



## METODIKA

MOŽNOSTI SLEDOVÁNÍ  
INTERNACIONALIZACE  
VEŘEJNÝCH VYSOKÝCH  
ŠKOL V ČR: DATOVÉ  
ZDROJE, LIMITY A NÁVRH  
ANALYTICKÝCH  
POSTUPŮ

---

Národní vzdělávací fond, o.p.s.  
2025

# Možnosti sledování internacionalizace veřejných vysokých škol v ČR: datové zdroje, limity a návrh analytických postupů

zpracoval: Národní vzdělávací fond, o.p.s.

2025

Projekt sdílených činností  
Strategická inteligence pro výzkum a inovace

## Obsah

1. Úvod .....	4
2. Zdroje dat a jejich využití .....	6
2.1 Výroční zprávy veřejných VŠ – tabulková příloha .....	6
2.2 Výroční zprávy veřejných VŠ – textová příloha .....	10
2.3 Polostrukturované rozhovory a expertní konzultace se zástupci VŠ .....	14
2.4 Řízená diskuse se zástupci VŠ a MŠMT .....	16
2.5 Databáze Scopus .....	16
3. Hodnocení zdrojů dat .....	17
3.1 Dostupnost dat .....	17
3.2 Srovnatelnost .....	18
3.3 Validita a reliabilita .....	19
3.4 Rizika a překážky využití zdrojů dat .....	20
4. Plán pro další období .....	21
5. Seznam zkratk.....	23
6. Použitá literatura .....	24
Příloha 1: Vizualizace sítě slov .....	26

## 1. Úvod

Tento metodický návrh vznikl v rámci projektu „Strategická inteligence pro výzkum a inovace“ (STRATIN+). Projekt STRATIN+ si klade za cíl poskytovat analytickou podporu pro efektivní realizaci politiky výzkumu, vývoje a inovací v ČR. V tomto kontextu představuje mapování úrovně a rozvoje internacionalizace vysokého školství zásadní úkol, neboť internacionalizace je katalyzátor transformace vysokých škol, zvyšování kvality výzkumu a posilování mezinárodní konkurenceschopnosti. Cílem metodického návrhu je zmapovat dostupné zdroje a připravit postupy využitelné ke kontinuálnímu systematickému posuzování úrovně internacionalizace veřejných vysokých škol v ČR.

Pro zachycení kvantitativní stránky internacionalizace VŠ byly využity zejména tabulkové přílohy výročních zpráv vysokých škol. Věnovali jsme se třem aspektům internacionalizace vysokých škol – zahraniční mobilitě akademických pracovníků, působení akademických pracovníků a výzkumníků s cizí státní příslušností na českých VŠ a využívání Joint/Double Degree programů. Pozornost jsme věnovali především mobilitě akademických pracovníků, kteří jsou klíčovým motorem pro transformaci, zvýšení kvality výzkumu a posílení mezinárodní konkurenceschopnosti celé instituce.

Rámcově jsme posuzovali možnosti využití dat z databáze Scopus pro zachycení mezinárodně zakotvených publikačních aktivit zaměstnanců českých vysokých škol. Hlubší analýza dat Scopus je plánovaná na další rok řešení projektu. Pracovali jsme také s textovými přílohami výročních zpráv VŠ, které pravidelně informují zejména o způsobech podpory vysokoškolské zaměstnanecké mobility a postupech začleňování zahraničních členů do akademické obce atd.

K doplnění již existujících dat je navržena v dalších letech řešení projektu realizace kombinovaného šetření založeného na polostrukturovaných kvalitativních rozhovorech a dotazníkovém šetření. Oba sběry dat budou zaměřeny primárně na získání zpětné vazby od představitelů vysokých škol k procesům, potřebám a překážkám internacionalizace. Dalším cílem sběrů dat je doplnění informací pro správné pochopení a interpretaci dat z již existujících zdrojů. Kombinace kvalitativních a kvantitativních metod umožňuje komplexnější pohled na problematiku a ověření dat triangulací.

Pozornost je věnována zachycení situace na veřejných vysokých školách. Výroční zprávy soukromých vysokých škol nejsou k dispozici v dostatečně standardizované podobě, která by umožňovala jejich hlubší analýzu, navíc jádro excelentních vědeckovýzkumných činností je v českém vysokoškolském sektoru soustředěno zejména na veřejných vysokých školách.

Následující tabulka shrnuje přehledně základní informace o prozkoumaných zdrojích dat, jaké poskytují informace, jejich přednosti a omezení (viz Tabulka 1).

**Tabulka 1: Přehled navrhovaných zdrojů dat pro posouzení úrovně internacionalizace VŠ**

Zdroj dat	Co měří / obsah dat	Úroveň detailu	Silné stránky	Limity / omezení
<b>Tabulkové přílohy výročních zpráv VŠ</b>	<ul style="list-style-type: none"> <li>• Mobilita akademických zaměstnanců podle VŠ a zemí</li> <li>• Joint/Double Degree programy</li> <li>• Počty studentů DDP</li> <li>• Zahraniční akademičtí pracovníci na českých VŠ</li> </ul>	<ul style="list-style-type: none"> <li>• Úroveň VŠ</li> <li>• U zahraničních akademiků fakulty</li> </ul>	<ul style="list-style-type: none"> <li>• Standardizovaný sběr dat MŠMT</li> <li>• Dlouhodobá dostupnost</li> <li>• Možnost časových řad</li> </ul>	<ul style="list-style-type: none"> <li>• Kvalita dat u mobility kolísá</li> <li>• Publikováno převážně v PDF</li> <li>• Omezené oborové členění</li> </ul>
<b>Textové přílohy výročních zpráv VŠ</b>	<ul style="list-style-type: none"> <li>• Postupy k podpoře internacionalizace</li> <li>• Podpora mobility studentů a zaměstnanců</li> <li>• Integrace zahraničních členů akademické obce</li> <li>• Další aktivity VŠ pro internacionalizaci</li> </ul>	<ul style="list-style-type: none"> <li>• Úroveň VŠ</li> </ul>	<ul style="list-style-type: none"> <li>• Kompletní a povinně zveřejňované texty</li> <li>• Konkretizace postupů a opatření</li> </ul>	<ul style="list-style-type: none"> <li>• Malá hloubka informací</li> <li>• Různá kvalita textů mezi VŠ</li> <li>• Reporting bias</li> </ul>
<b>Databáze Scopus</b>	<ul style="list-style-type: none"> <li>• Míra mezinárodní publikační spolupráce</li> <li>• Společné publikace se zahraničními institucemi</li> <li>• Možnost zahrnout ukazatel vědeckého dopadu článku (nutno ověřit)</li> </ul>	<ul style="list-style-type: none"> <li>• Úroveň VŠ</li> <li>• Úroveň širších oborů</li> <li>• Potenciálně nižší úroveň (nutné ověření)</li> </ul>	<ul style="list-style-type: none"> <li>• Mezinárodně uznávaná databáze</li> <li>• Ukazatel skutečné vědecké spolupráce</li> <li>• Možnost časových řad</li> </ul>	<ul style="list-style-type: none"> <li>• Změny v indexaci → horší srovnatelnost</li> <li>• Omezený přístup k detailním metadatům</li> </ul>
<b>Expertní rozhovory se zástupci (vybraných) VŠ</b>	<ul style="list-style-type: none"> <li>• Kontext a bariéry internacionalizace</li> <li>• Zkušenosti s mobilitou a DDP</li> <li>• Vysvětlení dat z jiných zdrojů</li> </ul>	<p>Úroveň bude stanovena podle potřeb zadavatele (možná je i větší podrobnost)</p>	<ul style="list-style-type: none"> <li>• Hlubkové kvalitativní informace</li> <li>• Možnost ověřit interpretace</li> <li>• Doplnění statistik</li> </ul>	<ul style="list-style-type: none"> <li>• Náročnost koordinace a zpracování</li> <li>• Riziko nízké ochoty ke spolupráci</li> <li>• Závislost na podpoře MŠMT</li> </ul>
<b>Dotazníkové šetření</b>	<ul style="list-style-type: none"> <li>• Potřeby a bariéry VŠ</li> <li>• Osvědčené postupy</li> <li>• Obecné trendy internacionalizace</li> </ul>	<ul style="list-style-type: none"> <li>• Možné cílit na různé úrovně</li> </ul>	<ul style="list-style-type: none"> <li>• Možnost plošného pokrytí sektoru</li> <li>• Standardizace → srovnatelnost mezi roky</li> <li>• Možnost cílit na různé pracovní role</li> </ul>	<ul style="list-style-type: none"> <li>• Riziko nízké návratnosti</li> <li>• Omezená hloubka</li> <li>• Závislost na podpoře MŠMT</li> </ul>

## 2. Zdroje dat a jejich využití

### 2.1 Výroční zprávy veřejných VŠ – tabulková příloha

Tabulkové přílohy výročních zpráv jsou zdrojem standardizovaných dat o veřejných vysokých školách. Jejich struktura je sjednocená metodickými pokyny MŠMT, mezi jednotlivými roky se může lišit. Navzdory možným změnám se jedná o nejspolehlivější, nejcelistvější a v čase nejstabilnější zdroj dat, který je k tématu internacionalizace VŠ k dispozici.

