



# Doporučení pro práci s bibliometrickými indikátory a jejich interpretaci

Toto doporučení je vydáváno pro pracovníky MU, kteří pracují s bibliometrickými indikátory (například impakt faktor, h-index, apod.). Zejména v situacích, kdy může bibliometrie ovlivnit hodnocení jednotlivce (příp. skupiny) je zapotřebí znát omezení bibliometrie a metodické základy využití bibliometrických indikátorů. Doporučení obsahuje obecné zásady práce s bibliometrickými indikátory a způsob nakládání s nimi. **Všechna doporučení vycházejí z mezinárodně obecně uznávaného metodologického konceptu a z dobré praxe expertů na scientometrii.**<sup>1</sup>

Bibliometrie je založena na datech z **bibliografických databází**. Nejvýznamnějšími producenty dat napříč obory jsou v současnosti společnosti:

1. Thomson Reuters (platforma Web of Science, WoS) a
2. Elsevier (platforma SCOPUS).

Masarykova univerzita pro snadnou tvorbu bibliometrických indikátorů z dat Web of Science a pro analýzy publikační činnosti využívá nástroj **InCites** společnosti Thomson Reuters. V současné době lze využít dvě verze (generace) systému – InCites 1<sup>2</sup> (přístup je omezen na úzkou skupinu uživatelů) a InCites 2<sup>3</sup>. Zdrojem pro analýzy v **InCites** je specializovaný dataset publikací a autorů Masarykovy univerzity, který je připravený Odborem výzkumu RMU na základě publikací vložených v IS MU, resp. dataset podle adresy ve WoS.

InCites umožňuje širokou škálu různých analýz na úrovni MU, pracovišť i jednotlivců. InCites umožňuje také základní benchmarking v porovnání se světovou produkcí. **Analýzu s využitím nástroje InCites zpracuje dle konkrétních požadavků Odbor výzkumu RMU.**

Kromě multioborových databází existuje řada oborově zaměřených databází, které lze ve vhodné situaci rovněž využít. Jsou to například [MathSciNet](#) (matematika), [SciFinder – CAS](#) (chemie, biochemie

---

<sup>1</sup> HICKS, D, et al. Bibliometrics: The Leiden Manifesto for research metrics. Nature. 2015, vol. 520, 7548, 429–431. doi: <http://dx.doi.org/10.1038/520429a>. Dostupné z: <http://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351>; San Francisco Declaration on Research Assessment (DORA), Dostupné z: <http://www.ascb.org/dora/>.

<sup>2</sup> Dostupné z <http://incites.isiknowledge.com>.

<sup>3</sup> Dostupné z <https://incites.thomsonreuters.com/#/analytics>.

a příbuzné obory), [PubMed](#) (medicínské obory), [ADS – Astrophysics Data System](#) (astrofyzika). Tyto databáze mají obvykle pro tvorbu bibliometrických indikátorů větší omezení, než tradiční producenti. Pro patenty existují databáze Derwent (součást Web of Science), [EPO- PATSTAT](#), [USPTO](#), [DEPATISnet](#), [WIPO](#).

V současné době se diskutuje o využití **Google Scholar** pro účely bibliometrie. Jedná se o bibliograficky specializovanou vyhledávací službu, která na základě vlastního algoritmu umožňuje zobrazovat reference k vybraným dokumentům. Pro bibliometrické analýzy je možné využít Google Scholar jen se značnými komplikacemi (viz oddíl „Práce s databázemi“), ale jeho výhodou je, že umožňuje sledovat ohlas produkce oborů, které mají v tradičních databázích slabé nebo žádné zastoupení (např. některé společenské a humanitní vědy).

Bibliometrických výstupů využívají s různou vahou také **žebříčky vysokých škol** (Leiden Ranking, THE World University Rankings, Shanghai Ranking, a jiné). V tomto kontextu se ale můžeme setkat s větší mírou mylné interpretace bibliometrických ukazatelů.

V dalším textu rámcově představíme jednotlivé databáze a způsob využití určitých konkrétních indikátorů.

