

MORE+BRAINS

Analýza nákladů a přínosů implementace PID v České republice



Foto: Martin Urbanec, CC-BY-SA, https://commons.wikimedia.org/wiki/File:Národní_technická_knihovna_v_Praze,_červenec_2020.jpg

15. prosince 2024



Spolufinancováno
Evropskou unií

MT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

NTK
50°8'14.083"N, 14°23'26.395"E
Národní technická knihovna
National Library of Technology

MORE+BRAINS

O zprávě

Autoři

Dr Phill Jones (hlavní autor), <https://orcid.org/0000-0003-0525-6323>

Josh Brown, <https://orcid.org/0000-0002-8689-4935>

Alice Meadows, <https://orcid.org/0000-0003-2161-3781>

Fiona Murphy, <https://orcid.org/0000-0003-1693-1240>

Citovat jako: Brown, J., Jones, P., Meadows, A., & Murphy, F. (2024) Analýza nákladů a přínosů implementace PID v České republice. <https://doi.org/10.48813/zm93-wy56>.

DOI anglické verze: <https://doi.org/10.48813/x4vd-yj13>

DOI české verze: <https://doi.org/10.48813/zm93-wy56>

Tato práce podléhá licenci [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0).

Publikace je financována z projektu [Czech Academic and Research Discovery Services \(CARDS\)](#), registrační číslo projektu: CZ.02.01.01/00/22_004/0004342.



Spolufinancováno
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



50°8'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology



Shrnutí

V rámci projektu CARDS (Czech Academic and Research Discovery Services)¹ zadala Národní technická knihovna (NTK) společnosti MoreBrains Cooperative zpracování analýzy nákladů a přínosů implementace vybraných perzistentních identifikátorů (PID) do různých typů informačních systémů v českém prostředí výzkumu, vědy a inovací (VaVal).

Perzistentní identifikátory zajišťují dlouhodobý a jedinečný digitální odkaz na entitu. Mezi entity, které jsou nedílnou součástí systému výzkumu, vývoje a inovací, patří organizace (univerzity, subjekty financující výzkum, výzkumná pracoviště atd.), lidé, kteří se podílejí na výzkumu (jako výzkumní pracovníci, technici, pracovníci podpory výzkumu, autoři, recenzenti atd.), přístroje, vybavení, projekty, granty a výstupy výzkumu. PID mohou odkazovat na samotnou entitu (pokud se jedná o digitální objekt) nebo na webovou stránku, která poskytuje přístup k entitě a informace o ní, nebo mohou poskytovat přístup k popisným metadatům o entitě.

Tato analýza plní trojí účel:

1. nezávisle vyhodnotit a kvantifikovat současné využití vybraných PID v prostředí českého výzkumu, vývoje a inovací (VaVal) k roku 2024;
2. posoudit potenciální přínos dalšího rozšiřování využití PID v porovnání s náklady na centralizovanou finanční, administrativní a metodickou podporu jejich využití prostřednictvím projektu CARDS;
3. za čtyři roky (2028) znovu provést analýzu nákladů a přínosů, zjistit, jakého pokroku bylo dosaženo, a formulovat další plány a strategie pro podporu PID po dokončení projektu CARDS.

Tato zpráva se zabývá prvními dvěma z uvedených cílů, neboť kvantifikuje současný stav PID v České republice. V období mezi jejím zveřejněním a opakovanou analýzou v roce 2028 se uskuteční série průzkumů a osvětových aktivit, které budou následně využity v konečném výstupu.

Digitální technologie mají ve výzkumu, vývoji a inovacích mnoho potenciálních výhod, přičemž jako hlavní přínos digitalizace procesů se často uvádí efektivita. Důkazy však hovoří o tom, že administrativní zátěž výzkumných pracovníků je i v dnešní době extrémně vysoká [1]. Náš výzkum ukázal, že informační systémy používané ve výzkumných institucích nejsou dostatečně propojené a ruční zadávání údajů se často opakuje napříč systémy [2].

Tato skutečnost je oficiálně uznávána jako problém českého výzkumu. Národní výzkumná a inovační strategie pro inteligentní specializaci České republiky 2021–2027 (dále jen „národní RIS3 strategie“) uvádí, že „problémem je také zastaralá legislativa a vysoká administrativní zátěž napříč celým VaV² systémem“ [3].

Národní politika výzkumu, vývoje a inovací České republiky 2021+ uvádí, že je třeba „snížit administrativní zátěž ve VaVal, zejm. u administrativy při podání návrhů projektů a vykazování a vyúčtování nákladů na základě stejných pravidel pro všechny poskytovatele“ [4]. Kromě toho

¹ Více o projektu Czech Academic and Research Discovery Services (CARDS) zde: <https://www.techlib.cz/cs/84666-czech-academic-and-research-discovery-services-cards>.

² Věda a výzkum (VaV).

Inovační strategie ČR 2019–2030 obsahuje prioritu „zamezit povinnosti opětovného poskytování dříve poskytnutých informací“ [5]. Metodika použitá pro zjišťování přínosů PID se zaměřuje na opakované využití metadat a na časové úspory a efektivitu, kterou může přinést, a proto úzce souvisí s výše popsány cíli národních strategií a politik. Širší adopce PID přispívá také k naplňování dalších priorit Národní politiky VaVal ČR 2021+, mezi které patří snaha o větší otevřenost výzkumu, rozvoj české znalostní ekonomiky a zvýšení dopadu a společenské relevance výzkumu.

Námi zpracovaná analýza ukazuje, že potenciální celkové finanční úspory z opakovaného použití metadat z registrů PID pro tři klíčové entity (publikace, projekty a granty) a z automatizace procesů jsou vyšší než finanční náklady na integrace a podporu PID. Náš model ukazuje, že **časové úspory plynoucí z adopce PID by mohly dosáhnout až 7 207 člověkodní ročně, což by odpovídalo čistým úsporám ve výši 13 214 713 Kč ročně.**

Kromě úspor dosažených díky PID a podpůrným službám poskytovaným projektem CARDS **přinášejí úspory také konsorcia ORCID a DataCite, jež začala fungovat v rámci národního programu podpory PID vedeného NTK.** Porovnáváme výdaje na individuální členství s poplatky za konsorcia a zjišťujeme, že při současném počtu členských institucí činí **úspora pro české instituce z těchto dvou konsorcií dohromady 3 050 240 Kč ročně.**

Dosažení těchto přínosů závisí na využití PID a integracích rozhraní API (Application Programming Interfaces) PID do různých systémů používaných pro správu výzkumu. Ty umožňují opakované využití metadat spojených s PID v každodenních pracovních postupech a při vytváření nových entit lze informace o nich zachytit jako metadata spojená s PID, která lze sdílet pro následné opakované použití. Zaměřujeme se proto na maximalizaci využití PID a doložení jejich budoucích přínosů a navrhujeme následující doporučení:

- **Provést navazující průzkum, který zjistí počet a rozsah současných a plánovaných integrací PID v českých institucích, a vyhodnotit pokrytí českých výzkumných entit v rámci registrů PID.** Stanovením základní úrovně adopce PID bude možné průběžné hodnocení pokroku.
- **Stanovit střednědobý až dlouhodobý cíl 80% pokrytí všech pěti prioritních PID, doplněný o cíl dosáhnout 60 % integrací PID v relevantních systémech institucí (rozuměno buď v absolutních číslech – počtu systémů – nebo v systémech obsahujících 60 % informací o výzkumu), čímž se zajistí dobrá návratnost investic do adopce PID.**
- **Propagovat význam PID mezi českými institucemi a výzkumnými pracovníky s cílem maximalizovat adopci a pokrytí klíčových prioritních PID.**
- **Zvýšit viditelnost českého výzkumu v celosvětových registrech PID s cílem maximalizovat jeho dohledatelnost, dosah a dopad.** To podpoří naplňování cílů klíčových národních strategií pro rozvoj a podporu výzkumu a vývoje v České republice, což přispěje k posílení mezinárodní viditelnosti a atraktivity českého výzkumu.
- **Implementovat PID do systémů národního významu, jako je IS VaVal, Národní repozitář a další národní systémy a služby.** Díky tomu i menší instituce, které nemají dostatečné technické kapacity nebo mají menší objem výzkumné činnosti, budou moci přímo využívat PID. To jim umožní šířeji sdílet informace o svých aktivitách a díky tomu budou moci účinněji a efektivněji přispívat k růstu české znalostní ekonomiky.

- **Provést průzkum s cílem identifikovat klíčové systémy, v nichž jsou shromažďovány informace o výzkumu** (např. žádosti o granty subjektů financujících výzkum, platformy pro vykazování výzkumných výsledků atd.), a přednostně se na ně zaměřit při integraci PID. To nejen zvýší objem metadat dostupných pro opakované použití prostřednictvím registrů PID, ale také zefektivní tyto služby.

I při konzervativních předpokladech, ze kterých v naší analýze vycházíme, je zřejmé, že přínosy adopce PID v České republice jsou značné. Reálné přínosy pravděpodobně přesáhnou úroveň odhadovanou v této zprávě, avšak jejich dosažení bude vyžadovat cílené strategické investice a systematickou podporu. Tyto investice a podpora by měly být zohledněny při tvorbě budoucích národních strategií a implementačních plánů. Kombinovaný přístup, který zahrnuje centrální služby, dlouhodobou technickou podporu, vzdělávání a zapojení komunit, zajistí spravedlivé a efektivní zavádění PID v České republice. Tento proces by měl být podložen spolehlivými důkazy o míře adopce PID a současné informační zátěži či neefektivním sběru informací. To umožní přesně zacílit aktivity na místa s největší potřebou nebo potenciálním dopadem, a maximalizovat tak dosah a rozsah české sítě PID.