Zaměřili jsme se na postižení úrovně zahraniční mobility vysoce kvalifikovaných vysokoškolských pracovníků VŠ (akademiků), sledování počtů akademických, vědeckých a odborných zaměstnanců s cizím státním občanstvím a na počty Joint/Double Degree programů a jejich studentů.

#### *2.1.1 Obousměrná mobilita akademických pracovníků v geografickém členění*

Tabulka 7.2 tabulkové přílohy výročních zpráv o činnosti VŠ zachycuje počty výjezdů a příjezdů akademických pracovníků členěné podle zemí. Byly sledovány celkové podíly vyjíždějících a přijíždějících akademiků vztažené k počtu zaměstnanců na jednotlivých VŠ, a dále geografické rozložení výjezdů a příjezdů v letech 2019 a 2023. Data z roku 2016 byla využita jen doplňkově, protože nebyla zcela logicky konzistentní (zřejmě v důsledku přechodu na novou metodiku vykazování).

Data umožňují určit podíl akademických pracovníků, kteří realizují výjezd do zahraničí, a podíl těch, kteří na danou vysokou školu přicházejí ze zahraničí, vždy ve vztahu k celkovému počtu zaměstnanců. Tento indikátor vypovídá o úrovni mezinárodní mobility dané instituce. Současně umožňuje zkoumat geografické směřování mobility. Údaje jsou dostupné na úrovni jednotlivých VŠ, což umožňuje sledovat vývoj geografického rozložení i intenzity mezinárodních kontaktů v čase.

V tabulce 7.2 jsou zachyceny výjezdy a příjezdy od 5 dnů trvání, tj. včetně relativně krátkých pobytů. Není zde možné zjistit ani délku a účel pobytu. Podle osloveného experta je však zahraniční mobilita zaměstnanců zpravidla krátkodobá. Ke spolehlivému zjištění těchto charakteristik by byl nutný vlastní sběr dat, nebo úprava požadavků MŠMT na statistiky zahraniční mobility.

Situaci v oborech je možné sledovat jen na úrovni specializací VŠ, tj. lze hledat charakteristické vzorce pro skupinu technických či uměleckých VŠ, nelze však oddělit humanitní, sociálněvědní či medicínské obory, protože toto oborové zaměření je patrné zpravidla až při pohledu na fakulty. Takto detailní data však nejsou plošně dostupná, což neumožňuje detailní posouzení mobility podle oborů. Rovněž není možné sledovat prostřednictvím těchto statistik úroveň mobilitní aktivity na podrobnější úrovni, např. na úrovni jednotlivých fakult.

Pro účely analýzy byly vypočítány:

- absolutní počty vyjíždějících a přijíždějících pracovníků
- podíly mobilních pracovníků vůči celkovému počtu akademických a vědeckých pracovníků na dané VŠ
- podíly výjezdů a příjezdů do/ze zemí nebo kontinentů ku celkovým počtům mobilit
- geografické rozložení příjezdů a výjezdů
- index mobility (údaj ukazuje, zda a v jaké míře převažují příjezdy nad výjezdy a naopak)

#### Hlavní zjištění:

Při porovnání výjezdů s příjezdy se ukazuje značná převaha výjezdů českých akademiků do zahraničí nad příjezdy zahraničních akademiků na české veřejné VŠ. Průměrný podíl vyjíždějících na akademických a výzkumných zaměstnancích VŠ dosahoval 35 %, zatímco u přijíždějících šlo pouze o 14 %. Příjezdová mobilita je značně omezená a zřejmě závislá na externích faktorech jako možnosti zahraničních institucí, atraktivita ČR, financování apod.

Vysoké školy se v míře internacionalizace výrazně liší. Významnými faktory jsou velikost (větší VŠ vykazují vyšší míru mobility), oborové zaměření (vyšší podíly u univerzálně zaměřených VŠ, výrazně nízká u umělecky zaměřených VŠ) a regionální vs. celostátní dosah VŠ (regionální školy mobilitu využívají méně).

Z hlediska geografického rozložení mobilit můžeme konstatovat, že výjezdová i příjezdová mobilita je z velké většiny zaměřena na evropské partnerské instituce. Více než 75 % všech výjezdů směřovalo do Evropy, zejména do Německa, Slovenska, Itálie, Španělska a Francie. Z mimoevropských oblastí byly častěji navštěvovány Spojené státy, Japonsko, Kanada a Čína. Příjezdová mobilita má podobné geografické rozložení, dominantní jsou opět evropské země (Polsko, Slovensko, Německo) a USA.

Některé malé regionální školy (např. SU, OU) mají relativně vysoký podíl příjezdů, který zřejmě souvisí s přeshraniční spoluprací (zejména s Polskem a Slovenskem).

Z faktorů pozitivně ovlivňujících mobilitu považujeme za vlivnou zejména geografickou blízkost, dostupnost financování (např. Erasmus+) a zavedené institucionální vazby. V případě příjezdů může být překážkou omezený zájem zahraničních akademiků, administrativní náročnost, nedostatek financí a slabší mezinárodní značka českých škol. Ověřené, a případně kvantifikované informace by však bylo možné získat prostřednictvím primárního sběru dat. Hlubší pohled na faktory ovlivňující oboustrannou mobilitu akademických pracovníků by bylo možné získat prostřednictvím rozhovorů či fokusních skupin.

Popis a základní vyhodnocení dat z výročních zpráv veřejných VŠ o zahraniční mobilitě akademických pracovníků - viz samostatné dokumenty:

*„Zahraniční mobilita akademických pracovníků – příjezdy a výjezdy podle VŠ“*

*„Zahraniční mobilita akademických pracovníků na veřejných VŠ – geografické rozložení příjezdů a výjezdů“*

### *2.1.2 Dynamika programů Joint/Double Degree: zapojení VŠ a studentská účast*

Tabulka 2.3 *Joint/Double/Multiple Degree studijní programy realizované se zahraniční VŠ* ve výročních zprávách veřejných VŠ poskytuje systematický a dlouhodobě dostupný zdroj informací o Double/Joint Degree programech. Lze z ní získat údaje o typu programu (bakalářský, magisterský, doktorský), zemi a instituci partnerské vysoké školy, počtu uzavřených smluv o realizaci programu a počtu českých studentů zapojených do Double/Joint Degree studia v daném roce.

Na základě těchto údajů je možné sledovat vývoj nabídky programů, míru jejich skutečného využívání studenty, zeměpisnou strukturu partnerských institucí a rozvoje této formy internacionalizace napříč lety. Tabulka zároveň umožňuje odlišit aktivní programy (s účastí studentů) od programů pasivních, které nebyly v daném roce využity. Vzhledem k tomu, že údaje jsou dostupné odděleně pro bakalářské, magisterské a doktorské studium, je tedy možné sledovat, jak se liší množství programů a míra jejich využívání podle úrovně studia.

Konkrétně umožňují data vyhodnotit:

- vývoj počtu DDP jednotlivých úrovní (bakalářské, magisterské, doktorské)
- rozdělení DDP na aktivní a neaktivní programy
- počet studentů DDP (bakalářské, magisterské, doktorské úrovně)
- podíl studentů DDP na celkovém počtu studentů jednotlivých úrovní
- země, ve kterých jsou lokalizované partnerské VŠ

Data byla zpracována formou deskriptivní analýzy za roky 2016, 2019 a 2023. Pro každý rok byl stanoven celkový počet DDP podle úrovně studia, počet aktivních a pasivních programů a počet studentů zapojených do studia v těchto programech. Z údajů o počtu studentů jsme dále vypočítali relativní ukazatele — zejména podíl studentů DDP na celkovém počtu studentů příslušné úrovně studia za pomoci dat z veřejně dostupné statistiky MŠMT.