## Obecné zásady a doporučení

1. Bibliometrie je nejspolehlivější pro účely porovnávání a analyzování dat na makro úrovních (země, případně instituce). Směrem k nižším jednotkám hodnocení (pracoviště, jednotlivci) se její vhodnost pro porovnávání a analyzování výzkumné výkonnosti snižuje a je vždy nutné je doplnit kvalitativním hodnocením<sup>4</sup> (např. problém tzv. „false positive“ – pokud se bibliometrickou metodou najdou rozdíly mezi dvěma pracovníky na pracovišti, nemusí to nutně znamenat, že je mezi nimi skutečný kvalitativní rozdíl). V případě analýz na individuální úrovni lze autorizací samotným výzkumníkem, či využitím pravidelně aktualizovaného profilu ResearcherID dosáhnout přesnějších vstupních dat.
2. Bibliometrie má sloužit jako jeden z podkladů pro následné kvalitativní hodnocení. Bibliometrii není vhodné přímo využívat jako jediný podklad pro hodnocení jednotlivců, či skupin, zejména v případech, kdy toto hodnocení může ovlivnit kariéru a personální otázky.
3. Analýza by měla vždy využívat několik zdrojů současně (zpravidla 3 zdroje). Kromě multioborových databází je doporučeno využít oborově zaměřenou databázi, pokud je to možné.
4. Výsledky bibliometrických analýz by měly být transparentní a všem srozumitelné. Hodnocená jednotka by měla mít vždy možnost data verifikovat. Pokud jsou bibliometrické analýzy jedním z podkladů pro hodnocení, měla by být kritéria transparentní a předem známá.
5. Pro získání věrohodné informace o sledované jednotce je zapotřebí využívat škálu relevantních indikátorů, nikoli pouze jeden indikátor. Pro zodpovědnou interpretaci údajů je třeba sledovat indikátory ve vzájemné souvislosti.

---

<sup>4</sup> Například v případě porovnání H-indexu je třeba zohlednit obor, věk, délku publikační aktivity, typy výstupů či vývoj publikační aktivity v čase.

6. V bibliometrii je nutné vždy srovnávat podobné s podobným (např. v rámci oboru). Nelze například porovnávat h-index vědce z oboru historie a z oboru biologie – obory mají diametrálně odlišné citační zvyklosti a strukturu obvyklých výstupů.
7. Excelenci je nutné hledat v relevantním prostorovém kontextu.
8. Metriky vztahující se k časopisům (např. impakt faktor) a metriky z něj odvozené je potřeba vyřadit z úvah o financování a personálních otázkách (zaměstnávání, povýšení) a k hodnocení jednotlivců.
9. Výkon je potřeba měřit a interpretovat v kontextu výzkumné strategie hodnocené jednotky.
10. Za účelem hodnocení výzkumu je nutné zvažovat kvalitu a dopad všech výstupů výzkumu, nejen publikací, přičemž je třeba brát v úvahu širokou škálu měřítek dopadu včetně kvalitativních indikátorů (např. vliv na používané metody a praxi).
11. Výzkum je potřeba hodnotit také na základě jeho přínosu, či dopadu, a nikoli pouze na základě časopisu, v němž je zveřejněn, nebo pouze na základě publikačních měřítek.
12. Nelze připustit, aby nevhodně zvolené indikátory přímo ovlivňovaly chování, či kariéry vědců. Informované kvalitativní hodnocení (např. panel/peer review) není ideální, ale stále nejlepší možný nástroj hodnocení a řízení vědy.
13. Data je potřeba vždy čistit a verifikovat. Čištění dat obvykle podléhá shody jmen, různé varianty adres a názvů institucí, obory, typy výsledků. Pokud nejsou data čistá, nejsou spolehlivým zdrojem pro analýzu.