Obsah

O zprávě	2
Autoři	2
Shrnutí.....	3
1 Úvod.....	7
2 Co znamenají PID a jaké mají výhody.....	7
2.1 PID přináší efektivitu	8
2.2 PID podporují širší priority	8
2.2.1 Otevřený výzkum.....	8
2.2.2 Budování znalostní ekonomiky.....	9
3 České výzkumné prostředí.....	11
3.1 Informační systém výzkumu, vývoje a inovací (IS VaVal)	11
3.2 Rozsah výzkumné činnosti	11
3.2.1 Jednoznačná identifikace výzkumných institucí	12
3.2.2 Odhad počtu výzkumných projektů	12
3.2.3 Počet vykázaných autorství	13
3.2.4 Počet vykázaných grantů.....	15
4 Analýza nákladů a přínosů	17
4.1 Celkové potenciální snížení nákladů	17
4.2 Institucionální náklady na implementaci	19
4.3 Náklady na národní program podpory PID	19
4.4 Cílová úroveň adopce	20
4.4.1 Co znamená adopce PID.....	20
4.4.2 Stanovení cílů pro adopci PID.....	21
4.5 Úspory díky činnosti národních konsorcií PID	24
5 Závěry.....	25
Seznam zdrojů.....	27
Příloha A – Výpočet nákladů a přínosů	29
Příloha B Datové zdroje	30
7.1 IS VaVal.....	30
7.2 Digital Science Dimensions.....	30
7.3 ROR	31

1 Úvod

Cílem projektu CARDS (Czech Academic and Research Discovery Services)³ je vybudovat centrální platformu pro správu všech typů informačních zdrojů pro české výzkumné instituce a podílet se na vytvoření společného zázemí pro implementaci datově orientovaných komponent strategie Open Science v ČR, především v rámci pilířů Open/FAIR Data a EOSC. To zahrnuje i podporu adopce PID a souvisejících služeb. V rámci tohoto projektu zadala Národní technická knihovna (NTK) společnosti MoreBrains Cooperative zpracování analýzy nákladů a přínosů vybraných perzistentních identifikátorů (PID) v informačních systémech v prostředí českého výzkumu.

Tato analýza plní trojí účel: 1) nezávisle vyhodnotit a kvantifikovat současné využití vybraných PID v prostředí českého VaVal k roku 2024; 2) posoudit potenciální přínos dalšího rozšiřování užití PID v porovnání s náklady na jejich centralizovanou finanční, administrativní a metodickou podporu prostřednictvím projektu CARDS; 3) za čtyři roky (2028) se k analýze nákladů a přínosů vrátit a zjistit, jakého pokroku bylo dosaženo, a na základě toho poté formulovat další plány a strategie podpory PID po ukončení projektu CARDS.

Tato zpráva se zabývá prvními dvěma z uvedených cílů, neboť kvantifikuje současný stav využití PID v České republice. V období mezi jejím zveřejněním a opětovným provedením analýzy v roce 2028 se uskuteční několik průzkumů a osvětových iniciativ, jejichž výsledky budou následně využity v konečném výstupu.

2 Co znamenají PID a jaké mají výhody

Perzistentní identifikátory (PID) poskytují dlouhodobý a jedinečný digitální odkaz na entitu. Mezi entity, které jsou nedílnou součástí systému výzkumu, vývoje a inovací, patří instituce (univerzity, zřizovatelé, výzkumná pracoviště atd.), lidé podílející se na výzkumu (jako badatelé, technici, pracovníci podpory vědy, autoři, recenzenti atd.), přístroje, vybavení, projekty, granty a výstupy výzkumu. PID mohou odkazovat na samotnou entitu (pokud se jedná o digitální objekt) nebo na webovou stránku, která poskytuje přístup k entitě a informace o ní, nebo mohou zajišťovat přístup k popisným metadatům o entitě.

V posledních 30 letech se výzkum stále více propojuje. Současně dochází k prudkému nárůstu rozsahu výzkumné činnosti, míry publikování a objemu dat pocházejících z výzkumu i dat o výzkumu. Propojené digitální odkazy a spolehlivá metadata mají proto nyní zásadní význam pro provádění a analýzu výzkumu a PID jsou všeobecně uznávány jako základní součást moderního digitálního výzkumného ekosystému. Svědčí o tom nárůst připravovaných národních strategií PID⁴ ve světě a rostoucí význam PID v realizaci klíčových strategických cílů národních i mezinárodních politik, jako je například EOSC (Evropský cloud pro otevřenou vědu) [6].

³ Více o projektu Czech Academic and Research Discovery Services (CARDS) zde:

<https://www.techlib.cz/cs/84666-czech-academic-and-research-discovery-services-cards>.

⁴ Více o pracovní skupině pro národní strategie PID v rámci RDA (Research Data Alliance) zde:

<https://www.rda-alliance.org/rationale/national-pid-strategies-interest-group>.

2.1 PID přinášejí efektivitu

Digitální technologie přinášejí do výzkumu, vývoje a inovací mnoho potenciálních výhod, přičemž za hlavní přínos digitalizace procesů je často považována vyšší efektivita. Fakta však ukazují, že administrativní zátěž výzkumných pracovníků je i v dnešní době extrémně vysoká [1]. Náš výzkum ukázal, že informační systémy používané ve výzkumných institucích jsou často zcela oddělené a ruční zadávání údajů se v nich mnohdy napříč různými systémy opakuje [2].

Tato skutečnost je oficiálně uznávána jako problém českého výzkumu. Národní výzkumná a inovační strategie pro inteligentní specializaci České republiky 2021–2027 (RIS3 strategie) uvádí, že „problémem je také zastaralá legislativa a vysoká administrativní zátěž napříč celým VaV systémem“ [3].

Národní politika výzkumu, vývoje a inovací České republiky 2021+ uvádí, že je třeba „snížit administrativní zátěž ve VaVal, zejm. u administrativy při podání návrhů projektů a vykazování a vyúčtování nákladů na základě stejných pravidel pro všechny poskytovatele“ [4]. Kromě toho Inovační strategie ČR 2019–2030 obsahuje prioritu „zamezit povinnosti opětovného poskytování dříve poskytnutých informací“ [5].

V předchozích analýzách nákladů a přínosů jsme ukázali, že opakované využití metadat z registrů PID může přinést významné úspory času a nákladů na běžné výzkumné procesy [2, 7, 8]. Národní platforma IS VaVal je primárním nástrojem pro vykazování výzkumu v České republice. Tato platforma využívá model, v němž je efektivita dosahována díky „znovuvyužití již jednou zadaných údajů v jiných informačních systémech veřejné správy (napojení na základní registry) a spolupráci při sdílení dat prvotně zadaných do IS VaVal dalšími orgány státní správy, změnách a rozšíření identifikátorů a obsahu záznamů...“ [9].

Opakované využití metadat z PID významně doplňuje tento osvědčený přístup a umožňuje zvýšit efektivitu v rámci institucí. Nárůst efektivity zvyšuje hodnotu stávajících investic, například do kontinuálního vývoje IS VaVal, a zároveň poskytuje argumenty pro investice do integrací a adopce PID. Je však důležité poznamenat, že přínosy PID a metadat znamenají mnohem více nežli zvýšení efektivity.

2.2 PID podporují širší priority

Zatímco zvýšení efektivity lze měřit z hlediska času a peněz, další přínosy plynoucí z komplexního zavedení PID se kvantifikují obtížněji, ale jsou neméně důležité. Úplnější, přesnější a včasější informace o výzkumných aktivitách, výsledcích výzkumu a jejich dopadech jsou klíčové v mnoha oblastech, jež jsou pro vládu a občany České republiky prioritní, včetně následujících příkladů.

2.2.1 Otevřený výzkum

PID jsou všeobecně uznávány jako základní stavební kámen otevřeného výzkumu. Registry PID obsahují značnou část metadat, která jsou pro otevřený přístup (open access; OA) velmi důležitá a která pomáhají sledovat a monitorovat pokrok v dosahování cílů OA. Z tohoto důvodu jsou PID klíčovou součástí technického rámce politik OA, jako je Plán S [10] nebo Doporučení Evropské komise o přístupu k vědeckým informacím a jejich uchovávání, které výslovně uvádí, že „členské státy by měly zajistit, aby instituce financující výzkum, které jsou příslušné pro správu veřejných

prostředků pro financování výzkumu, a akademické instituce, které získávají veřejné finanční prostředky, koordinovaně prováděly politiky a národní akční plány uvedené v bodě 1 na vnitrostátní úrovni tak, že... zajistí, aby byly publikace, které jsou výsledkem financování z veřejných prostředků, snadno identifikovatelné vhodnými technickými prostředky, mimo jiné prostřednictvím metadat připojených k elektronickým verzím výstupů výzkumu a trvalých identifikátorů...“ [11].

Otevřený výzkum však znamená více nežli přístup k publikacím a zasahuje do celého životního cyklu výzkumu. To s sebou přináší jak výzvy, tak i příležitosti.

V současné době nelze zjistit, kolik výzkumných dat vzniká v českých institucích; cenná data mohou ležet na pevných discích v zásuvkách psacích stolů nebo ve starých počítačích. Z tohoto důvodu Národní strategie otevřeného přístupu k vědeckým informacím České republiky na léta 2017–2020 zdůraznila význam ukládání, uchovávání a zpřístupňování dat prostřednictvím digitálních repozitářů: „Otevřená výzkumná data jsou zejména data v digitální podobě pocházející z výzkumných projektů (z experimentů, šetření a měření, a to včetně tzv. metadat [dat, která poskytují informaci o jiných datech] a podrobností o zpracování dat) dostupná bez omezení online všem potenciálním uživatelům“ [12].

Zatímco systematické používání PID pro datové soubory umožní získat jasnou představu o objemu, povaze a rozsahu datových sad produkovaných výzkumnými pracovníky v ČR, politika odkazuje na propojení těchto datových souborů s výzkumnými přístroji atd., které byly použity při jejich tvorbě. Toho lze snadno dosáhnout pomocí PID, které lze vložit do metadat datové sady. Tato propojení výzkumných výstupů s přístroji, osobami, finančními prostředky nebo projekty jsou rovněž zásadní pro integritu výzkumu. Kombinace otevřených výzkumných postupů a otevřených PID nabízí obrovskou příležitost jak zlepšit přehled a povědomí o dopadech a výsledcích národních i mezinárodních investic do českého výzkumu.

