zkoumaného období došlo k téměř 10% nárůstu počtu institucí, které poplatky za APC v dotazníkovém šetření uvedly. Mezi instituce, které uvedly nulové výdaje za poplatky APC, patří především knihovny, muzea, fakultní nemocnice, galerie a grantová agentura apod.

Rok sběru	APC: ANO	APC: NE	Bez odpovědi	Uvedli poplatky za APC (podíl z oslovených)
2019	69	29	19	59 %
2020	72	40	9	60 %
2021	81	39	5	65 %
2022	-	-	-	-
2023	89	32	9	68 %

4 Analýza získaných dat

4.1 Celkový objem vynaložených prostředků – vývoj v letech 2018-2022



* údaje o předplatném časopisů mimo CzechELib nejsou pro r. 2018 k dispozici

Obr. 1: Celkový objem finančních prostředků. Data nezahrnují předplatná EIZ jiných typů než časopisy mimo smlouvy NTK/CzechELib

Vypovídací hodnota údajů z let 2018 až 2019 je nižší:

- hodnoty za CzechELib jsou nižší, protože neobsahují zejména nečasopisecké EIZ, které v rámci udržitelnosti OP VaVpl a doběhu některých projektů LR byly placeny mimo CzechELib,
- údaje o předplatném za r. 2018 nejsou k dispozici.

Trendy:

- CzechELib – ostatní EIZ: **stabilní**, bez skokových nárůstů
- CzechELib – časopisy primárních vydavatelů: **mírný růst** (mj. přesuny z předplatného mimo CzechELib)
- předplatné časopisů mimo CzechELib: **mírný pokles**
- APC: téměř **exponenciální růst** v letech 2018-2020

4.2 Režim předplatného

Výdaje na individuální předplatná časopisů mimo CzechELib (stav r. 2022):

- více než 600 vydavatelů
- více než 2000 titulů

Celkový roční objem finančních prostředků vynakládaných na předplatné e-časopisů v ČR byl v letech 2020-2022 cca 400 mil. Kč ročně. Z celkových ročních výdajů na předplatné e-časopisů je v r. 2020 cca 87,5 % hrazeno prostřednictvím CzechELib a cca 12,5 % z individuálního předplatného; v r. 2022 jsou však podíly již cca 89 % a 11 %, **celková roční částka** (součet výdajů na předplatné časopisů v rámci CzechELib a mimo něj) je však v letech 2020-2022 prakticky **konstantní**.

To potvrzuje pozorovaný trend – po r. 2020 docházelo ke konsolidaci a přesunu zejména dražších titulů nebo kolekcí pod smlouvy CzechELib (ovlivněné zejména koncem udržitelnosti OP VaVpl, ale i jiných projektů). Je znát **mírný pokles celkové částky vydané za předplatná mimo CzechELib**, jakož i mírný setrvalý pokles v počtu vydavatelů a titulů.

4.3 APC

Přibližné počty položek za celé sledované období:

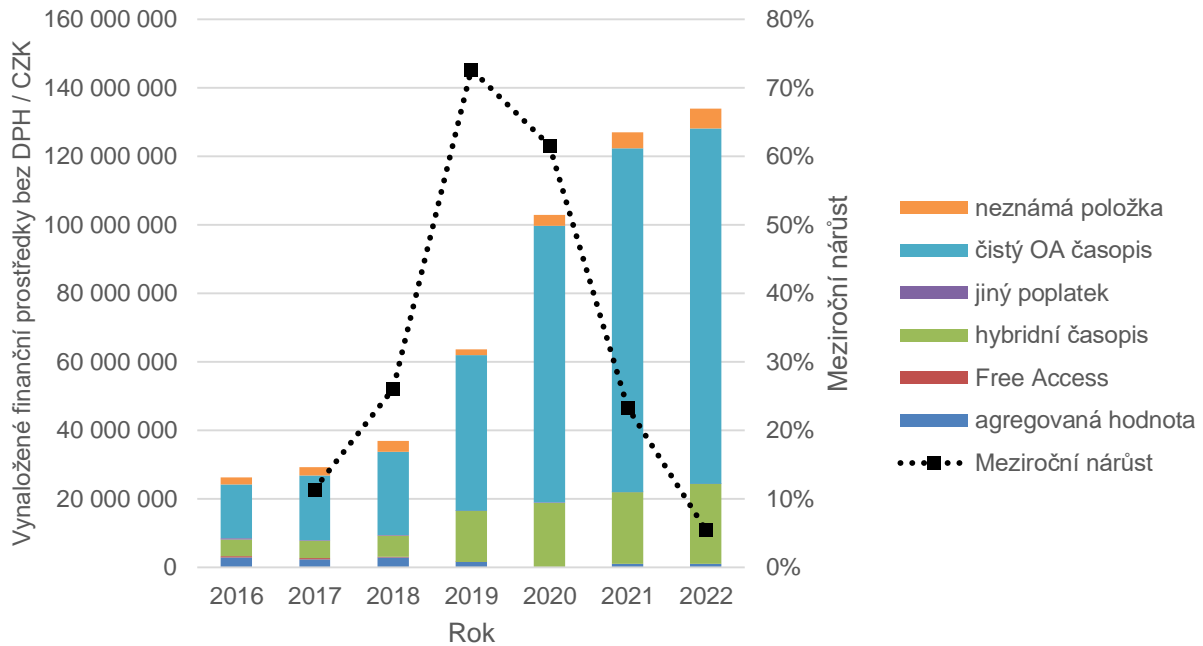
- počet článků: přes 14 300 (reálné číslo je vyšší, někde byly dodány jen agregované hodnoty),
- počet časopisů: přes 2 500,
- počet vydavatelů: přes 500.