## Zdroje dat: Web of Science, SCOPUS, Google Scholar

Politika producenta dat obvykle determinuje obecně přejímaná kritéria kvality a excelence. Při práci s databázemi je proto potřeba počítat s těmito vlastnostmi a omezeními:

Web of Science	SCOPUS
Přísná kritéria pro indexaci (obsahuje méně zdrojů). Velmi omezeně jsou indexovány knihy. Databáze má velké omezení pro použití v případě určitých humanitních a sociálních věd. V případě analýzy citovaných zdrojů lze využít „Cited Reference Search“, který obsahuje také časopisy a výsledky, jež nejsou indexovány ve WoS, např. knihy. Je například vhodné využít Cited Reference Search pro zjištění přesnějšího h-indexu, zejména pro sociální a humanitní vědy.	Mírněji nastavená kritéria pro indexaci (obsahuje více zdrojů). Velmi omezeně jsou indexovány knihy. Databáze má velké omezení pro použití v případě určitých humanitních a sociálních věd. Oproti WoS je však zastoupení humanitních a sociálních věd větší.
Velký podíl časopisů pocházejících z anglo-amerického prostředí, obsahuje malý podíl lokálních zdrojů.	Více „evropsky“ orientovaná databáze s větším podílem lokálních zdrojů.
WoS není vhodným zdrojem dat pro všechny obory. Společenské a humanitní obory až na výjimky nemají dostatečné zastoupení zdrojů a tím ani potřebné množství zdrojových záznamů.	SCOPUS má obecně širší záběr, je vhodnějším zdrojem pro společenské a humanitní vědy a pro konferenční příspěvky.
Data je potřeba vždy čistit a verifikovat. Čištění	Data je potřeba vždy čistit a verifikovat. Čištění

dat obvykle podléhají shody jmen, různé varianty adres a názvů institucí, obory.	dat obvykle podléhají shody jmen, různé varianty adres a názvů institucí, obory.
Využívá impakt faktor pro měření popularity zdrojů.	Využívá SJR faktor (SCImago Journal Rank) pro měření popularity zdrojů.

Využití Google Scholar pro účely bibliometrie je stále velmi problematické, je zapotřebí data podrobit masivní kritice a čištění. Pro obory nepokryté tradičními databázemi (viz výše) je však Google Scholar vhodné vzít v úvahu.

Při práci s Google Scholar je potřeba počítat s těmito vlastnostmi a omezeními:

<b>Google Scholar omezení</b>	<b>Google Scholar výhody</b>
Netransparentnost – využívá algoritmus, který vyhodnocuje, co je vědecký příspěvek. Tento algoritmus není zveřejněn, je tedy nutné manuálně rozhodnout, které výsledky vyhledávání jsou relevantní (vědecké x populární).	Široká nabídka typů dokumentů – kromě tradičních „articles“, „books“, také reporty, prezentace, závěrečné práce, učební materiál.  Široké geografické a oborové pokrytí. Pro sociální a humanitní vědy perspektivní zdroj.
Google Scholar ve výsledcích vyhledávání může poskytovat různé verze téhož výstupu, podle toho, kde všude je výstup zveřejněn.	Poskytuje doplňkové služby (profil uživatele, základní metriky citovanosti).
Není specifikováno, která místa jsou zahrnuta do vyhledávání. Není také možná kontrola nad změnami webového obsahu, tzn. stejné vyhledávání může vrátit různé výsledky v závislosti na čase.	Prohledává i repozitáře.  Nabízí fulltext přístup (pokud je dostupný).
Manipulovatelnost dat (možnost přihlášení se k cizímu výsledku, fake articles).	Google Scholar je moderní výzva – velké pole pro vylepšování.
Není kontrola čistoty dat – nutnost extrémní opatrnosti při práci s daty zejména na úrovni jednotlivce. Data je potřeba vždy čistit a verifikovat.	Pro analýzy výsledku vyhledávání Google Scholar slouží software Harzing's Publish or Perish, dále různé aplikace pro Android a iOS a přidružené funkcionality (např. UGRinvestiga); doplňky pro prohlížeče.
	Propojení vyhledávání akademických výstupů se sociální sítí (pokud mají spoluautoři profil na Google Scholar) a dalšími funkcionalitami. Mají-li autoři profil na Google Scholar, je její použití výrazně spolehlivější, protože to znamená, že sami autoři verifikovali svoje publikace.
	Významný nástroj pro zvýšení viditelnosti zvláště pro mladé vědecké pracovníky.