2.2.2 Budování znalostní ekonomiky

Jádrem RIS3 strategie je snaha vybudovat v České republice silnou znalostní ekonomiku: „Dlouhodobá strategická vize formuluje základní směr rozvoje ČR s důrazem na sféru znalostní ekonomiky a na transformaci hospodářství tak, aby rostla konkurenceschopnost založená na inovacích...“ [13].

Zároveň RIS3 strategie přiznává, že „řízení VaVal se zatím dostatečně neopírá o strategickou inteligenci, tj. o strategické informace o stavu a dynamice vývoje systému VaVal v ČR a jeho jednotlivých částí, o nových trendech a potřebách, na něž by měl výzkum z hlediska společenské relevance reagovat, jakož i o dopadech realizovaných politik a opatření“ [3].

Má-li být vybudována dlouhodobě silná znalostní ekonomika, je nezbytné mít k dispozici strategické poznatky a podklady. Národní politika výzkumu, vývoje a inovací si klade mj. za cíl: „Zajistit strategické řízení systému VaVal na základě doložitelných faktů (tzv. „evidence-based“ politiky)“ [4]. Mnohé z ukazatelů kvality, které tvoří základ těchto strategických cílů, lze nejúčinněji sledovat pomocí PID (například ROR pro afiliace mezinárodních partnerů, ORCID pro identifikaci spolupracovníků a sledování publikací, geografických lokací atd. na globální úrovni).

Investice do centralizované podpory PID urychlují jejich adopci a integrace do systémů. Rozšířené pokrytí PID následně povede ke zvýšení efektivity, díky čemuž se náklady na integrace více než vrátí. Kromě těchto úspor navíc využití PID zlepší strategické poznatky a přehled o výzkumné činnosti a dosažených výsledcích a může přispět k naplňování hlavních národních priorit v řadě oblastí.

3 České výzkumné prostředí

3.1 Informační systém výzkumu, vývoje a inovací (IS VaVal)

V posledních letech byly v České republice učiněny významné kroky ke sjednocení vykazování a shromažďování informací o výzkumné činnosti veřejně přístupným a transparentním způsobem. Národní systém pro evidenci výzkumné činnosti Informační systém výzkumu, vývoje a inovací (IS VaVal) [14] je veřejný informační systém, který shromažďuje, zpracovává, využívá a poskytuje informace o výzkumu, vývoji a inovacích financovaných z veřejných prostředků ČR. Systém se skládá ze čtyř vzájemně propojených databází, které zahrnují:

- CEA (Centrální evidence aktivit VaVal) - programy VaVal,
- VES (Evidence veřejných soutěží ve VaVal) - veřejné soutěže,
- CEP (Centrální evidence projektů) - udělené granty,
- RIV (Rejstřík informací o výsledcích) - výstupy výzkumu.

Databáze CEA, VES a CEP jsou naplňovány poskytovateli finančních prostředků a provozovatelem IS VaVal, kteří vkládají údaje do aplikace nazvané RoP (administrační rozhraní pro poskytovatele podpory). Pro naplnění databáze RIV jsou údaje o výzkumné činnosti vykazovány jednotlivými výzkumnými institucemi, jež jsou financovány z veřejných zdrojů a jež své údaje zadávají prostřednictvím aplikace s názvem VaVER (editační rozhraní pro příjemce). Z těchto údajů jsou vytvářeny dávkové XML soubory, které jsou zasílány jednotlivým českým agenturám financujícím výzkum. Ty pak shromažďují a třídí údaje o výstupech výzkumu ze všech institucí, které od nich dostávají finanční prostředky a konsolidují je do nových XML souborů. Tyto soubory jsou následně zaslány provozovateli IS VaVal, který je nahraje do databáze. Výsledkem je veřejně přístupný zdroj, jenž propojuje programy s výzvami k financování, udělenými granty a výslednými výstupy výzkumu.

3.2 Rozsah výzkumné činnosti

Hodnota PID vyplývá z různých případů použití popsanych v oddíle 2. Patří mezi ně snížení administrativní zátěže, zvýšení automatizace a zlepšení kvality, včasnosti a úplnosti dat a metadat pro rozhodování na národní, institucionální i individuální úrovni. Tato analýza se zaměřuje na okamžité a přímé úspory nákladů a času spojené se snížením zátěže výzkumných pracovníků a pracovníků podpory výzkumu, ke kterému dochází díky minimalizaci nutnosti ručně opakovaně zadávat metadata do více informačních systémů pro účely zveřejňování, správy a vykazování výzkumu.

Náklady na zbytečné ruční opakované zadávání údajů a následné úspory nákladů závisí na rozsahu činnosti, počtu entit a rozsahu metadat, které tyto entity vzájemně propojují, včetně:

- výzkumných institucí,
- výzkumných projektů v rámci institucí,
- financovaných grantů,
- výzkumných publikací,
- autorství publikací.

3.2.1 Jednoznačná identifikace výzkumných institucí

Databáze RIV obsahuje názvy a jedinečná IČO⁵ institucí, které vykazují publikace a další výstupy svých výzkumných pracovníků. V současné době nejsou IČO přímo přiřazena k identifikátorům Registru výzkumných organizací (ROR)⁶, avšak IS VaVal plánuje zavedení ROR v roce 2025 [15]. Do té doby lze mapování provést vyhledáním názvů institucí v databázi ROR. Na základě součtu unikátních IČO v databázi RIV za období 2018–2022 lze identifikovat 320 jedinečných institucí provádějících výzkum v ČR, které vykazují publikace prostřednictvím IS VaVal.

Pro účely níže uvedené analýzy nákladů a přínosů jsme pracovali s počtem 110 institucí, na který se zaměřuje projekt CARDS. Víme, že 92 % výsledků výzkumu v databázi RIV vykazuje právě 110 institucí s nejvyšším objemem výsledků. Navíc ne všechny české instituce provádějící výzkum budou potřebovat vlastní integrace PID, protože budou využívat několik národních služeb se zabudovanou podporou PID (většina z nich je v současné době ve vývoji v rámci iniciativy EOSC [16]), například Národní repozitář.

3.2.2 Odhad počtu výzkumných projektů

Projekty jsou důležitou součástí správy informací o výzkumu, neboť představují práci odvedenou v rámci výzkumných institucí. Neexistuje však shoda na tom, jak projekty definovat, popisovat nebo počítat a různé zainteresované strany chápou projekty odlišně. Například poskytovatelé finančních prostředků často definují projekty jako udělené granty, ale výzkumné projekty v rámci institucí mohou být financovány z několika grantů, nebo naopak nemusejí být financovány vůbec. V naší analýze definujeme projekty jako samostatné části výzkumné práce v rámci institucí, protože to více odpovídá tomu, jak jsou výzkumné činnosti v institucích prováděny.

V celosvětovém měřítku se výzkumné projekty sledují špatně. Informace o nich jsou často rozptýleny v rámci institucí v elektronických laboratorních denících (ELN), ekonomických systémech institucí či systémech pro evidenci výzkumné činnosti (CRIS). Nový identifikátor výzkumné činnosti (RAiD)⁷ tento problém řeší tím, že projekty kodifikuje jako entitu s definovaným schématem metadat, což bude v budoucnu velmi cenné pro sledování a hodnocení výzkumu. Prozatím předpokládáme, že počty projektů odpovídají úrovni financování výzkumu. Dřívější odhady počtu aktivních projektů ve Spojeném království (UK) [7] byly upraveny na základě zveřejněných statistik o financování výzkumu podle OECD⁸. Na základě tohoto přístupu odhadujeme, že v České republice je rámcově ve sledovaném období řešeno 9,5 % z celkového počtu výzkumných projektů ve Spojeném království, tj. 4 754 aktivních projektů, které je třeba vykazovat. Některé z těchto projektů mohou představovat víceleté aktivity, jiné mohou být mnohem kratší. Náš odhad však není

⁵ Identifikační číslo osoby IČO je v České republice osmimístné identifikační číslo právnické osoby, podnikající fyzické osoby nebo organizační složky státu.

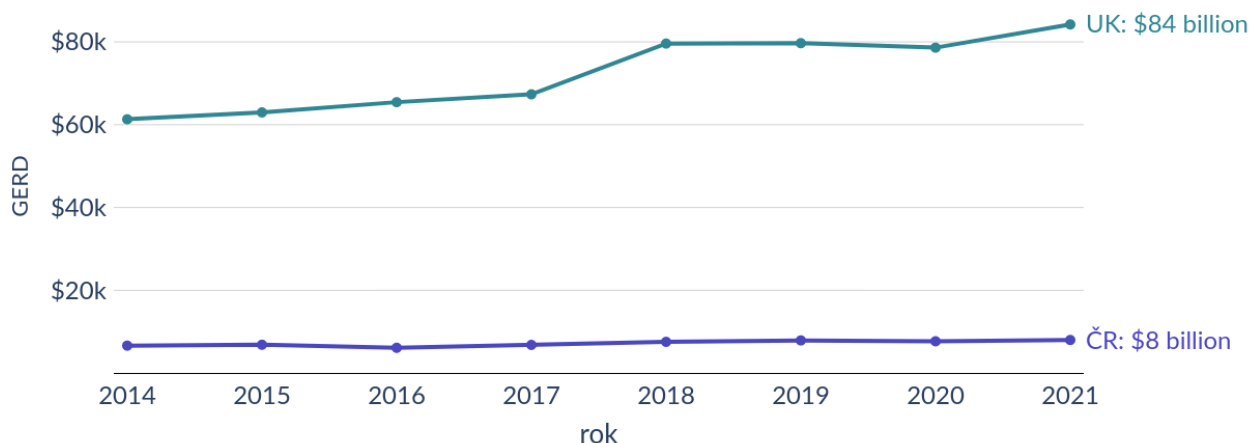
⁶ Registr výzkumných organizací (Research Organization Registry, ROR) je globální, komunitou vedený registr otevřených trvalých identifikátorů výzkumných institucí, který se postupně stává faktickým standardem pro PID institucí. Další informace o ROR naleznete na adrese: <https://ror.org/>.