Výdaje spojené s APC dosáhly v r. 2022 **téměř 40 %** z předplatného časopisů primárních vydavatelů v rámci CzechELib, **průměrná výše APC** bez DPH byla cca **33 000 CZK**.

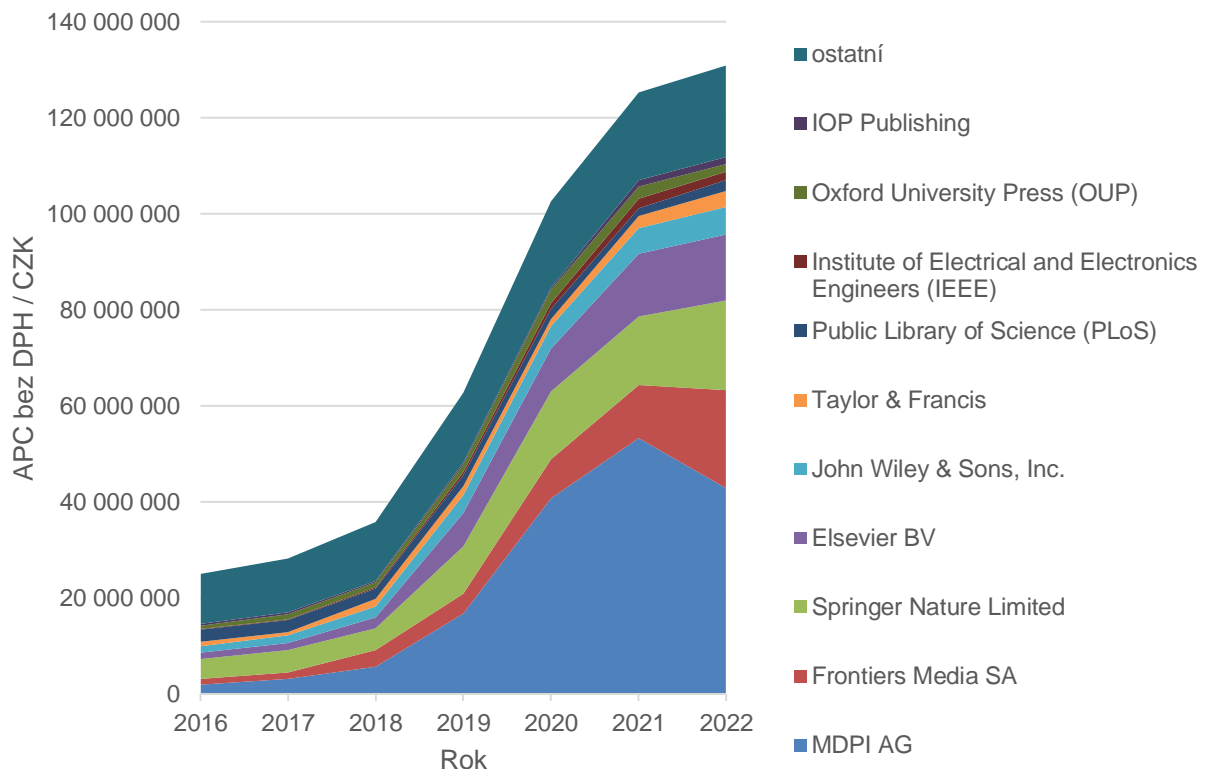
Od roku 2018 do roku 2020 bylo možno zaznamenat téměř exponenciální růst (viz **Obr. 2**), od r. 2021 se zdá, že dochází k jisté stabilizaci a zastavení enormního růstu. Nutno znovu podotknout, že hodnoty objemu financí jsou pouze hrubá čísla a jejich přesnou skutečnou výši nelze na straně CzechELib zcela ověřit, ale sebraná data potvrdila předpověď významného nárůstu vynaložených APC (viz Přílohu 4 pilotní zprávy¹).