# Zásady práce s vybranými bibliometrickými indikátory

## 1. Impakt faktor

Impakt faktor (Thomson Reuters) byl původně vytvořen jako nástroj, který měl pomáhat knihovníkům vyhledávat vlivné časopisy vhodné k nákupu. Měří vlivnost (popularitu) časopisu, nikoli dopad vědy článku nebo jednotlivce. To vyplývá z metody výpočtu:

$$\text{Impakt faktor v roce } X = \frac{\text{všechny citace všech článků v daném časopise získané v letech } X-1 \text{ a } X-2}{\text{počet „citovatelných“ výstupů (obvykle article, note a review), které byly v daném časopise publikovány v letech } X-1 \text{ a } X-2.}$$

V současnosti je impakt faktor (IF) velmi vlivným údajem, neboť je pro svou jednoduchost a srozumitelnost plošně využíván, někdy ovšem v nesprávných souvislostech.

Ve prospěch IF platí, že (ve většině oborů) existuje korelace mezi kvalitou a IF. **Časopisy s vysokým IF mají nejnáročnější peer review, které stále zdokonalují.** Při hodnocení jednotlivce se samozřejmě posuzuje v jakých časopisech (s jakým IF) daný pracovník publikuje. Nicméně nelze to použít jako hlavní indikátor kvality, vždy je třeba vzít úvahu další bibliometrické údaje a kvalitativní posouzení (viz dále).

V neprospěch IF platí, že výše IF není pro autory jedinou motivací. Někteří autoři cíleně publikují v časopisech v nižším IF, pokud se například jedná o oborově specifické časopisy s cílem zasáhnout svoje publikum. V mnoha oborech se však výše IF nesleduje vůbec a i pro špičkové vědce není výše impakt faktoru relevantní.

V případě výzkumu s rychle stárnoucími daty může být také motivací cílit na méně prestižní časopis s kratším a jednodušším procesem publikování. Některé obory (např. informatika) také publikují své výsledky primárně prostřednictvím konferencí a sborníků z konferencí, protože zde je nejrychlejší proces publikování. **Proto by měl být IF používán pouze k účelům, ke kterým je vhodný, tj. k evaluaci časopisů. Využití impakt faktoru jako jediného indikátoru pro hodnocení článků, institucí, či osob je odmítáno odbornou scientometrickou komunitou, ale i samotným producentem dat.**<sup>5</sup>

Impakt faktor má řadu omezení:

- IF poutá pozornost k časopisu, ale neznamená, že bude článek v tomto časopise více citován.
- Počet citovatelných článků (jmenovatel) může být ovlivněn velikostí časopisu.
- Výhodu mají starší časopisy oproti novým a časopisy s vyšší periodicitou oproti časopisům, které vycházejí méně často.
- IF lze manipulovat pomocí redakční politiky.<sup>6</sup> Výhodu mají review časopisy (články typu review jsou obvykle více citovány).

<sup>5</sup> Například The State of Journal Evaluation. Dostupné z: <http://stateofinnovation.thomsonreuters.com/how-to-measure-up-impact-factor-2015>.

<sup>6</sup> The PLoS Medicine Editors. The impact factor game. PLoS Medicine. 2006, 3, 6, e291, doi: <http://dx.doi.org/10.1371/journal.pmed.0030291>.

- e) Nevyváženost (IF je silně ovlivnitelný menšinou vysoce citovaných článků).<sup>7</sup>
- f) Vlastnosti IF jsou specifické pro různé obory. Nelze mezi sebou porovnávat absolutní hodnoty impakt faktorů časopisů z různých oborů.
- g) Netransparentní proces určování kritérií pro výpočet IF.

IF je kalkulován na 3 desetinná místa, což evokuje dojem přesnosti údaje. Z výše uvedených vlastností ovšem vyplývá, že výpočet může ovlivňovat řada „měkkých“ faktorů a proto zejména mezi časopisy s drobným rozdílem v absolutní výši IF nelze pouze na základě výše IF rozhodnout o jejich kvalitě.