⁷ RAiD je jedinečný identifikátor pro výzkumné projekty a činnosti, který vyvinula organizace Australian Research Data Commons (ARDC). Rychle se prosazuje na mezinárodní úrovni a je podpořen normou ISO. Další informace naleznete na adrese <https://raid.org/>.

ISO 23527:2022 Information and documentation — Research activity identifier (RAiD) [Internet]. 2022 [citováno 11. 10. 2024]. Dostupné z: <https://www.iso.org/standard/75931.html>.

⁸ Údaje OECD jsou k dispozici zde: <https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm>.

vázán na datum zahájení nebo ukončení projektu, ale je odvozen z porovnání úrovně financování výzkumu s referenční hodnotou (úrovní projektové činnosti ve Spojeném království).



Obrázek 1: Celkové náklady na VaV (GERD) ve Spojeném království a v České republice. Tyto údaje byly použity jako ukazatele velikosti výzkumného prostředí při výpočtu škálovacího koeficientu pro odhad počtu projektů.

3.2.3 Počet vykázaných autorství

Jak bylo popsáno v kapitole 3.1, databáze RIV obsahuje záznamy o publikacích, které jsou výsledkem výzkumu financovaného z veřejných prostředků ČR. U každého záznamu instituce uvádí počet svých výzkumných pracovníků spojených s danou publikací. RIV obsahuje širokou škálu typů výstupů, včetně patentů, softwaru, prototypů a několika dalších. Pro účely naší analýzy jsme se však zaměřili pouze na následující typy výstupů:

- články v recenzovaných odborných časopisech,
- odborné knihy,
- kapitoly v odborných knihách,
- články/stati ve sbornících.

Ne všechny záznamy v databázi RIV mají přiřazený identifikátor DOI – za období 2018–2022 obsahuje DOI přibližně 66 % z 273 291 záznamů vykázaných výsledků. Pro odhad počtu publikací jsme proto použili DOI, pokud bylo k dispozici. V případech, kdy DOI k dispozici nebylo, jsme použili názvy, abychom publikace vzájemně deduplikovali a jednoznačně identifikovali.

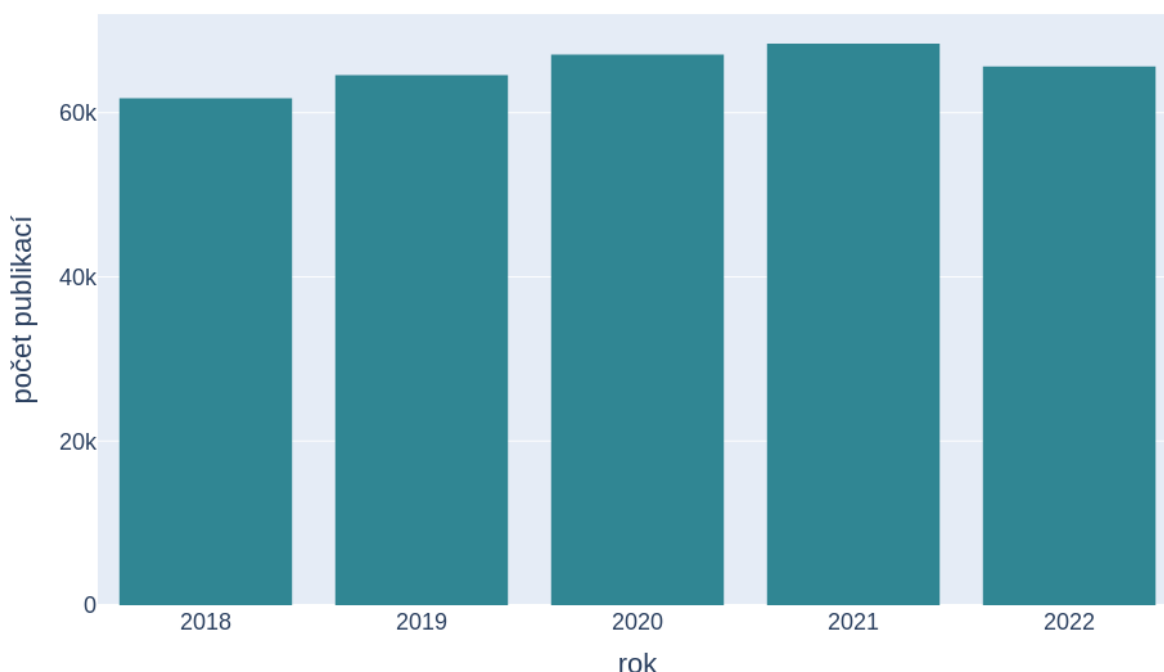
Do databáze RIV se povinně vykazují publikace financované z českých veřejných zdrojů, avšak nejsou v ní spolehlivě zahrnuty publikace pocházející z výzkumu, který není financován z veřejných prostředků, nebo je přímo financován ze zahraničních zdrojů, např. z Evropské unie. Pokud bychom tedy brali v úvahu pouze údaje z databáze RIV, nezapočetali bychom celou administrativní zátěž, které čelí výzkumní pracovníci a pracovníci podpory výzkumu. Abychom tuto skutečnost částečně korigovali, zahrnuli jsme do naší analýzy rovněž údaje o publikacích z databáze Digital Science Dimensions (Dimensions).⁹ Dimensions je databáze publikací, udělených grantů, patentů,

⁹ Daniel W. Hook, Simon J. Porter a Christian Herzog, 'Dimensions: Building Context for Search and Evaluation', *Frontiers in Research Metrics and Analytics*, č. 3, str. 23, srpen 2018, <https://doi.org/10.3389/frma.2018.00023>.

výzkumníků a institucí, která je vytvořena z kombinace otevřených i licencovaných proprietárních zdrojů dat.

Abychom snížili riziko dvojího započítání publikací, sestavili jsme dotaz API pomocí jazyka Dimensions Search Language (DSL) za účelem získání publikací autorů spojených s českými institucemi, jež mají výzkumné granty financované organizacemi mimo Českou republiku. Například článek nalezený v databázi Dimensions, který byl publikován autorem z Univerzity Karlovy a financován Evropskou komisí, byl započítán, zatímco článek téhož autora financovaný Grantovou agenturou ČR nikoli.

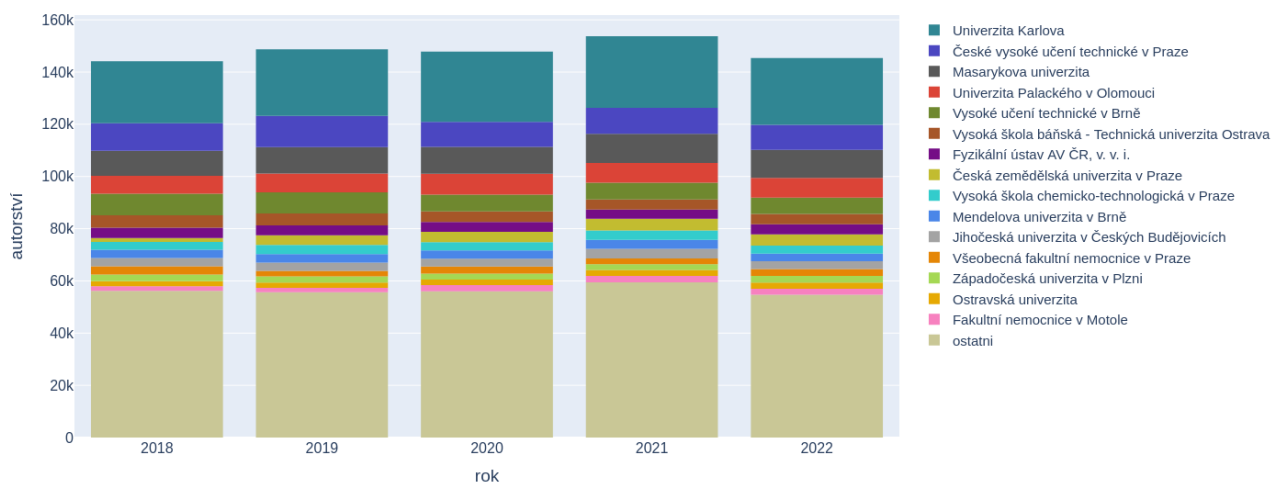
Na základě sloučení údajů z databází RIV a Dimensions odhadujeme, že v České republice vzniká přibližně 67 000 publikací ročně, jak ukazuje obrázek 2. Konečný odhad je pravděpodobně podhodnocený, protože z dat Dimensions jsme mohli zahrnout pouze publikace, u nichž existuje explicitní vazba mezi publikací a jiným než českým poskytovatelem financí. Nemohli jsme tedy započítat publikace, které jsou výsledkem výzkumné činnosti bez finanční podpory, nebo publikace, u nichž se v datovém souboru Dimensions nepodařilo identifikovat zdroj financování, například kvůli špatně formátované sekci o poděkováních a využitých zdrojích financování.



Obrázek 2: Sloučením dat RIV a údajů o mimočeských poskytovatelích finančních prostředků z databáze Dimensions odhadujeme, že v ČR vzniká přibližně 67 000 publikací ročně.

Pro naši konečnou analýzu potřebujeme znát nejen počet publikací, ale také počet autorství, protože časové a finanční náklady na opakované zadávání metadat článků jsou úměrné počtu autorů, kteří musí uvádět, že jsou s každou publikací spojeni. Naštěstí je tento výpočet díky struktuře RIV poměrně jednoduchý, protože každá česká instituce vykazuje počet autorství u každé publikace. U dat Dimensions jsme provedli parsování autorství a afiliací a odfiltrovali jsme autory bez české afiliace.

Obrázek 3 ukazuje odhadovaný počet autorství na základě sloučení údajů RIV a Dimensions za jednotlivé roky, rozdělený podle institucí; 15 institucí s nejvyšším počtem autorství je znázorněno samostatnými barvami. Protože neexistuje přímé mapování identifikátorů Dimensions (Dimensions ID) s IČO, namapovali jsme jak Dimensions ID, tak názvy institucí z údajů RIV na ROR pomocí rozhraní ROR API.