Hlavními faktory dynamiky růstu/poklesu je víceméně publikační činnost u pěti vydavatelů – MDPI, Frontiers Media, Springer Nature, Elsevier a John Wiley & Sons, kteří v r. 2022 stáli za cca 85 % všech výdajů za APC – viz **Obr. 3**.

Z **Obr. 3** také vyplývá, že dochází ke změně podílu na trhu – po letech exponenciální expanze dochází poprvé k odlivu autorů od MDPI, zato k silnějšímu nárůstu u Frontiers Media a také Springer Nature.



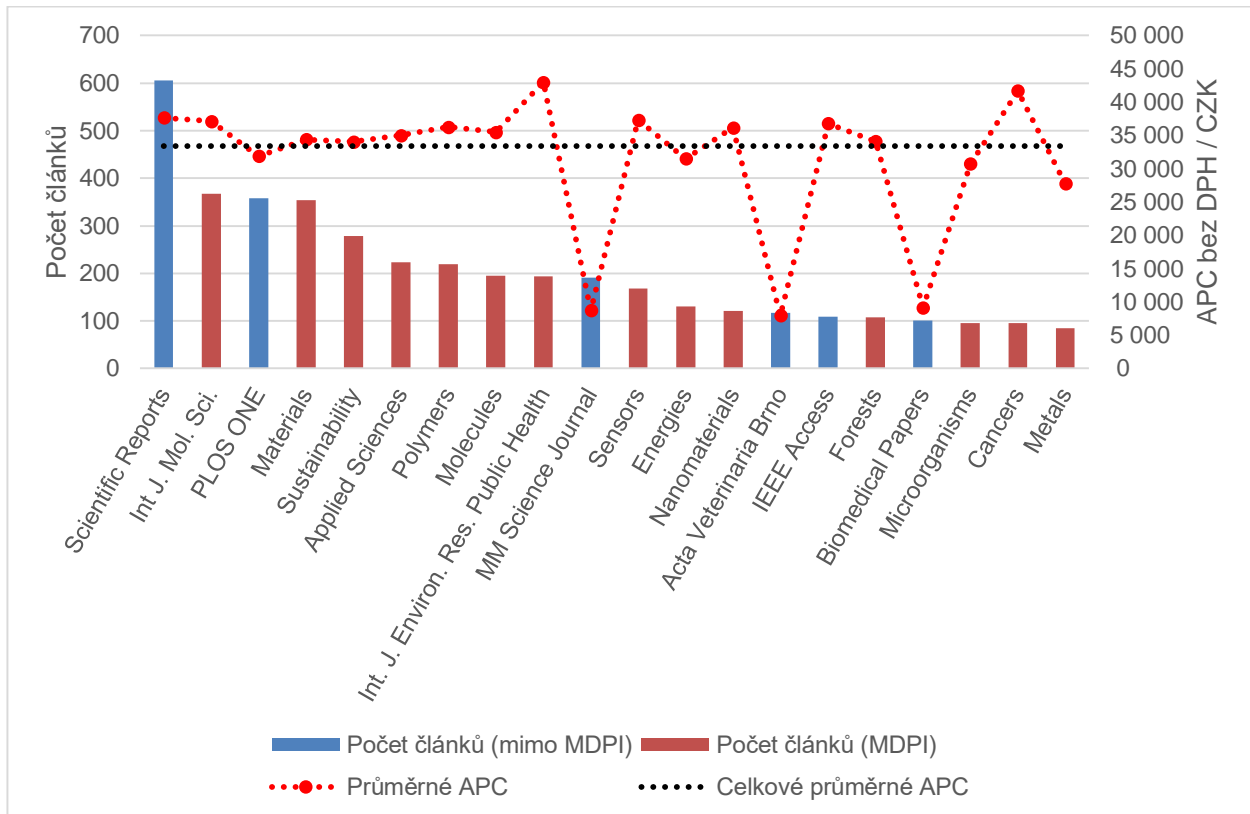
Obr. 2: Nahlášené platby zahrnovaly kromě APC (představovány zejména položkami „hybridní časopis“, „čistý OA časopis“ a „agregovaná hodnota“) i poplatky, které nespadají do definice kategorie APC („Free Access“, „jiný poplatek“)



Obr. 3: Top 10 vydavatelů (podle pozice v r. 2022). Zahrnuje pouze údaje, kde bylo možné určit vydavatele (tj. v záznamu nechyběla metadata k publikaci)

Nárůst celkových výdajů je způsoben zejména razantním nárůstem počtu publikací, nikoli meziročním nárůstem výše jednotlivých APC poplatků.