Pro umožnění meziškolních srovnání jsme data agregovali podle velikosti školy a jejího oborového zaměření. Výsledky byly prezentovány pomocí indikátorů, grafů a tabulek, které zachycují trendy a rozdíly mezi jednotlivými VŠ.

Analýza přinesla přehled o rozvoji Double/Joint Degree programů na veřejných vysokých školách v ČR. Identifikovali jsme růst nabídky programů ve všech stupních studia, nejvýrazněji u doktorských a bakalářských DDP, a zároveň jsme popsali rozdíl mezi formálním počtem programů a jejich skutečným využíváním. Ukázali jsme, že přibližně čtvrtina až třetina programů zůstává každoročně neaktivní a že míra účasti studentů je velmi nízká — u magisterského studia byla v roce 2003 2,5 % studentů, u bakalářských a doktorských programů méně než 1 %.

Z analýzy dále vyplynulo, že nejčastějšími partnerskými zeměmi jsou Francie a Německo, že neaktivnější v zapojování studentů jsou středně velké VŠ, a že oborově univerzální školy (např. JU, TUL, UK, MUNI) dosahují vyšší míry využívání DDP než školy úzce specializované, s výjimkou VŠE. Umožnili jsme také identifikovat školy, které DDP rozvíjejí (např. VŠE, JU, TUL), i školy, které je v praxi nevyužívají.

Ze statistik není možné zjistit, které faktory ovlivňují rozdíly mezi jednotlivými VŠ a jejich skupinami, například z jakého důvodu je relativně nejvíce těchto programů ve skupině středně velkých VŠ, nebo proč je tento typ programu nejrozšířenější v magisterském stupni studia. Bez dalšího zjišťování informací a dotazování není možné s jistotou tyto faktory určit. Plánované šetření umožní hlubší pochopení informací získaných ze statistik.

Z dat nelze sledovat úspěšnost tohoto typu studií, nelze z nich ani získat informace o účasti studentů Double/Joint Degree programů z partnerských institucí. Není tedy možné posoudit, nakolik se jedná o reciproční výměnu studentů.

Z hlediska kvality dat lze předpokládat, že statistiky jsou přesné a spolehlivé, protože pracují s údaji vedenými v centrálně vedených informačních systémech VŠ a MŠMT.

Joint/Double Degree programy jsou jedním z důležitých aspektů internacionalizace. Ke zvýšení úrovně internacionalizace přispívají navázáním kontaktů na úrovni organizačních jednotek VŠ, zaměstnanců i studentů.

Podrobná analýza viz samostatný dokument: „Joint/Double degree studijní programy na veřejných vysokých školách“

### *2.1.3 Přehled počtů akademických a vědeckých zaměstnanců VŠ s cizím státním občanstvím*

Tabulka 6.5 výročních zpráv VŠ zachycuje přepočtené počty akademických a vědeckých zaměstnanců VŠ, kteří mají cizí státní občanství a propojují tak VŠ se zahraničním prostřednictvím svých kontaktů, zkušeností a znalostí. Země původu jsou zachyceny jen částečně, tabulka zachycuje počty zaměstnanců z Německa, Polska, Rakouska a Slovenska, se kterými české VŠ spolupracují nejintenzivněji. Dále jsou již uváděny jen počty zaměstnanců z ostatních států EU a z ostatních států mimo EU. Údaje z tohoto datového zdroje neumožní podrobnější klasifikaci zemí původu zaměstnanců.

Pracovali jsme s daty z let 2019 a 2023, zprávy z roku 2016 mají jinou, méně podrobnou strukturu a vykazují fyzické počty osob, které nelze srovnávat s přepočtenými počty zaměstnanců. Srovnatelná struktura dat vhodná pro založení časové řady je k dispozici až od roku 2018.

Vysoké školy uvádí údaje na úrovni fakult, což otevírá možnosti zachycení situace v některých oborových skupinách, které často (ne však vždy) odpovídají fakultám.

Nevýhodou tohoto zdroje dat je nemožnost získat údaje o skutečném počtu fyzických osob, které na VŠ působí, které by lépe vypovídaly o jejich přínosu k internacionalizaci vysokoškolského prostředí. Velmi podrobné členění akademických a vědeckých pracovníků umožňuje sledovat, jaké akademické hodnosti a pozice cizinci zastávají.

## 2.2 Výroční zprávy veřejných VŠ – textová příloha

### 2.2.1 Vymezení analytického korpusu a předmětu sledování

Textové přílohy výročních zpráv veřejných VŠ jsou standardizované prostřednictvím otázek MŠMT, na které musí VŠ v textové příloze odpovědět. Podrobnější struktura a obsah odpovědi je ponechána na uvážení VŠ. Internacionalizaci je věnována 7. kapitola. V letech 2019–2024 obsahovala otázky týkající se podpory mobility studentů i zaměstnanců, mechanismů zajišťování integrace zahraničních členů akademické obce a dalších významných aktivit posilujících internacionalizaci<sup>1</sup>. Formulace otázek pro další roky aktuálně není známa a může se změnit.

Textové přílohy výročních zpráv uvádí pro období 2019–2024 popisy aktivit vysokých škol v rámci následujících okruhů:

- a) Strategie VŠ k podpoře účasti studentů v mobilitních programech,
- b) Strategie VŠ k podpoře zahraniční mobility zaměstnanců,
- c) Procesy, programy a nástroje integrace zahraničních členů akademické obce,
- d) Klíčové aktivity VŠ posilující internacionalizaci jejich činností, vč. zapojení do mezinárodních konsorcií a programů.

V tomto období navrhujeme analyticky zkoumat následující témata:

1. Strategie podpory zahraniční mobility zaměstnanců
  - Tato doména představuje klíčový faktor pro rozvoj internacionalizace, neboť umožňuje mapování mechanismů vedoucích k dalšímu rozvoji internacionalizace prostřednictvím mezinárodní zkušenosti akademiků, výzkumníků a pedagogů, kteří se podílejí na chodu a operativním řízení VŠ.
2. Procesy integrace zahraničních členů akademické obce

---

<sup>1</sup> Do roku 2018 byly otázky pouze 2<sup>1</sup> a týkaly se: 1) podpory účasti studentů na zahraničních mobilitních programech a 2) způsobů zajištění integrace zahraničních členů akademické obce. Od roku 2022 přibyla otázka na zkušenosti s implementací kombinovaných a virtuálních mobilit.

- Tato oblast poskytne vzhled do stavu a vývoje podpůrných mechanismů pro zahraniční akademiky a studenty na českých VŠ
3. Další významné aktivity posilující internacionalizaci VŠ
- Tento okruh má potenciál zachytit další aktivity, programy či nástroje přispívající k internacionalizaci.

### *2.2.2 Pilotní evaluace metod zpracování přirozeného jazyka (NLP)*

Vzhledem k časové náročnosti analytického zpracování většího objemu textů bylo věnováno relativně mnoho času prozkoumání metod a nástrojů, které by bylo možné pro tento úkol využít. K analýze lze využít buď tradiční kvalitativní metodologii, která spočívá v manuálním, případně počítačově podpořeném kódování a vyhledání témat v textu zkušeným výzkumníkem, nebo lze sáhnout k pokročilým metodám zpracování přirozeného jazyka (Natural Language Processing – NLP).

Z důvodu efektivity analytických prací byl nejprve ověřován potenciál a výsledky metod zpracování přirozeného textu (NLP) - tematického modelování<sup>2</sup>, které již delší dobu získávají na významu i v sociálních vědách (např. pro metaanalýzu textů, viz M. Baranowski<sup>3</sup> nebo N.C. Lindstedt<sup>4</sup>).

Pilotně jsme ověřovali tematické modelování textů (Topic Modelling) za pomoci vlastního software pro zpracování textů s využitím algoritmů NMF (Non-negative Matrix Factorization), LDA (Latent Dirichlet Allocation) a BERTopic.