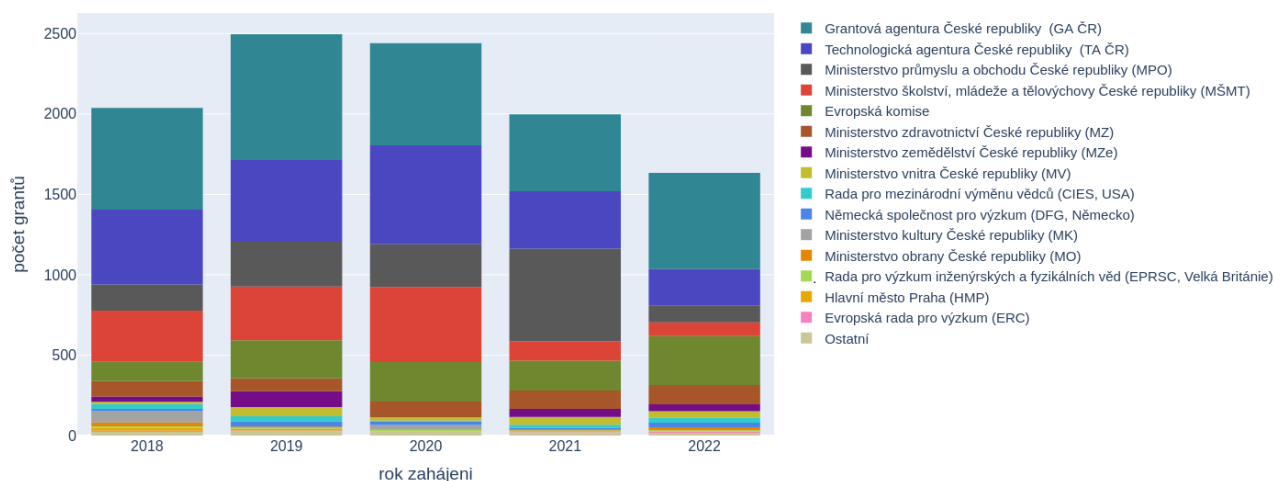
Obrázek 3: Sloučením RIV a údajů o mimočeských poskytovatelích financí z databáze Dimensions a připočtením autorů afiliovaných s českými institucemi odhadujeme přibližně 149 000 autorství ročně.

Jak je patrné z obrázku 3, zprůměrováním za tři roky v letech 2020 až 2022 získáme hodnotu 148 992 autorství.

3.2.4 Počet vykázaných grantů

Databáze CEP obsahuje záznamy o grantech udělených českými veřejnými poskytovateli finančních prostředků českým výzkumným institucím. Abychom získali úplnější obraz o českých výzkumných grantech, zkombinovali jsme údaje z CEP s údaji z databáze Dimensions, která explicitně zahrnuje údaje o projektech VaV přímo financovaných EU jako jsou Horizon Europe, H2020 a jejich předchůdci. Tyto údaje odpovídají datům v databázi CORDIS¹⁰.

¹⁰ Více informací o databázi CORDIS naleznete zde: <https://cordis.europa.eu/>.



Obrázek 4: Počet grantů získaných českými institucemi na základě údajů obsažených v databázi CEP v kombinaci s údaji z databáze Dimensions.

Údaje o grantech z CEP jsme zkombinovali s údaji z Dimensions o grantech udělených českým institucím poskytovateli financí, kteří nesídlí v České republice. Na základě průměru za tři roky (2020–2022) odhadujeme, že každý rok je uděleno 2 024 grantů.

4 Analýza nákladů a přínosů

Analýza použitá v této zprávě vychází z metodiky, kterou jsme vyvinuli pro posouzení udržitelnosti programu na podporu adopce PID pro organizaci Jisc ve Spojeném království [17]. Od té doby byla použita v podobném kontextu pro Austrálii v práci financované organizací Australian Research Data Commons (ARDC) a Australian Access Federation (AAF) [2] a rovněž pro Irsko v práci financované organizací National Open Research Forum (NORF) [8] (více informací o metodice viz uvedené zprávy). Výpočty použité v této analýze jsou podrobně uvedeny v příloze A.

V analýze jsme vyčíslili přímé úspory nákladů, kterých bude dosaženo při integraci registrů metadat PID do informačních systémů tím, že se odstraní ruční opakované zadávání údajů o výzkumné činnosti. Tyto integrace by umožnily jednoznačnou identifikaci grantů, projektů a publikací, čímž by se odstranily problémy spojené s deduplikací při analýze výkonnosti výzkumu a výzkumných činností. Tyto tři entity byly použity pro naši analýzu jako hlavní, protože metadata s nimi spojená jsou široce opakovaně používána, což umožňuje kvantifikaci přínosů. PID pro další entity spojené s výzkumem, jako jsou osoby nebo instituce, umožňují přesnou identifikaci příslušných grantů, publikací atd. a směřování informací do institucionálních systémů. K přínosům těchto PID patří úspora času při zadávání metadat, o níž pojednává tato zpráva, jakož i další typy zvýšení efektivity procesů, pro které však nebyly při této analýze k dispozici podklady.

Kromě toho lze informace, jako jsou názvy, autoři, afilice, abstrakty, data atd., automaticky vkládat do informačních systémů institucí, poskytovatelů finanční podpory a vládních institucí. Tyto úspory srovnáváme s náklady na implementaci integrací PID ve 110 institucích, na něž se zaměřuje projekt CARDS. Nakonec představujeme dva scénáře: jeden, v němž výzkumné instituce hradí samy veškeré náklady spojené s integracemi PID do svých vlastních systémů, a druhý, v němž využijí centrálního programu podpory PID, který běží v NTK. Kromě toho vypočítáváme dosavadní úspory dosažené díky konsorciím ORCID a DataCite, jež jsou podporována projektem CARDS, a to porovnáním nákladů konsorcia s náklady, které by vznikly institucím, kdyby se staly individuálními členy.

4.1 Celkové potenciální snížení nákladů

Pro výpočet předpokládaných úspor nákladů byly použity odhady ročních počtů projektů, grantů a autorství vypočtené v části 3.2. Odhady pro každou z těchto tří entit jsme vynásobili odhadovaným počtem případů, kdy výzkumní pracovníci a pracovníci podpory výzkumu zadávají informace do různých systémů sami. Vycházíme přitom z našich vlastních předchozích průzkumů pracovních postupů v akademickém výzkumu [2]. Mezi tyto systémy patří systémy pro evidenci výzkumné činnosti (CRIS), interní výkaznictví, repozitáře, finanční systémy a další.

Celkové promarněné náklady spojené s opakovaným zadáváním metadat byly vypočteny tak, že se množství času, který tento proces zabere [17, 18], vynásobilo průměrnými mzdami a režijními náklady akademických výzkumných pracovníků a univerzitních pracovníků podpory výzkumu, přičemž tyto údaje byly převzaty ze systému ISPV (Informace a statistiky o průměrném výděлку).¹¹

¹¹ Údaje z Informačního systému o průměrném výděлку (ISPV) jsou dostupné ke stažení zde: <https://www.ispv.cz/cz/Vysledky-setreni/Aktualni.aspx>. Pro analýzu byly použity průměrné mzdy pro „Odborné pracovníky v administrativě a správě organizací“ (ISCO 3343) a „Učitelé na vysokých a vyšších odborných školách“ (ISCO 2310).

Celkové potenciální úspory za sektor							
	Počet	# opakovaných zadávání	# minut /opak. zadávání	Roční úspory času (člověkodny)	Náklady (za autora a minutu)		Roční finanční úspory
Metadata publikací	148,992	3.1	6.73	6,476	8.58 Kč	- pracovník podpory výzkumu	26,666,605 Kč
					10.45 Kč	- výzkumník	32,482,306 Kč
					Průměr		29,574,456 Kč
Metadata grantů	2,024	3.25	10	137	8.58 Kč	- pracovník podpory výzkumu	564,315 Kč
					10.45 Kč	- výzkumník	687,387 Kč
					Průměr		625,851 Kč
Metadata projektů	4,754	6	10	594	8.58 Kč	- pracovník podpory výzkumu	2,447,128 Kč
					10.45 Kč	- výzkumník	2,980,821 Kč
					Průměr		2,713,975 Kč
Celkové potenciální roční úspory díky automatickému zadávání klíčových metadat pomocí propojení přes API:				7,207	32,914,281 Kč		

Tabulka 1: Celkové potenciální úspory času a peněz při automatické synchronizaci všech metadat k publikacím, grantům a projektům do informačních systémů institucí zapojených do výzkumu v České republice.

4.2 Institucionální náklady na implementaci

Implementace PID musí být pro výzkumné instituce finančně udržitelná. V rámci několikaletých mezinárodních konzultací [20] bylo identifikováno a následně ověřeno v analýzách pracovních postupů v akademickém výzkumu [21] pět prioritních PID, které jsou nezbytné pro zachycení informací o výzkumné činnosti a pro jejich automatický přenos mezi systémy:

- Digitální identifikátory objektů (DOI) pro granty všeho druhu,
- Identifikátory ORCID (Open Researcher and Contributor Identifier) pro osoby podílející se na výzkumu,
- Identifikátory výzkumné činnosti (RAiD) pro výzkumné projekty,
- Identifikátory ROR (Research Organization Registry) pro instituce,
- DOI pro výstupy výzkumu (zejména články a data).

Naše odhady nákladů na institucionální implementaci vycházejí z práce Johnsona a kol. [18], kteří došli k závěru, že implementace ORCID zabere přibližně 40 člověkodnů. Tato časová zátěž zahrnovala nejen technickou integraci, ale také tvorbu dokumentace a zajištění osvěty a vzdělávání. Ve skutečnosti by v případě instituce implementující několik PID bylo možné odvedenou práci a získanou zkušenost spojenou s každou jednotlivou integrací alespoň částečně znovu využít, zejména v oblasti školení zaměstnanců, osvěty a komunikace. Kromě toho by některé integrace byly méně náročné než ORCID. Například pro většinu institucí by integrace s ROR (což je centrálně spravovaná databáze institucí) vyžadovala pouze schopnost číst údaje z registru ROR na rozdíl od složité integrace s ORCID (do které se musí zapojit více zúčastněných stran, včetně koncových uživatelů). Odhadujeme proto, že integrace všech pěti prioritních PID by stála přibližně čtyřnásobek nákladů na integraci ORCID. Náklady na implementaci jsme rozložili do pěti let a opět jsme vycházeli z údajů o mzdách v systému ISPV a zprůměrovali jejich hodnotu odpovídající řadě příslušných technických rolí (více informací viz příloha A).