Průměrná výše APC (viz výše) se blíží typické výši APC v časopisech vydavatelství MDPI a časopisu Scientific Reports, díky váze (= množství publikací) těchto časopisů (viz **Obr. 4**).



Obr. 4: Top 20 titulů (podle počtu článků za celé sledované období). Zahrnuje pouze údaje, kde bylo možné určit titul (tj. v záznamu nechyběla metadata k publikaci). Těchto 20 titulů odpovídá cca 30 % všech článků a taktéž cca 30 % všech výdajů za APC (za celé období). Červeně označené tituly patří vydavatelství MDPI AG

5 Závěr

Zatímco sebraná data o výdajích na předplatné mají relativně vysokou věrohodnost i přesnost, údaje o APC stále poskytují jen velmi nepřesný a neúplný obraz skutečného stavu.

Přes postupná dílčí zlepšení v evidenci a sběru dat na jednotlivých institucích (viz Subjektivní hodnocení kvality získaných dodávaných dat) stále přetrvávají principiální problémy (viz Přílohu 4 pilotní zprávy¹) na mnoha institucích.

Dosavadní přístup tak stále zůstává v suboptimálním nastavení, a to zejména pro následující dva významné případy:

- NTK/CzechELib: Pro model platby za vydavatelské zpracování publikačních výstupů (a následný otevřený přístup) je bezpodmínečně nutné znát data o publikační aktivitě českých autorů velmi přesně, protože ta jsou významným parametrem obchodního modelu vydavatelů.
- Poskytovatelé grantů: Náklady na Open Access publikování jsou často nedílnou (a dnes již nezanedbatelnou) položkou rozpočtů jednotlivých grantových projektů, aktuální důvěryhodná data umožní lépe plánovat a odhadovat alokaci potřebných částek jak na úrovni příjemců, tak na úrovni samotných poskytovatelů.

Pokud nepřistoupíme k efektivní formě centrálně řízeného sběru dat, budeme se muset smířit se silně kolísající kvalitou (a tedy nízkou vypovídací hodnotou) sebraných dat a výrazným zpožděním ve sledování dat (zdlouhavý sběr dat na institucích a následné nutné čištění dat). V jednáních s vydavatelem budeme stále ve stavu „důkazní nouze“, což může v konečném výsledku citelně oslabovat vyjednávací pozici a působit negativně na cenu.

Naštěstí se (zatím) nepotvrdila obava z výrazně klesající ochoty institucí každoročně pracně dohledávat potřebné údaje, tato ochota je však stále v některých případech zajištěna zejména díky požadavku „shora“ (žádost RVVI, resp. ministryně pro vědu, výzkum a inovace).

Vzhledem k předchozí řadě centralizovaných sběrů dat lze důvodně předpokládat, že většina institucí již ve svých účetních a vykazovacích systémech vytvořila předpoklady pro vykazování nákladů na publikování, tj. APC společně s vykazováním dotyčných výsledků výzkumu. To by vedlo k podstatně přesnějším a zejména včasným výsledkům, které by následně umožnily přesněji zaměřit vyjednávání NTK, alokaci finančních prostředků ad.

NTK/CzechELib proto žádá Radu pro VaVal:

- Aby umožnila zavést systematický monitoring spojený s vykazováním údajů do RIV, což znamená přehodnotit prozatím odmítavé stanovisko k takto centralizovanému sběru dat.
- Pokud to by se ukázalo jako neprůchodné, aby alespoň nadále podporovala centralizovaný sběr dat v aktuální podobě.

Jako jedno z možných opatření je též centralizace proplácení APC podle vzoru FWF Austrian Science Fund⁵.