Výsledky byly obtížně interpretovatelné, témata a s nimi propojená klíčová slova neposkytla dostatečně diferencovaný pohled na tematiku internacionalizace. Hlavní příčinou jsou obsahově a tematicky homogenní vstupní dokumenty. K podobnému závěru jsme došli i při zpracování textů metodami NMF, LDA a BERTopic za pomoci prostředí jazyka R.

Dále byl testován modul IBM SPSS Statistics pro analýzu textu (Acrea Text Analytics), konkrétně vyhledávání klíčových slov. I v tomto případě však nástroj extrahoval příliš obecná klíčová slova, která nedokázala dostatečně detailně zachytit programy, metody a zkušenosti VŠ. Aktuální verze tohoto programu proto pro účely monitoringu není vhodná.

Grafické znázornění vztahů mezi slovy definujícími témata pomocí vizualizace sítě slov) byla poslední testovaná metoda z oblasti NLP, ani zde nebyly výsledky dostačující. (bližší postup a výsledky viz Příloha 1: Vizualizace sítě slov.)

---

<sup>2</sup> Přehled těchto metod lze nalézt např. ve článku I. Vayanskeho a S.A.P. Kumara „A review of topic modeling methods“ z roku 2022.

<sup>3</sup> M. Baranowski: Epistemological aspect of topic modelling in the social sciences: Latent Dirichlet Allocation, 2022.

<sup>4</sup> N.C. Lindstedt: Structural Topic Modeling For Social Scientists: A Brief Case Study with Social Movement Studies Literature, 2005–2017; 2019.

### 2.2.3 Vyhodnocení metod analýzy textových příloh a metodologický návrh pro budoucí postup

Pilotní evaluace metod zpracování přirozeného jazyka (NLP), zejm. tematického modelování, ukázala jejich neadekvátnost pro tento analytický úkol. Vývoj těchto metod a možnosti jejich aplikace však budeme nadále sledovat. Pro aktuální fázi však navrhujeme kvalitativní analýzu dokumentů, která zajistí vysokou míru validity, spolehlivosti a přesnosti výsledků. Navrhovaný postup zahrnuje:

1. Induktivní kódování/tematická analýza: systematické čtení a kódování textů zkušenými analytiky s cílem indukovat specifická témata, strategie a iniciativy. Při práci několika analytiků na omezeném množství textů bude vytvořen seznam kódů a témat, založený na diskusi vedoucí ke shodě.
2. Explicitní kódovací schéma: Vytvoření precizního kódovacího schématu diskursivní metodou umožní hlubší porozumění textu, než které nabízí extrakce klíčových slov. Kódovací schéma (codebook) bude upravováno i v průběhu další analýzy textů, a to na základě diskuse a shody všech zúčastněných analytiků.
3. Křížové ověření (Inter-coder Reliability): prostřednictvím kódování vzorku textů více analytiky bude zajištěna reliabilita kódování
4. Kvalifikace, znalosti a zkušenosti výzkumníků se zkoumaným tématem jsou dalším aspektem posilujícím realibilitu a validitu analýzy a jejích výsledků.

Analýza bude zaměřena na tematické kódování s primárním cílem identifikace a kategorizace zejména konkrétních programů, opatření a nástrojů na podporu zahraniční mobility a internacionalizace.

Tento úkol vyžaduje zvolit přístup umožňující do určité míry srovnávat situaci mezi jednotlivými lety, školami a jejich skupinami. Zpracování textů tedy bude z hlediska metodologie kvalitativního výzkumu blíže přístupům řazeným do skupiny „small q“<sup>5</sup>, tedy mezi přístupy označované jako (post)pozitivistické. Při vytváření základního kódovacího schématu (codebook) se budeme opírat o znalosti získané v předešlých projektech a v průběhu sekundární analýzy a konzultací s experty.

Jako podpůrný nástroj pro kódování a stanovování témat lze využít některý z dostupných LLM a opřít se o předešlé zkušenosti etablovaných výzkumníků. Využití LLM pro podporu

---

<sup>5</sup> Rozlišení přístupu zakotveného v rámci teoreticky podloženého kvalitativního interpretačního paradigmatu (např. Grounded Theory) zaměřeného na induktivní přístup, porozumění, významy, kontext a reflexivitu, „big Q“, od přístupů využívající kvalitativní data spíše za soubor technik použitelných v rámci převážně kvantitativního a deduktivního rámce, „small q“, zavedli nejprve v 80. letech Kidder a Fine (viz jejich článek „Qualitative and Quantitative Methods: When Stories Converge“ z roku 1987).

tematického kódování sociální vědci již testují (např. Stefan De Paoli<sup>6</sup>). Alternativou je i využití aplikace na kódování za pomoci LLM (např. QualiGPT<sup>7</sup> nebo TALLMesh<sup>8</sup>, testoval např. tým z Penn State University<sup>9</sup>). Supervize lidského výzkumníka je však stále nutná nejen pro dozor, ale i pro prohloubení interpretace.

Výstupy budou mít podobu textové zprávy obsahující popis:

- a) Strategie podpory zahraniční mobility zaměstnanců v daném roce a porovnání s předcházejícími lety (které typy strategií jsou využívány, které ustupují, které jsou využívány nově)
- b) Procesy integrace zahraničních členů akademické obce (které procesy, programy a postupy k podpoře zahraničních členů akademické obce jsou využívány, které ustupují, které jsou využívány nově)
- c) Další významné aktivity posilující internacionalizaci VŠ (přehled typů aktivit k podpoře internacionalizace VŠ a případně jejich hodnocení, aktivity nové a ustupující)
- d) V závislosti na konkrétních výstupech bude text zahrnovat i případné další získané kontextové a jiné informace, které výzkumníci vyhodnotí jako relevantní pro řízení rozvoje internacionalizace VŠ.

#### *2.2.4 Vyhodnocení datového zdroje a jeho využitelnosti*

Klíčovým metodologickým omezením je nízká analytická hodnota dat. Texty často uvádějí mnoho aktivit, smluv a opatření, ale postrádají hlubší kontextuální bohatost a vysvětlení procesů či dopadů potřebné pro hlubší kvalitativní analýzu. Skutečnost, že MŠMT, pro které jsou výroční zprávy zpracovávány a kterému jsou předávány, je zároveň dohledovým orgánem pro VŠ, může způsobovat zkreslení textů zpráv (reporting bias): texty zřídka obsahují negativní hodnocení, převažuje neutrální sentiment. To může snižovat validitu závěrů.

Analyticky hodnotné jsou konkrétní uváděné přínosné programy, postupy a partnerství, které zde lze nalézt, zejména v oblasti mobility zahraničních pracovníků, která jinde není dostatečně zpracovávána. Přes svá omezení mají zprávy vysokou komplementární hodnotu. Poskytují deskriptivní přehled konkrétních programů a partnerství.

---

<sup>6</sup> Viz článek „Further Explorations on the Use of Large Language Models for Thematic Analysis. Open-Ended Prompts, Better Terminologies and Thematic Maps“. In *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 25, č. 3, 2024. Dostupné na: <https://doi.org/10.17169/fqs-25.3.4196>

<sup>7</sup> Viz [GitHub - KindOPSTAR/QualiGPT: QualiGPT: An easy-to-use tool for qualitative research](#) .

<sup>8</sup> O aplikaci TALLMesh a možnostech jejího využití viz preprint článku autorů De Paoli a Fawzi „TALLMesh: a simple application for performing Thematic Analysis with Large Language Models“.

<sup>9</sup> Preprinty článků z let 2023 a 2024, které shrnují zkušenosti s využitím této aplikace prostřednictvím API přístupu k LLM GPT, jsou dostupné pod názvy „QualiGPT: GPT as an easy-to-use tool for qualitative coding“ a „When Qualitative Research Meets Large Language Model: Exploring the Potential of QualiGPT as a Tool for Qualitative Coding“.

Reliabilitu v kvalitativním výzkumu definuje Silverman<sup>10</sup> jako „stupeň konzistence, s jakým různí pozorovatelé nebo tentýž pozorovatel v různých situacích případy zařadí do též kategorie“<sup>11</sup>. Zajištění maximální možné reliability lze docílit v první fázi při tvorbě kategoriálního systému (codebook) diskusemi nad nalezenými kategoriemi v rámci zpracovatelského týmu a posouzením výsledného textu expertem.