Jak již bylo zmíněno, předkládáme dva scénáře – jeden, ve kterém počítáme náklady na implementaci PID bez centrální podpory z programu rozvoje PID při NTK, a druhý, jenž tuto podporu využije. Projekt CARDS hraje při snižování nákladů spojených s implementací na institucionální úrovni velmi důležitou roli, která spočívá v osvětě, vzdělávání a sdílení znalostí. Abychom zohlednili tento pozitivní efekt na institucionální úrovni, předpokládáme, že s podporou na národní úrovni by se doba potřebná na integraci pěti prioritních PID zkrátila na čas odpovídající pouhému dvojnásobku doby potřebné pro jednu integraci ORCID.

4.3 Náklady na národní program podpory PID

V úvahu je třeba vzít i náklady na samotný projekt CARDS, které zahrnují provoz českých konsorcií ORCID a DataCite, jakož i poskytování řady podpůrných a komunitních služeb. Při zvažování nákladů na implementaci jsme se proto zaměřili na náklady na zaměstnance, kteří budou poskytovat podpůrné služby, včetně koordinátora, technického manažera komunity, manažerů pro osvětu a komunikaci, právníků a dalších podpůrných pracovníků.

Nezahrnuli jsme náklady na poplatky samotných konsorcií, protože ty jsou posuzovány samostatně v části, v níž se zabýváme úsporami konsorcií.

4.4 Cílová úroveň adopce

PID představují propojené entity, a protože usnadňují sdílení informací, platí, že čím více institucí využívá integrace PID, tím cennější tyto integrace pro každou instituci jsou. To znamená, že dílčí ekonomický přínos není lineárně závislý na počtu integrací PID. Naopak, pokud je počet institucí s integracemi PID nízký, přínos každé nové integrace je omezený. S růstem sítě institucí s integracemi PID se však přínos pro nově integrující instituce zvyšuje, až nakonec dosáhne stabilní úrovně. K modelování přínosu zvyšujícího se s rostoucí adopcí PID jsme použili logistickou funkci, což je běžná metoda používaná k modelování takových efektů [22].

4.4.1 Co znamená adopce PID

Adopce PID zde znamená využívání PID a metadat v systémech s cílem ušetřit čas a úsilí při opakovaném využívání metadat a automatizaci procesů. Adopce závisí na dvou vzájemně závislých faktorech: pokrytí PID a integracích PID.

Pokrytí PID je zásadní pro podporu integrací PID. Pokud má ORCID iD pouze 5 % výzkumných pracovníků, budou instituce jen velmi málo motivovány k tomu, aby věnovali zdroje na integraci rozhraní API ORCID do svých systémů. Čím vyšší je úroveň pokrytí PID, tím vyšší je potenciální hodnota integrace PID. Integrace PID umožňuje plně využívat výhody PID, umožňuje opakované použití metadat, čímž šetří čas a peníze výzkumných pracovníků a pracovníků podpory výzkumu.

Pokrytí PID

Pokrytí lze posuzovat a hodnotit dvěma způsoby:

1. Pokud lze stanovit celkový počet výskytů dané entity (např. počet publikací nebo grantů), můžeme tento celkový počet porovnat s počtem PID v příslušném registru či registrech. Tuto míru pokrytí PID pak lze vyjádřit jako procento z celkového počtu známých entit, např. 90 % českých výzkumných institucí má ROR ID a nedávná analýza založená na údajích RIV naznačila, že adopce ORCID iD mezi výzkumnými pracovníky v České republice se blíží 60 % [23].
2. Pokud tyto informace nejsou k dispozici (např. nemáme spolehlivé údaje o celkovém počtu datových sad v daném čase), můžeme místo toho sledovat změny v počtu PID registrovaných pro každou entitu v průběhu času. To lze měřit sledováním absolutního počtu registrovaných PID a počtu registrací za rok pro každou z entit.

Tyto informace pak mohou být využity při stanovení realistických cílů pro zlepšení pokrytí. Například vzhledem k tomu, že v současné době je míra rozšíření ORCID již 60 %, zdá se dosažení 80 % jako realistický cíl. V případě, že není znám celkový počet entit, je zapotřebí stanovit cíle pokrytí PID pro každou třídu entit na základě dostupnosti PID pro tuto entitu, současného počtu registrovaných PID a rozsahu úsilí či změn potřebných k široké implementaci těchto PID.

Integrace PID

Integrace PID nastává, když má instance softwaru (konkrétní systém) aktivní připojení k rozhraní API registru PID a načítá metadata z registru, přidává metadata k záznamům nebo obojí. Úroveň integrace se měří počtem aktivních propojení API a rozsahem jejich používání v době šetření.

4.4.2 Stanovení cílů pro adopci PID

V současné době nemáme přístup k údajům, které by umožnily vyhodnotit počet a rozsah integrací PID v informačních systémech v ČR. Jedinou spolehlivou metodou, která byla vyvinuta pro hodnocení integrací, je použití průzkumů. Tento přístup je osvědčenou metodou zjišťování úrovně integrací a využívání PID a byl vyvinut a zdokonalen při řešení projektů vedených OCLC a EuroCRIS [24], na nichž se podílely i týmy společnosti MoreBrains pro Spojené království [25] a Irsko [26].

V této fázi naší práce jsme se zaměřili na pokrytí PID, ale v dalších fázích bude zapotřebí provést průzkum za účelem změření úrovně integrací PID. Pokud bude průzkum proveden opakovaně, může být použit k měření změn v těchto úrovních, a tudíž ke zjištění pokroku na cestě k dosažení cílů adopce.

V tabulce č. 2 jsou uvedeny čisté souhrnné úspory všech českých výzkumných institucí plynoucí z implementace pěti prioritních PID. Jsou uvedeny dva scénáře. V prvním jsou čisté úspory z oddílu 4.1 poníženy náklady na implementaci v každé ze 110 institucí bez finanční podpory z národního programu financovaného z projektu CARDS. V tomto případě úspory z omezení opakovaného zadávání údajů stačí na pokrytí nákladů na implementaci na úrovni těsně pod 60 % adopce. Při 80% adopci činí úspora 7 870 137 Kč ročně. Ve druhém scénáři, který počítá s podporou z projektu CARDS, úspory převyšují náklady při úrovni adopce mezi 50 a 55 %; při 80% adopci dosahují úspory 12 245 728 Kč.

Jak bylo uvedeno výše, úspěch adopce PID závisí na dostatečně vysokém pokrytí, které motivuje k integracím, a rovněž na takové úrovni integrace, která umožňuje opakované použití metadat ve velkém měřítku. Mají-li úspory převýšit investice do integrací PID a do podpůrných služeb, je nutné dosáhnout 55 % možného opakovaného použití metadat. Proto navrhuje, aby byl stanoven cíl pro integrace alespoň v této výši, upravený podle úrovně pokrytí.

Přínosy zavedení PID lze zhodnotit, pokud se podaří zachytit dostatečně vysoké procento výzkumné činnosti českých institucí v systémech s aktivní integrací PID. Dále je nutné, aby měl dostatečný počet vstupů a výstupů této výzkumné činnosti PID. Zaměřením se na integrace PID v institucích, které produkují nejvíce výzkumné činnosti, se dosáhne cíle neefektivněji. Tím však hrozí, že menší výzkumné instituce nebo entity s menším objemem výzkumné činnosti zůstanou pozadu. Tyto instituce mohou mít prospěch ze zavedení PID díky tomu, že budou využívat integrace PID v národních systémech, které mohou zefektivnit práci jejich výzkumníků a pracovníků podpory výzkumu.

Pokud bude dosaženo námi doporučeného cíle 80% pokrytí prioritních PID, pak by k dosažení 55% adopce bylo třeba dosáhnout integračního cíle 68,75 % (80 % z 68,75 je 55). Jak je uvedeno výše, tento přístup předpokládá vážení podílu, kterým každá integrace přispěla k dosažení cíle podle objemu výzkumné činnosti zaznamenané v informačních systémech s integracemi PID.

Bez přesných a aktuálních informací o současné úrovni institucionální adopce není možné pro ni stanovit realistický cíl nebo časový plán. Proto doporučujeme, aby byl tento cíl znovu posouzen i s přihlédnutím k výsledkům průzkumu plánovaného na rok 2025. Vzhledem k tomu, že plány na integrace mohou trvat několik let a vyžadují dlouhodobější soustředění a závazek, navrhovali bychom také začlenit dohodnuté cíle do budoucích národních strategií a plánovacích dokumentů.

Pokud centrální národní služby dokáží zajistit přístup k výhodám PID velké skupině institucí, pak by cílová úroveň institucionální adopce mohla být nižší. Tuto hodnotu bude nutno určit na základě úrovně výzkumné činnosti v institucích, které implementují PID do svých systémů, a v těch, které využívají centrální služby.