6 Přílohy

6.1 Příloha 1: Zpracování dat

6.1.1 Unifikace a čištění primárních dat

Přes zlepšující se situaci v kvalitě dat oproti prvotnímu pilotnímu sběru¹ bylo přesto nutno provést rozsáhlé a důkladné čištění dat, které zahrnovalo zejména:

- kontrolu finančních údajů:
 - detekce a ruční oprava chyb v číselných formátech
 - identifikace a oprava chybně zadaných měn
- verifikaci identifikátorů (DOI v datech o APC, ISSN v datech o předplatném):
 - oprava syntaktických chyb strojově
 - ověření správnosti DOI / ISSN, které nebyly nalezeny v databázi CrossRef
 - doplnění chybějících DOI / ISSN pomocí jiného identifikátora článku
- sjednocení formulace odpovědí beze změny jejich významu
- ověření věcné správnosti u navzájem si odporujících odpovědí.

Data byla průběžně propojována s databází CrossRef⁶, a to za účelem:

- efektivnější opravy chybných identifikátorů,
- následné standardizace klíčových údajů o časopisech (vydavatel a název titulu),
- obohacení záznamů obsahujících DOI o další informace, zejména typ licence.

Data pro články nebo časopisy, které nejsou indexovány v CrossRef, byla doplněna a standardizována ručně.

6.1.2 Unifikace finančních údajů pro následné analýzy

Kurz: pro potřeby srovnání jednotlivých položek (včetně meziročního) byly všechny finanční údaje převedeny na CZK, a to kurzem deviza prodej k 2. 1. 2023 (tento kurz slouží jako základ pro stanovení tzv. „bezpečného kurzu“ používaného v rámci CzechELib).

DPH: všechny částky byly převedeny na hodnoty bez DPH (očistění od vlivu měnící se DPH). V případě údajů o APC, kde byla uvedena pouze hodnota vč. DPH byla uvažována jednotná sazba 21 % (služba).

Výhodou výše uvedených aproximací je mj. možnost průhlednějšího meziročního srovnání plateb v cizích měnách – očistění od fluktuace kurzu CZK a DPH a díky tomu **snadnější sledování trendů**.

Nevýhodou je vznik mírných nepřesností – např. položky uvedené v CZK (naprostá většina údajů o předplatných), placené v minulých letech, používaly jiný (neúvedený) kurz CZK, nebo v případě položek s DPH mohla být situace podstatně komplexnější.

Vzniklé nepřesnosti se však budou pohybovat spíše v řádu procent a vzhledem k tomu, že cílem sběru dat není účetní přesnost, nýbrž analýza trendů, je tato přesnost zcela přijatelná.

6.2 Příloha 2: Seznam zkratek a termínů

EIZ (Elektronický informační zdroj)	elektronické knihy, časopisy a databáze
OA (Open Access)	otevřený přístup, neboli způsob publikování, který zajišťuje okamžitý, trvalý, svobodný a bezplatný přístup k vědeckým publikacím
Transformační smlouva	smlouva zajišťující převedení části poplatku za čtení na poplatek za publikování
Diamond časopis	náklady na publikování jsou hrazeny samotným vydavatelem nebo jeho (časopisu) sponzorem
Gold OA časopis	časopis publikující výhradně v režimu Open Access, náklady na publikování jsou hrazeny autorem či jeho institucí
Hybridní časopis	časopis, který je jako celek v režimu předplatného (část publikací je přístupná pouze pro předplatitele), ale jednotlivé publikace mohou být zveřejněny v režimu OA za poplatek APC
APC (Article Processing Charge)	publikační poplatek, který hradí autor u publikace, jež má být publikována v režimu OA
DOI (Digital Object Identifier)	identifikátor digitálního objektu ⁷

7 Odkazy

1. Dibuszová, E.; Svoboda, M.; Jirát, J. *Finanční analýza prostředků v oblasti přístupu k vědeckým informacím: Zpráva o plnění Opatření 8 Akčního plánu pro implementaci Národní strategie otevřeného přístupu České republiky k vědeckým informacím na léta 2017 – 2020*; 2020; DOI: <https://doi.org/10.48813/9e4m-4m21>
2. O CzechELib. <https://www.czechelib.cz/cs/21-o-czechelib> (accessed 2.11.2023).
3. OpenAPC. <https://openapc.net/> (accessed 2.11.2023).
4. *The new university Open Access checklist (A practical guide on implementation)*; European University Association 2021; <https://eua.eu/downloads/publications/2021%20oa%20checklist%20final.pdf>.
5. Peer-Reviewed Publications. <https://www.fwf.ac.at/en/research-funding/fwf-programmes/peer-reviewed-publications> (accessed 6.11.2023).
6. Metadata Retrieval. <https://www.crossref.org/services/metadata-retrieval/> (accessed 2.11.2023).
7. What is a DOI? <https://www.doi.org/the-identifier/what-is-a-doi/> (accessed 2.11.2023).