Validita bude zajišťována expertními zkušenostmi zpracovatelského týmu a konzultacemi s externími experty.

### 2.3 Polostrukturované rozhovory a expertní konzultace se zástupci VŠ

Polostrukturované rozhovory a expertní konzultace se zástupci VŠ jsou integrální součástí metodiky směřující k přesnějšímu zachycení a hlubšímu pochopení problematiky internacionalizace VŠ. Cílem je lépe pochopit procesy spojené s internacionalizací na VŠ, zejména s ohledem na jejich potřeby, možnosti a bariéry. Umožní to také získávat zpětnou vazbu i k dalším prováděným analýzám.

Záměrem je realizace polostrukturovaných rozhovorů, kterými lze dobře zachytit situaci na jednotlivých VŠ, a které poskytnou MŠMT souhrnný strukturovaný pohled na možnosti další podpory a rozvoje internacionalizace VŠ. Rozhovory budou zaměřeny na:

- a) otázky potřeb, bariér a možností rozvoje internacionalizace z pohledu vysoké školy (Takto zaměřené dotazy poskytnou, zejména v případě opakovaného šetření, komunikační kanál pro vyhledání možností optimalizace a odstranění bariér internacionalizace VŠ.)
- b) získání doplňkových a kontextových informací pro hlubší interpretaci dat z ostatních datových zdrojů. (Analýza dat z tabulkových příloh výročních zpráv VŠ ukazuje, že pro jejich správnou interpretaci je třeba získat doplňkové informace z terénu a hlouběji pochopit celý kontext, realitu, kterou data odráží, i proces jejich sběru.)

V průběhu kalendářního roku je možno realizovat 6-12 rozhovorů se zástupci vysokých škol, a to tak, aby bylo pokryto co nejširší spektrum typů vysokých škol z hlediska oborového zaměření (specializace versus univerzální zaměření; různé oborové skupiny), velikosti a regionalit (VŠ spíše regionálně zaměřené versus VŠ s celostátní působností). V případě zájmu a aktivní podpory plošného získávání zpětné vazby ze strany státních orgánů je možno rozhovory realizovat rozhovory se zástupci všech VŠ při širším záběru otázek na možnosti, bariéry a potřeby pro internacionalizaci VŠ.

---

<sup>10</sup> David Silverman takto přistupuje k reliabilitě v kvalitativním výzkumu ve své učebnici kvalitativního výzkumu „Ako robiť kvalitatívny výskum“. Odkazuje se přitom na text M. Hammersleyho „Reading Ethnographic Research: a Critical Guide“ z roku 1990.

<sup>11</sup> Hammersley, 1990 s. 67

Struktura témat s nabídkou zpřesňujících otázek bude zachycena ve scénáři polostrukturovaného rozhovoru, který bude uspořádaný podle tematických okruhů tak, aby sloužil zároveň jako podklad pro konzistentní selektivní záznam odpovídající výzkumným otázkám a tématům.

Nárys výzkumných otázek:

- a) Jaké bariéry omezují či přímo ohrožují internacionalizaci Vaší VŠ?
- b) Co Vaše VŠ potřebuje pro úspěšný rozvoj internacionalizace?
- c) Jaký způsob podpory internacionalizace se na Vaší VŠ osvědčil?
- d) Které možnosti a cesty k dalšímu rozvoji internacionalizace Vaší VŠ zůstávají nevyužity a proč?

Nárys tematických okruhů:

- a) mobilita vysoce kvalifikovaných zaměstnanců,
- b) mobilita studentů,
- c) Joint/Double Degree programy,
- d) studijní programy v AJ,
- e) meziuniverzitní a mezifakultní smlouvy a spolupráce,
- f) mezinárodní projekty,
- g) financování mobility a mezinárodních projektů,
- h) jiná témata, která jsou podstatná pro danou VŠ

Rozhovory budou vedeny přednostně osobně, v případě přání respondenta online. S online rozhovory (Teams) existují dobré zkušenosti, neboť se již široce etablovaly a v rámci šetření zvyšují ochotu expertů podílet se na výzkumu.

Expertní konzultace je koncipována jako flexibilní nástroj pro vyjasnění dílčích otázek a potíží, a upřesnění interpretace získaných dat. V srpnu 2025 byla realizována konzultace se zaměstnancem rektorátu jedné z pražských VŠ, díky níž byly získány důležité informace o způsobu evidence dat zejména o zahraniční mobilitě zaměstnanců. Z konzultace mj. vyplynulo, že další sběr dat by mohl být na VŠ vnímán jako zátěž a realizace rozhovorů či dotazníkových šetření vyžaduje jednoznačnou podporu a komunikaci ze strany MŠMT.

Rizika:

- Potřeba diskutovat příliš mnoho témat povede k nutnosti hledat na každé VŠ několik osob znalých tematiky a ochotných se šetření zúčastnit.
- Tým může být donucen redukovat množství témat nebo rozdělit realizaci šetření do více fází a období.
- Neochota VŠ spolupracovat by ohrozila buď validitu a reliabilitu informací, nebo realizaci šetření vůbec.

## 2.4 Řízená diskuse se zástupci VŠ a MŠMT

Řízená diskuse nad výsledky zpracování dat z výročních zpráv, dalších zdrojů a z expertních rozhovorů se jeví jako efektivní forma ověření a získání doplňkových informací a zároveň příležitost k hlubší diskusi se zástupci vysokých škol a MŠMT. Její realizace je plánována v roce 2026. Tímto způsobem je možno získat informace o bariérách, potřebách a způsobech internacionalizace na českých VŠ, a zároveň bude zajištěno relativně přímé předání zkušeností, potřeb, bariér a příležitostí, které vysoké školy vnímají v souvislosti se snahou o rozvoj internacionalizace na jejich pracovištích. Součástí diskuse bude i mapování postupů, které se VŠ osvědčily k podpoře rozvoje mezinárodních vztahů (např. Erasmus pobyty, které jsou od zaměstnanců velmi poptávané).

Diskuse bude řízena výzkumníkem s dlouholetou zkušeností s výzkumnými metodami sběru dat a znalostí problematiky vysokého školství. Podkladem bude scénář řízené diskuse založený na výzkumných otázkách.

Z řízeného workshopu bude zpracován tematicky členěný záznam, výstupem bude zpráva s popisem a interpretací klíčových získaných poznatků. Podstatným výsledkem bude také navázaná komunikace mezi aktéry z vysokých škol a státní správy.

---

## 2.5 Databáze Scopus

Spoluautorství na vědeckých člancích představuje jeden z důležitých indikátorů výzkumné spolupráce. Odráží nejen formální partnerství, ale především reálné zapojení týmů, sdílení expertízy a společné vytváření vědeckého poznání. Na základě provedených rešerší byla pro tuto analýzu vybrána databáze Scopus. Cílem je měřit a vyhodnotit rozsah publikační spolupráce se zahraničními vědci.

Scopus je celosvětově uznávaná a doplňovaná databáze, která obsahuje vysoký podíl světových vědeckých článků. Ve srovnání s WOS nabízí podstatně lepší pokrytí vědecké publikační produkce<sup>12</sup>, a je tedy vhodnějším zdrojem pro zachycení míry spolupráce zaměstnanců jednotlivých českých VŠ na přípravě vědeckých článků se spoluautory ze zahraničních institucí. Tento údaj zachytí jeden z aspektů internacionalizace VŠ, propojení a spolupráci na úrovni společného zpracování vědeckých projektů a jejich výstupů – článků. Za tímto typem aktivity je prokazatelná spolupráce se zahraničními institucemi, a to relativně intenzivní, protože vytvoření vědeckého článku je dlouhodobý proces.

---

<sup>12</sup> Viz článek P. Mongeona a A. Paul-Huse z roku 2016 (The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>)

Pro realizaci analýzy lze zvolit přístup k databázi Scopus prostřednictvím rozhraní Scopus API a jeho implementace v prostředí R pomocí balíčku rscopus. Tento přístup umožní automatizovaně stahovat potřebná metadata o publikacích a jejich afiliacích. Získaná data je možno následně zpracovat a vyhodnotit v softwaru R, který nabízí široké možnosti pro datovou analýzu, vizualizaci i tvorbu reprodukovatelných výstupů.