Čisté úspory na základě úrovně adopce								
Úroveň adopce	0 %	20 %	40 %	50 %	55 %	60 %	80 %	90 %
Efektivní přínos (s ohledem na křivku adopce)	0.0 %	4.7%	26.9 %	50.0 %	62.2 %	73.1 %	95.3 %	98.2 %
Roční časové úspory v člověkodnech za celý sektor	0	342	1938	3604	4486	5269	6865	7078
Každoroční úspory nákladů (přínos)	-	1,560,989 Kč	8,852,014 Kč	16,457,141 Kč	20,487,802 Kč	24,062,268 Kč	31,353,293 Kč	32,322,278 Kč
Scénář 1: Bez centrální podpory z programu PID NTK								
Náklady na implementaci za rok za všechny instituce	(23,483,156 Kč)	(23,483,156 Kč)	(23,483,156 Kč)	(23,483,156 Kč)	(23,483,156 Kč)	(23,483,156 Kč)	(23,483,156 Kč)	(23,483,156 Kč)
Čisté úspory	(23,483,156 Kč)	(21,922,167 Kč)	(14,631,142 Kč)	(7,026,015 Kč)	(2,995,354 Kč)	579,112 Kč	7,870,137 Kč	8,839,122 Kč
Scénář 2: S centrální podporou z programu PID NTK								
Náklady na implementaci za rok za všechny instituce	(11,741,578 Kč)	(11,741,578 Kč)	(11,741,578 Kč)	(11,741,578 Kč)	(11,741,578 Kč)	(11,741,578 Kč)	(11,741,578 Kč)	(11,741,578 Kč)
Náklady na program PID NTK	(7,365,987 Kč)	(7,365,987 Kč)	(7,365,987 Kč)	(7,365,987 Kč)	(7,365,987 Kč)	(7,365,987 Kč)	(7,365,987 Kč)	(7,365,987 Kč)
Čisté úspory	(19,107,565 Kč)	(17,546,576 Kč)	(10,255,551 Kč)	(2,650,424 Kč)	1,380,237 Kč	4,954,703 Kč	12,245,728 Kč	13,214,713 Kč

Tabulka 2: Souhrnné čisté náklady a/nebo přínosy všeobecné adopce pěti prioritních PID v českých institucích. S podporou centrální služby PID poskytované NTK se integrace PID dokáže samofinancovat od úrovně 50 % až 55 % adopce.

4.5 Úspory díky činnosti národních konsorcií PID

Vedle úspor nákladů, které přináší PID a podpůrné služby poskytované v rámci projektu CARDS, dochází ke snížení nákladů také díky konsorciím ORCID a DataCite, která začala fungovat pod vedením NTK v rámci programu podpory PID. Tabulka 3 uvádí náklady, které by vznikly 11 členům konsorcia DataCite a 24 členům konsorcia ORCID, kdyby byli individuálními členy.¹² Porovnáme-li tyto náklady s poplatky konsorcií, zjistíme, že při současném počtu členských institucí tato dvě konsorcia ušetří českým institucím 3 050 240 Kč ročně.

Čisté úspory díky členství v konsorciích			
Národní konsorcia	Počet institucí	Náklady na instituci	Náklady na všechny instituce
Náklady na individuální členství			
DataCite	11	(50,000 Kč)	(550,000 Kč)
ORCID	24	(207,000 Kč)	(4,968,000 Kč)
Výdaje celkem			(5,518,000 Kč)
Náklady na členství v konsorciu			
DataCite	N/A	N/A	(50,000 Kč)
ORCID	24	(100,740 Kč)	(2,417,760 Kč)
Náklady celkem			(2,467,760 Kč)
Roční úspory díky oběma konsorciím			3,050,240 Kč

Tabulka 3: Dosavadní úspory díky konsorciím DataCite a ORCID financovaným z projektu CARDS
(úspory se budou zvyšovat s rostoucím počtem členských institucí).

¹² V době psaní tohoto článku (prosinec 2024) bylo 11 členů konsorcia DataCite a 24 členů konsorcia ORCID. Jakmile se ke konsorciím připojí další instituce, roční úspory se budou zvyšovat.

5 Závěry

Z naší analýzy je zřejmé, že investice do podpory adopce PID v českých institucích přinese významnou návratnost. Rozšířením členství v konsorciích, technické podpory, osvěty a vzdělávání v komunitě bude centrální národní podpora PID (koordinovaná NTK v rámci projektu CARDS) a adopce PID/služeb dostupných v Národní datové infrastruktuře zásadní pro maximalizaci těchto výnosů a jejich dosažení v co nejkratším časovém horizontu.

U některých entit (například organizací a osob) máme k dispozici dostačující odhady současné úrovně pokrytí PID. Důrazně doporučujeme provést navazující průzkum, který by zjistil počet a rozsah současných a plánovaných integrací PID v českých institucích, a vyhodnotit pokrytí českých výzkumných entit v rámci registrů PID. Tím se stanoví výchozí úroveň adopce PID, což umožní průběžné hodnocení pokroku.

Potenciální celkové finanční úspory z opakovaného použití metadat a automatizace procesů jsou vyšší než finanční náklady na integrace a podporu PID. Z našich modelů vyplývá, že při 55% adopci PID budou čisté úspory činit 1 380 237 Kč. S rostoucí adopcí se zvyšují i čisté úspory – na 13 214 713 Kč při 90% adopci PID. Tyto roční finanční úspory jsou výsledkem úspory času stráveného administrativními úkony a mohly by snadno přesáhnout 7 207 člověkodní za rok. Z toho vyplývá, že cíl zvýšit adopci PID má zásadní význam pro splnění priorit české politiky v oblasti snižování administrativní zátěže a posilování znalostní ekonomiky, jak je stanoveno ve strategiích RIS3 [27] a NP VaVal 2021+ [4]. S ohledem na výsledky naší analýzy doporučujeme stanovit cíl 80% pokrytí všech pěti prioritních PID posílený o cíl integrovat PID do nejméně 60 % institucionálních informačních systémů v prostředí výzkumu (rozuměno buď v absolutních číslech - počet systémů - nebo v systémech obsahujících 60 % informací o výzkumu), což by stačilo pro zajištění návratnosti investic.

Kromě toho širší pokrytí PID pro granty, výstupy, instituce, osoby a projekty umožní výrazně zlepšit porozumění výzkumné činnosti a jejím dopadům. Schopnost lépe zjišťovat a kvantifikovat výstupy výzkumu povede ke strategickým poznatkům, například umožní identifikovat trendy ve výzkumné spolupráci nebo vznikající odborné oblasti nových inovativních témat, a tím podpoří lepší investice a výraznější sociální a ekonomický dopad výzkumu.

Zvyšování viditelnosti českého výzkumu v globálních registrech PID rovněž přispěje k dosažení klíčových cílů národní politik. K nim patří strategické cíle výzkumu a vývoje, které spočívají ve zvýšení kvality a mezinárodní excelence výzkumu a vývoje v České republice, ve zvýšení otevřenosti a atraktivity ČR pro mezinárodní výzkum a vývoj, jakož i v intenzivnějším zapojení výzkumu a vývoje v ČR do Evropského výzkumného prostoru. Zvyšování dohledatelnosti a dosahu českého výzkumu také maximalizuje jeho dopad a pomůže naplnit cíl RIS3 strategie, kterým je růst kvality a společenské relevance veřejného výzkumu.

Co nejúplnější a nejkomplexnější adopce PID by proto měla být cílem další práce zainteresovaných stran v českém výzkumném ekosystému, včetně všech výzkumných institucí (velkých i malých), stejně jako poskytovatelů finančních prostředků a služeb, které čeští výzkumní pracovníci využívají. Doporučujeme, aby vedle centrální podpory PID byly perzistentní identifikátory implementovány také do národních systémů jako jsou IS VaVal, Národní repozitář a další národní systémy a služby, především ty, které mají být vybudovány v rámci iniciativy EOSC CZ. Tím bude zajištěno, že i menší instituce, instituce s nedostatečnými technickými kapacitami a instituce s malým objemem

výzkumné činnosti budou moci přímo využívat PID, šířeji sdílet informace o své činnosti, zvýšit její efektivitu a více přispívat k růstu české znalostní ekonomiky.

Na podporu tohoto úsilí doporučujeme provést další průzkum, jenž bude identifikovat klíčové systémy, v nichž se shromažďují informace o výzkumu (např. žádosti o granty od poskytovatelů finančních prostředků, platformy pro vykazování atd.), a přednostně se na ně zaměřit při integraci PID. Tím se jednak zvýší objem metadat, která jsou k dispozici pro opakované použití prostřednictvím registrů PID, a jednak se zvýší efektivita národních služeb. V této oblasti již bylo dosaženo značného pokroku díky založení konsorcií DataCite a ORCID a díky plánované integraci ROR.

I při konzervativních odhadech použitých v této analýze je zřejmé, že přínosy zavedení PID v České republice jsou značné. Reálné přínosy pravděpodobně přesáhnou úroveň odhadovanou v této zprávě, ale mají-li být skutečně realizovány, budou vyžadovat cílené strategické investice a systematickou podporu. Kombinovaný přístup, který zahrnuje centrální služby, dlouhodobou technickou podporu, vzdělávání a zapojení komunit, napomůže spravedlivé a účinné adopci PID v České republice. Tento přístup by měl být podložen spolehlivými údaji o míře adopce PID a o současném zatížení neefektivním shromažďováním informací, aby bylo možné přesně zacílit činnosti na místa s největší potřebou nebo potenciálním dopadem a maximalizovat tak dosah a rozsah české sítě PID.