Z dostupných údajů v databázi Scopus lze využít informace o počtu publikací autorů z českých vysokých škol, které vznikly ve spolupráci s autory ze zahraničních institucí. Tento počet nám poslouží jako indikátor mezinárodní spolupráce.

Údaje lze s jistotou zachytit na úrovni jednotlivých VŠ, získání dat na úrovni nižších organizačních jednotek (fakulty, ústavy, pracoviště) bude ověřováno v příštím roce. Zároveň se pokusíme popsat i stav spolupráce v rámci jednotlivých vědních oborů či alespoň širších vědních domén, u nichž by měla být taková identifikace na základě dostupných údajů rovněž proveditelná. Bude ověřena také možnost zapojení dalších dostupných scientometrických ukazatelů, které umožňují posoudit dopad publikovaných článků. Tyto metriky mohou pomoci lépe porovnat výsledky mezi jednotlivými institucemi a doplnit tak informace o rozsahu jejich mezinárodní spolupráce.

K vyhodnocení možností, které nabízí data dostupná o článcích a jejich autorech jsme zatím použili data z roku 2023. Jejich další zpracování a plánované opakování výpočtů umožní založit časovou řadu.

Pro doplnění analýzy mezinárodní spolupráce je záměrem využít také síťovou analýzu, která umožní vizualizovat a kvantifikovat vazby mezi českými vysokými školami a jejich zahraničními partnery. V této síti budou uzly představovat instituce nebo země a hrany spoluautorství na vědeckých článcích, přičemž jejich váha odráží počet společných publikací. Výsledná analýza by umožnila identifikovat klíčové partnerské školy, intenzitu spolupráce i strukturu mezinárodních výzkumných sítí.

Výstupem bude za každou VŠ souhrnný ukazatel, který vyjadřuje počet článků publikovaných v daném roce a vzniklých ve spolupráci autorů s afiliací k této instituci a spoluautorů působících v zahraničních organizacích. Tento ukazatel tak poskytne přehled o míře mezinárodní publikační spolupráce jednotlivých škol.

## 3. Hodnocení zdrojů dat

### 3.1 Dostupnost dat

Tabulkové přílohy VZ: všechny výroční zprávy veřejných VŠ jsou dostupné online na stránkách vysokých škol, zpravidla ve formátu pdf. Tabulková a textová příloha je strukturovaná podle standardizovaného požadavku MŠMT, a je tedy dobře použitelná jako zdroj dat pro analýzy. Problém je ztížená dostupnost dat z tabulkových příloh ve formátu

pdf, ze kterého se data získávají problematicky. Budoucí analýzy by zefektivnilo poskytnutí tabulek ve vhodnějším formátu (např. jako excelové tabulky).

Textové přílohy VZ: jejich dostupnost těchto zpráv je dobrá, jsou součástí školami povinně zveřejňovaných výročních zpráv VŠ.

Scopus: obsahuje značný podíl světových vědeckých článků (i ve srovnání s WOS). Pro účely námi plánovaných analýz je dostupnost dat z této databáze dobrá, možný je i API přístup.

Řízený workshop: vzhledem k tomu, že tato metoda sběru dat vyžaduje přítomnost zástupců VŠ a MŠMT pouze na omezenou dobu, může dojít k potížím se získáním dostatečného množství zástupců vysokých škol. Bylo by vhodné podpořit účast VŠ ze strany MŠMT. Z dosavadních konzultací víme, že žádost MŠMT by musela být dobře odůvodněna se zdůrazněním přínosů pro VŠ (např. možnost sdělit ministerstvu své dosud nenaplněné potřeby).

Data z rozhovorů se zástupci VŠ: partnery pro rozhovory je možné vybírat z větší skupiny potenciálních respondentů, což zjednoduší jejich získávání pro spolupráci. I zde by bylo vhodné oslovovat zástupce VŠ souběžně se žádostí od MŠMT.

Celkově je dostupnost dat z výročních zpráv i z databáze Scopus dobrá, data jsou veřejně přístupná a jejich užívání není právně omezené ani licencované. Data z výročních zpráv jsou úplná, lze je získat od všech VŠ. Poněkud omezená je interpretovatelnost dat z tabulkových příloh výročních zpráv. Podrobné metodiky vysvětlující význam jednotlivých tabulek a další detailnější informace s daty a jejich významem související se nám nepodařilo získat.

## 3.2 Srovnatelnost

Tabulkové přílohy VZ: vzájemná srovnatelnost údajů mezi školami i meziroční je dobrá. Mezi lety může docházet ke změnám struktury požadovaných dat, základní informace potřebné pro zachycení zahraniční mobility, počtů programů a studentů Joint/Double Degree programů a zahraničních pracovníků však zůstávají spíše dlouhodobě jednotné, nejpozději od roku 2018. Rizikem jsou budoucí změny ve struktuře vykazovaných dat ze strany MŠMT.

Textové přílohy VZ: srovnatelnost textových příloh je spíše dobrá, jsou odpovědi na jednotně formulované otázky MŠMT. Tyto otázky se v průběhu času mohou měnit, k tematickému okruhu Internacionalizace byla od roku 2022 přidána otázka na smíšené virtuální a fyzické mobility. Formulace otázky je však jediným požadavkem na jednotnost odpovědí. Z analýzy textů vyplývá, že každá VŠ volí jiný přístup ke zpracování odpovědí, délku odpovědi i přístup k výběru informací k zařazení do textové přílohy. To porovnatelnost dat do určité míry omezuje. Domníváme se však, že i přesto lze nalézt hlavní témata týkající se internacionalizace VŠ a sledovat jejich vývoj v čase.

Databáze Scopus: Její datová základna je jedinečnou možností získat informace o mezinárodním provázání českých VŠ. Srovnatelnost dat ze Scopusu mezi jednotlivými roky by však v případě změn v databázi mohla způsobit potíže.

Scopus průběžně rozšiřuje a upravuje svou indexaci, což může ovlivnit počet záznamů i strukturu metadat (např. přiřazení institucí či zemí). V průběhu let se také mění metodika sjednocování názvů institucí (tzv. affiliation profiles), což může způsobit, že spolupráce českých vysokých škol se zahraničními partnery nejsou v různých letech zachyceny zcela konzistentně. Nové časopisy mohou být zařazovány a vyřazovány, což se rovněž promítá do časových srovnání. Tyto překážky však lze zohlednit v interpretaci dat, pravděpodobnost zásadních problémů není velká.

Při analýze trendů mezinárodní spolupráce je také možné se zaměřit spíše na relativní změny a dlouhodobé tendence než na absolutní hodnoty a jejich srovnání mezi jednotlivými roky.

Řízený workshop: srovnatelnost v případě rozhodnutí o každoročním konání workshopu bude zajištěna zařazením dlouhodobě využitelných témat k diskusi. Je však třeba počítat s tím, že meziroční srovnatelnost nebude stoprocentní. Pravděpodobně budou přibývat v následujících letech nové podstatné otázky a ubývat starší již nerelevantní témata. Lze však zpracovávat dlouhodobý přehled diskutovaných témat, bariér atd.

Data z rozhovorů se zástupci VŠ: Srovnatelnost kvalitativních dat je složitější, předpokládáme, že se budou objevovat silná témata, jejichž vývoj bude v rozhovorech možné sledovat v průběhu času i mezi VŠ.

### 3.3 Validita a reliabilita

V kvalitativní analýze se v současnosti posuzuje kvalita zpracování dat reprezentovaná reliabilitou a validitou zejména skrze reflexivitu, systematický postup, inter kódovou shodu, členské ověření, triangulaci a využití nástrojů pro dokumentaci analytických kroků. Tyto principy lze plně aplikovat i při analýze dokumentů a vyhodnocení realizovaných rozhovorů. Kvalita dat bude zajištěna zejména systematickým přístupem k vytváření kódovacího schématu (zvýšení konzistence a věcné validity kódů), společnou prací při kódování a interpretaci dat (zvýšení inter-rater reliability a objektivitu interpretace), zapojením externích expertních spolupracovníků (minimalizace potenciální zaujatosti, ověření validity) a v neposlední řadě i formou metodologické triangulace<sup>13</sup>, při které se budou vzájemně doplňovat a potvrzovat data získaná polostrukturovaných rozhovorů a diskusního workshopu.

---

<sup>13</sup> Formy triangulace popisuje J. Hendl v knize Kvalitativní výzkum: datová triangulace, triangulace výzkumníků, metodologická triangulace, teoretická triangulace. Všechny tyto přístupy zvyšují

V případě statistik z výročních zpráv VŠ je podstatné zejména ověření logické validity, tj. konzistence dat zejména mezi jednotlivými roky. Ta není u všech VŠ optimální, v některých případech se nám nepodařilo ani po konzultaci najít vysvětlení pro značné výkyvy v meziročních rozdílech – jedná se tedy pravděpodobně o chyby v datech. To se týká zejména mobility zaměstnanců. Na základě rešeršemi a konzultacemi zjištěných informací je míra, v jaké tyto statistiky zachycují skutečný stav mobility zaměstnanců, ovlivněna způsobem získávání dat uvnitř VŠ. Také jsme narazili na některé nejasnosti v pokynech k vyplnění tabulkové přílohy<sup>14</sup>

Nepřesnosti mohou pocházet zejména z nedostatečné pozornosti věnované statistikám mobility zaměstnanců (a statistikám zaměstnanců obecně) ve srovnání se studenty. Zatímco mobilita studentů je jednotně zachycována v IT systémech škol, interní informační systémy zachycující údaje o zaměstnancích jsou nejednotné a tento údaj zpravidla nezachycují.

Potíže s kvalitou těchto dat uváděla i Zpráva z výzkum ZMOVYS<sup>15</sup> realizovaného v letech 2008-2009: „...pokyny pro zpracování informací o uskutečňování zahraniční mobility jsou natolik nejednoznačné, že nelze spoléhat ani na jejich validitu, ani na jejich reliabilitu“<sup>16</sup>.

Tabulky s výkazy o počtu studijních programů a studentů v Joint/Double Degree programech (tabulka 2.3: Joint/Double Degree studijní programy realizované se zahraniční VŠ) čerpají z dat o studentech a studijních programech. V těchto údajích jsme nenalezli podstatné problémy, jejich kvalitu považujeme za dobrou.

Kvalitu dat o zahraničních pracovnících zaměstnaných na českých vysokých školách (tab. 6.5) lze považovat z důvodu původu dat ve mzdových údajích vysokých škol za spolehlivé. Nepříznivé pro účely analýzy je využívání přepočtených stavů zaměstnanců (od roku 2018), které neumožňuje dobře rozlišit, kolik fyzických osob – cizích státních příslušníků se v akademickém prostředí reálně pohybuje.

Vzhledem ke smíšené metodice bude v zajištění kvality dat hrát klíčovou roli triangulace dat, umožněná porovnáním a doplněním závěrů ze statistik, textů a vlastního šetření, a dále expertní a kolektivní týmová validace výstupních textových interpretací dat.

### 3.4 Rizika a překážky využití zdrojů dat

Velké problémy při zpracování dat se vyskytují v případě statistik o zahraniční mobilitě zaměstnanců. Variabilita v jejím vykazování je poměrně velká a není vždy dobře

---

<sup>14</sup> Na dvou místech je rozdílný údaj o minimální délce zahraničních pobytů zaměstnanců: údaj „alespoň 5 dní“ uvádí MŠMT u vyjíždějících a přijíždějících akademických a ostatních pracovníků v poznámce pod tabulkou 7.2. Údaj „delší než 5 dní“ je uveden na prvním listu metodiky ke zpracování tabulkové přílohy zprávy o činnosti VŠ v popisu tabulky 7.2.

<sup>15</sup> Více o projektu viz <https://www.csvs.cz/ukoncene-projekty/zmovys/>.

<sup>16</sup> Závěrečná zpráva z výzkumu ZMOVYS, str. 113, online: <https://csvs.cz/wp-content/uploads/zaverecnavyzkumnazpravazmovys.pdf>

vysvětlitelná. Data z roku 2016, zřejmě po zavedení nové metodiky, vykazují těžko interpretovatelné výkyvy<sup>17</sup>.

Ze strany MŠMT jsou jasně dané údaje a struktury sbíraných dat pro VZ veřejných VŠ, avšak uvnitř jednotlivých VŠ jsou postupy sběru dat nejednotné a v některých případech i vnitřně roztržité. Na řadě škol zatím jednotný elektronický systém nezachycuje údaje o mobilitě zaměstnanců.

Změny v metodice, které nastaly například od roku 2019 u tabulky 6.5 (zaměstnanci s cizí státní příslušností) jsou příkladem rizika, které ohrožuje využití dat z tabulkových příloh výročních zpráv VŠ jako vstupu pro časovou řadu. V případě, že MŠMT rozhodne o změně parametrů tabulky, nebo o jejím zrušení, nebude možné v časové řadě pokračovat.

Překážkou pro efektivní získávání a zpracovávání dat z výročních zpráv je i formát zveřejňovaných výročních zpráv, zejména tabulkové přílohy, které jsou nejčastěji zveřejněny jako pdf soubory, ačkoli MŠMT od vysokých škol dostává vhodnější excelovské tabulky.

Podobně je tomu i u textových částí výročních zpráv, jejichž obsah je sice od roku 2019 stabilní, nicméně pokládané otázky stanovuje MŠMT, a v případě jejich změny nebude možné pokračovat s porovnáním výsledků v čase.

U řízené diskuse s experty VŠ a MŠMT je největším rizikem jejich neochota či nedostatek času a nemožnost spolupracovat. Toto riziko lze snížit jasným a dobře komunikovaným záměrem ze strany MŠMT.

Rizikem pro úspěšné provádění analýzy dat ze Scopusu a jejího opakovaného provádění v dalších letech jsou změny v databázi Scopus, její struktury, definicích kategorií jednotlivých dat (např. afiliace), ale také případné změny v rozsahu veřejnosti poskytovaných dat.

## 4. Plán pro další období

Další postup v rámci monitoringu internacionalizace vysokých škol bude zaměřen jak na další rozvoj metodiky, tak na realizaci pilotních analýz.

V oblasti kvantitativních dat získaných z výročních zpráv VŠ půjde zejména o zavedení a pravidelné doplňování časových řad a zpracování textového komentáře k nově získaným

---

<sup>17</sup> Proto jsme tento rok využili jen jako doplněk a pracovali jsme s lety 2019 a 2023.

datům. Rovněž u textových dat budou pokračovat práce na tematické analýze textů VŠ o internacionalizaci v dlouhodobém časovém horizontu.

Předpokládá se příprava a realizace řízené diskuse se zástupci VŠ a MŠMT o stavu internacionalizace, potřebách VŠ, bariérách a osvědčených postupech. Diskuse bude uskutečněna souběžně s polostrukturovanými rozhovory se zástupci škol, které budou mít stejné tematické zaměření doplněné ještě o konkrétní otázky týkající se ostatních využívaných datových zdrojů. Řízená diskuse a rozhovory se budou vzájemně doplňovat.

Bude realizována analýza dat získaných z databáze Scopus, která umožní získat informace o míře spolupráce českých veřejných VŠ se zahraničními institucemi na vědeckých článcích. Analýza proběhne na úrovni VŠ, testovány budou možnosti zajištění dat na úrovni částí VŠ a začlenění hodnocení kvality do analýzy.

Kromě pokračování ve zpracování dat z dosud prozkoumaných zdrojů budou vyhledávány a ověřovány další vhodné zdroje dat, např. ty, které souvisejí s hodnocením kvality VŠ (sebeevaluační zprávy, veřejně dostupné podklady a výstupy k hodnocení kvality VŠ).

Závěrem bude zhodnocena kvalita a využitelnost analýzami získaných dat, informací a závěrů a navrženy úpravy, zejména v pokračování časových řad, ale také nové zdroje, nebo potřebné doplnění informací.