Seznam zdrojů

1. Toma Susi, Shaul Shalvi, Mangala Srinivas. I'll work on it over the weekend': high workload and other pressures faced by early-career researchers. Nature. 17. 6. 2019; <https://doi.org/10.1038/d41586-019-01914-z>.
2. Josh Brown, Phill Jones, Alice Meadows , Fiona Murphy. Incentives to invest in identifiers: A cost-benefit analysis of persistent identifiers in Australian research systems [Internet]. Zenodo; 30. září 2022 [citováno 10. 12. 2024]. Dostupné z: <https://doi.org/10.5281/zenodo.7100578>.
3. Ministerstvo průmyslu a obchodu. Národní výzkumná a inovační strategie pro inteligentní specializaci České republiky 2021–2027: (Národní RIS3 strategie) [Internet]. [citováno 11. 12. 2024] str. 36. Dostupné z: <https://www.ris3.cz/analyzy-a-dokumenty/zakladni-dokumenty>.
4. Rada pro výzkum, vývoj a inovace a Ministerstvo školství, mládeže a tělovýchovy. Národní politika výzkumu, vývoje a inovací České republiky 2021+ [Internet]. [citováno 11. 12. 2024]. Dostupné z: <https://vyzkum.gov.cz/FrontClanek.aspx?idsekce=913172>.
5. Karel Havlíček. Inovační strategie České republiky 2019–2030 [Internet]. ISBN: 978-80-7440-242-5: Rada pro výzkum, vývoj a inovace ; 2019 [citováno 11. 12. 2024]. Dostupné z: <https://vyzkum.gov.cz/FrontClanek.aspx?idsekce=866015>.
6. Evropská komise, Generální ředitelství pro výzkum a inovace. EOSC Executive Board. A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC). [Internet]. 2020 [citováno 10. 12. 2024]. Dostupné z: <https://data.europa.eu/doi/10.2777/926037>.
7. Josh Brown, Phill Jones, Alice Meadows, Fiona Murphy. Revised cost-benefit analysis for the UK PID Support Network [Internet]. Zenodo; listopad 2022 [citováno 25. 1. 2024]. Dostupné z: <https://doi.org/10.5281/zenodo.7356219>.
8. Phill Jones, Alice Meadows, Fiona Murphy, Josh Brown, Aislinn Conway, Lisa Griffith. Efficiency and insight: a cost-benefit analysis for a central service to support persistent identifier implementation in Ireland. 2024 [citováno 10. 5. 2024]; Dostupné z: <https://doi.org/10.7486/DRI.nz80kt123>.
9. Rady pro výzkum, vývoj a inovace. Koncepce Informačního systému výzkumu, experimentálního vývoje a inovací na období let 2021 až 2025 [Internet]. červenec 2020 [citováno 11. 12. 2024] str. 3. Dostupné z: <https://vyzkum.gov.cz/FrontClanek.aspx?idsekce=913128>.
10. cOAlition S. Technical Guidance and Requirements | Plan S [Internet]. [citováno 2. 2. 2024]. Dostupné z: <https://web.archive.org/web/20240202114216/https://www.coalition-s.org/technical-guidance-and-requirements>.
11. Evropská komise. Doporučení komise (EU) 2018/790 o přístupu k vědeckým informacím a jejich uchování [Internet]. 25. 4. 2018 str. 14–15. Dostupné z: <https://eur-lex.europa.eu/legal-content/CS/TXT/PDF/?uri=CELEX:32018H0790>.
12. Úřad vlády České republiky. Národní strategie otevřeného přístupu České republiky k vědeckým informacím na léta 2017-2020 [Internet]. [citováno 11. 12. 2024]. Dostupné z: <https://vyzkum.gov.cz/FrontClanek.aspx?idsekce=876326>.
13. Ministerstvo průmyslu a obchodu. Národní výzkumná a inovační strategie pro inteligentní specializaci České republiky 2021–2027: (Národní RIS3 strategie) [Internet]. [citováno 11. 12. 2024 Dec 11] str. 40. Dostupné z: <https://www.ris3.cz/analyzy-a-dokumenty/zakladni-dokumenty>.
14. Marek Blažka, Libor Kraus. Guide to the system of public support for research, development and innovation in the Czech republic – 2022 [Internet]. 2022 [citováno 10. 10. 2024]. Dostupné z: https://www.comtesfht.cz/media/document/pruvodce_2022-en-a-isbn-kor.pdf.

15. Rada pro výzkum, vývoj a inovace. Usnesení k jednotlivým bodům programu 405. zasedání Rady pro výzkum, vývoj a inovace [Internet]. 25.10. 2024 [citováno 10. 12. 2024]. Dostupné z: <https://vyzkum.gov.cz/FrontClanek.aspx?idsekce=1020795>.
16. Matěj Antol, Jiří Marek, Michaela Capandová, Jaroslav Juráček, Luděk Matyska. "EOSC CZ: Towards the development of Czech national ecosystem for FAIR research data" [Internet]. arXiv; 2024 [citováno 8. 1. 2025]. Dostupné z: <https://doi.org/10.48550/arXiv.2402.13343>.
17. Josh Brown, Phill Jones, Alice Meadows, Fiona Murphy, Paul Clayton. UK PID Consortium: Cost-Benefit Analysis [Internet]. Zenodo; červen 2021 [citováno 10. 12. 2024]. Dostupné z: <https://doi.org/10.5281/zenodo.4772627>.
18. Rob Johnson, Helen Henderson, Hazel Woodward. Institutional ORCID Implementation And Cost Benefit Analysis Report [Internet]. Zenodo; červenec 2015 [citováno 29. 1. 2024]. Dostupné z: <https://doi.org/10.5281/zenodo.1445290>.
19. Mona Hide Klausen. "Even Minor Integrations Can Deliver Great Value – A Case Study." *Procedia Computer Science*. 2017; 106:153–159. <https://doi.org/10.1016/j.procs.2017.03.011>.
20. Josh Brown. Developing a persistent identifier roadmap for open access to UK research [Internet]. Zenodo; červenec 2019 [citováno 10. 12. 2024]. Dostupné z: <https://doi.org/10.5281/zenodo.5609266>.
21. Josh Brown, Phill Jones, Alice Meadows, Fiona Murphy. PID-optimised workflows: A vision of a more efficient future [Internet]. Zenodo; září 2022 [citováno 10. 12. 2024]. Dostupné z: <https://doi.org/10.5281/zenodo.7085489>.
22. Dmitry Kucharavy, Roland De Guio. "Application of S-shaped curves". *Procedia Engineering* 9. 2011: 559–572. <https://doi.org/10.1016/j.proeng.2011.03.142>.
23. Jan Dvořák, Tomáš Chudlarský, Josef Špaček. "ORCID Adoption Monitoring Through a National and an Institutional CRIS". *Procedia Computer Science*. 1. 1. 2024;249:286–295. <https://doi.org/10.1016/j.procs.2024.11.074>.
24. Rebecca Bryant, Anna Clements, Pablo de Castro, Joanne Cantrell, Annette Dortmund, Jan Fransen, Peggy Gallagher, Michele Mennielli. Practices and Patterns in Research Information Management: Findings from a Global Survey. 2018 [citováno 10. 12. 2024]. Dostupné z: <https://doi.org/10.25333/BGFG-D241>.
25. Josh Brown, Alice Meadows. Persistent identifiers adoption and awareness survey report [Internet]. Jisc; 7. říjen 2020 [citováno 10. 12. 2024]. Dostupné z: <https://repository.jisc.ac.uk/id/eprint/8107>.
26. Josh Brown, Aislinn Conway, Michelle Doran, Lisa Griffith, Phill Jones, Alice Meadows, Fiona Murphy. Interoperability, Openness, and Impact: Recommendations and Roadmap for an Irish National PID Strategy [Internet]. Digital Repository of Ireland; 2024 [citováno 11. 12. 2024]. Dostupné z: <https://doi.org/10.7486/DRI.sn00qt29n>.
27. Ministerstvo průmyslu a obchodu. Národní výzkumná a inovační strategie pro inteligentní specializaci České republiky 2021–2027: (Národní RIS3 strategie) [Internet]. [citováno 11. 12. 2024]. Dostupné z: <https://www.ris3.cz/analyzy-a-dokumenty/zakladni-dokumenty>.

Příloha A – Výpočet nákladů a přínosů

Viz samostatný soubor:

Příloha A – Výpočet nákladů a přínosů, <https://doi.org/10.48813/3vry-ag05>.

Příloha B Datové zdroje

7.1 IS VaVal

Data z obou databází (RIV, CEP) poskytl Alexandr Kuchynka z NTK ve formě souborů CSV a XLXS v kódování UTF-8.

7.2 Digital Science Dimensions

Data z databáze Dimensions byla získána na základě licence prostřednictvím jejího rozhraní API. Dokumentaci naleznete zde: <https://www.dimensions.ai/resources/api-documentation>. Rozhraní Dimensions API používá vlastní dotazovací jazyk nazvaný Dimensions Search Language (DSL), konkrétní dotazy použité pro tuto analýzu jsou uvedeny níže:

Publikace s autory z českých institucí, financované jinými než českými subjekty

Rozhraní Dimensions API je omezeno na 50 000 výsledků na dotaz a 1 000 výsledků na stránku. Aby bylo zajištěno dodržení těchto limitů, byla data pro každý rok dotazována zvlášť a výsledky byly stahovány po 1 000 záznamech.

Například první stránka výsledků za rok 2018 byla vyžádána pomocí následujícího dotazu DSL.

```
DSL.search publications
where research_org_countries.id="CZ"
not funder_countries.id ~ "CZ"
and year = 2018
return publications [basics+doi+funders]
limit 1000 skip 0
```

Výzkumné granty udělené českým institucím jinými než českými poskytovateli financí

Počet grantů za období 2018–2022 byl nižší než 50 000, takže nebylo nutné provádět samostatný dotaz pro každý rok. Níže je uveden dotaz DSL použitý pro první stránku výsledků.

```
search grants
where research_org_countries.id="CZ"
and start_year in [2018:2022]
and research_org_types~"Education"
not funder_org_countries.id="CZ"
return grants [basics+funder_orgs+research_orgs+research_org_types]
limit 1000 skip 0
```

Data v IS VaVal a Dimensions mohou být jako u všech databází v průběhu času revidována a zdokonalena a mohou být k dispozici nové údaje. V zájmu reprodukovatelnosti byl skript použitý k analýze dat v Dimensions spuštěn třikrát v průběhu jednoho týdne od 18. do 25. listopadu 2024.

7.3 ROR

Data ROR jsou otevřeně dostupná prostřednictvím rozhraní API. Dokumentaci naleznete zde:

<https://ror.readme.io/>.

Identifikátory ROR byly získány buď na základě identifikátorů GRID, nebo na základě názvů institucí pomocí jednoduchého vyhledávacího dotazu. Je vyžadováno kódování URL. Například ROR ID pro České vysoké učení technické v Praze lze získat z datové struktury JSON vrácené následujícím URL.

```
https://api.ror.org/organizations?query=%22%C4%8Cesk%C3%A9%20vysok%C3%A9%20u%C4%8Den%C3%AD%20technick%C3%A9%20v%20Praze%22
```