## 5. Seznam zkratek

DDP – Double degree programme - studijní program realizovaný se zahraniční VŠ

JU – Jihočeská univerzita

LLM – Velký jazykový model (Large Language Model) v oblasti umělé inteligence

MŠMT – Ministerstvo školství, mládeže a tělovýchovy

MUNI – Masarykova univerzita

NLP - zpracování přirozeného jazyka (Natural Language Processing)

OU – Ostravská univerzita

SU – Slezská univerzita

TUL – Technická univerzita Liberec

UK – Univerzita Karlova

VŠ – vysoké školy

VŠE – Vysoká škola ekonomická v Praze

## 6. Použitá literatura

Baranowski, M.: Epistemological aspect of topic modelling in the social sciences: Latent Dirichlet Allocation. In: *Przegląd krytyczny* 2022, vol. 4, nr. 1, str. 7 – 16. Dostupné na: [https://www.researchgate.net/publication/362839762\\_Epistemological\\_aspect\\_of\\_topic\\_modelling\\_in\\_the\\_social\\_sciences\\_Latent\\_Dirichlet\\_Allocation#full-text](https://www.researchgate.net/publication/362839762_Epistemological_aspect_of_topic_modelling_in_the_social_sciences_Latent_Dirichlet_Allocation#full-text)

Braun, V.; Clarke, V.: Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, č. 2, 2006, str. 77–101. Dostupné na: <https://doi.org/10.1191/1478088706qp063oa>.

Braun, V.; Clarke, V.; Weate, P.: Using Thematic Analysis in Sport and Exercise Research. In *Routledge Handbook of Qualitative Research in Sport and Exercise*, editovali Brett Smith a Andrew C. Sparkes. London 2016, nakladatelství Taylor & Francis, Routledge. Dostupné z: <https://uwe-repository.worktribe.com/output/3123247>

CSVŠ, UJAK: Zahraniční mobilita vysokoškolských pracovníků. Závěrečná výzkumná zpráva. Praha 2009, dostupné na: <https://csvs.cz/wp-content/uploads/zaverecnavyzkumnazpravazmovys.pdf>

De Paoli, S.: Further Explorations on the Use of Large Language Models for Thematic Analysis. Open-Ended Prompts, Better Terminologies and Thematic Maps". In *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 25, č. 3, 2024. Dostupné na: <https://doi.org/10.17169/fqs-25.3.4196>.

De Paoli, S.; Fawzi, A.: TALLMesh: a simple application for performing Thematic Analysis with Large Language Models. arXiv:2504.13892. Preprint, arXiv, 5. duben 2025. Dostupné z: <https://doi.org/10.48550/arXiv.2504.13892>.

Hammersley M. *Reading Ethnographic Research: a Critical Guide*. Londýn 1990, Longmans.

Hendl, J.: *Kvalitativní výzkum. Základní teorie, metody a aplikace*. Čtvrté, přepracované a rozšířené vydání. Portál 2016.

Kidder, L. H., Fine, M.: Qualitative and Quantitative Methods: When Stories Converge. In: *New Directions for Program Evaluation* 1987, č. 35 (1987), str. 57–75. <https://doi.org/10.1002/ev.1459>.

Lindstedt, N.C.: Structural Topic Modeling For Social Scientists: A Brief Case Study with Social Movement Studies Literature, 2005–2017. 2019, in: *Social Currents*, 6(4), str. 307–318. Dostupné na: <https://doi.org/10.1177/2329496519846505>.

Minksová, L., Provázková, K.: Struktura výjezdů vysokoškolských pracovníků na zahraniční pracovní cesty a jejich důvody k účasti i neúčasti na zahraniční mobilitě, *Aula*. 2010, 18, 1, s. 21–28.

Minksová, L., Provázková, K.: Zahraniční mobilita vysokoškolských pracovníků: rezervy z pohledu účastníků a organizátorů, *Aula*. 2010, 18, 1, s. 28–37.

Mongeon, P.; Paul-Hus, A.: The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228; 2016. Dostupné na: <https://doi.org/10.1007/s11192-015-1765-5>

Silge, J.; Robinson, D.: Text mining with R: A tidy approach. O'Reilly Media, 2017.

Silverman, D.: *Ako robiť kvalitatívny výskum: praktická príručka*. Ikar, 2005. ISBN 055109044.

Vayansky, I.; Kumar, S.A.P.: A review of topic modeling methods. In: *Information Systems* 94, 101582, 2022. Dostupné na: <https://doi.org/10.1016/j.is.2020.101582>

Výroční zpráva o činnosti za rok 2018, Masarykova univerzita. Dostupé na: [https://www.muni.cz/media/3169388/mu\\_vzoc\\_2018-cz.pdf](https://www.muni.cz/media/3169388/mu_vzoc_2018-cz.pdf)

Výroční zprávy o činnosti veřejných VŠ z let 2016, 2019 a 2023

Zhang, H., Wu, C., Xie, J., Kim, C., & Carroll, J. M.: QualiGPT: GPT as an easy-to-use tool for qualitative coding. *arXiv preprint 2023*, dostupné na: <https://arxiv.org/abs/2310.07061>

Zhang, H., Wu, C., Xie, J., Rubino, F., Graver, S., Kim, C., Carroll, J. M., & Cai, J.: When Qualitative Research Meets Large Language Model: Exploring the Potential of QualiGPT as a Tool for Qualitative Coding. *arXiv preprint 2024*, dostupné na: <https://arxiv.org/abs/2407.14925>.

### **Webové zdroje:**

<https://www.csvs.cz/ukoncene-projekty/zmovys/> - projekt ZMOVYS – zahraniční mobilita pracovníků vysokých škol

<https://msmt.gov.cz/vzdelavani/vysoke-skolstvi/vyrocní-zpravy-o-cinnosti-vysokych-skol> - Výroční zprávy vysokých škol, stránka MŠMT

## Příloha 1: Vizualizace sítě slov

Ve zpracování textových dokumentů zpráv jsme vycházeli z přístupů navržených v knize: J. Silgeho a D. Robinsona z roku 2017 „Text mining with R: A tidy approach“. Cílem bylo identifikovat hlavní tematické okruhy a významová spojení v textových odpovědích jednotlivých škol. Postup zahrnoval následující kroky:

### 1) Předzpracování textu

Textová data byla importována z CSV souboru a transformována do formátu „tidy“ pomocí balíčků tidyverse a tidytext. Texty byly tokenizovány na jednotlivá slova (funkce `unnest_tokens()`) a očištěny o česká i anglická stop-slova (slova bez významové informační hodnoty), včetně rozšířeného seznamu běžných slov (spojky, zájmena, předložky, částice). Výsledkem byl soubor jednoslovných tokenů spojených s identifikátorem školy.

### 2) Váhování slov pomocí TF-IDF

Pro zjištění relativní důležitosti slov v rámci jednotlivých škol byla použita metoda TF-IDF (Term Frequency – Inverse Document Frequency) implementovaná funkcí `bind_tf_idf()` z balíčku tidytext. TF-IDF umožňuje potlačit běžně se vyskytující výrazy a zvýraznit slova, která jsou typická pro konkrétní dokument (školu). Následně byla vybrána sada nejvýraznějších slov s nejvyšší hodnotou TF-IDF.

### 3) Korelační a síťová analýza slov

Pro identifikaci slov, která se často vyskytují společně v rámci textů, byla využita korelační analýza spoluvýskytů pomocí funkce `pairwise_count()` z balíčku widyr. Na základě četnosti společného výskytu dvojic slov byla sestavena síťová reprezentace s uzly (slova) a hranami (spoluvýskyty). Síť byla analyzována a vizualizována pomocí knihoven `igraph` a `ggraph`, přičemž pro detekci tematických shluků byla použita metoda Louvain community detection (`cluster_louvain()`).

### 4) Analýza víceslovných výrazů (n-gramy)

Pro doplnění jednoslovné analýzy byla vytvořena i sada n-gramů (nejčastější dvouslovné a tříslavné fráze), opět pomocí `unnest_tokens()` s parametrem `token = "ngrams"`. Tento postup umožnil identifikovat ustálené víceslovné výrazy a fráze typické pro daný kontext (např. „projektová výuka“, „spolupráce žáků“).

### 5) Tematická analýza LDA

Pro explorativní identifikaci tematických oblastí byl aplikován topic modeling metodou Latent Dirichlet Allocation (LDA) pomocí knihovny `topicmodels`. Model LDA rozdělil korpus do předem zvoleného počtu témat, přičemž každé téma bylo popsáno souborem slov s nejvyšší pravděpodobností výskytu. Záměrem bylo ověřit, zda je možno tímto způsobem porovnat témata napříč školami a popsat hlavní obsahové