Sčítání impakt faktorů časopisů pro účely hodnocení určité jednotky (tzv. **kumulativní impakt faktor**) je výraznou ukázkou zneužití impakt faktoru pro nevhodné účely. Tento údaj konstatuje, jaký je součet impakt faktorů časopisů, ve kterém daný autor publikoval. Nevypovídá ale nic o vědeckých kvalitách autora ani o dopadu jeho článků. Hodnota tzv. kumulativního IF stoupá s věkem autora. Tento údaj rovněž nelze k ničemu vztáhnout, porovnávat, a to ani v rámci oboru.

#### Příklad hodnocení dvou kolegů na základě kvantitativních údajů:

Kolega A	Kolega B
<ul style="list-style-type: none"> <li>- 1 článek z roku 2012 v časopise s IF 12, multioborový časopis.</li> <li>- Počet citací: 9.</li> <li>- Konsorcium 55 autorů.</li> </ul>	<ul style="list-style-type: none"> <li>- 1 článek z roku 2012 v časopise s IF 4, specializovaný oborový časopis.</li> <li>- Počet citací: 30.</li> <li>- Bilaterální spolupráce se zahraniční univerzitou.</li> <li>- Článek je v odborné komunitě sdílen prostřednictvím příspěvků na Twitteru.</li> </ul>
H-index 1	H- index 1

Tento jednoduchý vzorový příklad ukazuje, že nelze posuzovat pouze na základě jediné hodnoty. Pokud vedoucí pracoviště bude posuzovat zaměstnance podle IF časopisu, ve kterém publikuje, bude se jevit kolega A v žebříčku výše a případné personální důsledky by nepůsobily spravedlivě. Publikace v časopise s vysokým IF lze sice vnímat jako úspěch, ale je potřeba posoudit okolnosti a dopad článků samotných, kdo na tom úspěchu má hlavní podíl, tj. zahrnout kvalitativní hodnocení, případně více indikátorů. Po prozkoumání všech okolností by hodnocení dopadu vědy zřejmě vyznělo lépe pro kolegu B. Reálné situace jsou obvykle mnohem složitější.

## 2. h-index

H-index je indikátorem, který měří produktivitu a impakt. Hodnota  $h$  udává počet článků, které mají  $h$  a více citací. Ani h-index není metrika vhodná pro všechny účely. Jeho výhodou je, že stírá zřetel mimořádně citovaným článkem a články, které ještě nebyly citovány. V případě sledování trendu h-indexu v čase lze rozeznat konsistentně kvalitní výzkum v rámci oboru. Je rovněž užitečný pro sledování „rising stars“ v jednotlivých oborech. Omezení h-indexu jsou dána jeho silnou závislostí na oborových specifikách citovanosti.

<sup>7</sup> Editorial. Not so deep impact. Nature 2005, 435, 1003–1004. doi: <http://dx.doi.org/10.1038/4351003b>.

Vlastnosti h-indexu:

- Liší se napříč různými databázemi, ale i v rámci jedné databáze (pro jednu osobu lze spočítat několik h-indexů, dle zvolené databáze a filtru). V rámci WoS lze vyhledat v určitých případech tři rozdílné h-indexy: 1.) z Web of Science Core Collection, 2.) při volbě All Databases a 3.) ručně z Cited Reference Search. Pro získávání h-indexu není jednoznačně platná metodika, proto je při jeho uvádění potřeba vždy uvést zdroj, kde byl h-index získán.
- Také Google Scholar uvádí h-index, jež je typicky vyšší než u WoS, protože Google Scholar zahrnuje mnohem větší objem zdrojů s různou kvalitou.
- Je oborově závislý, nelze srovnávat jeho hodnoty mezi odlišnými obory.
- Nikdy nemůže klesnout v závislosti na čase.<sup>8</sup>

### 3. Ostatní bibliometrické indikátory

Volba sledovaných indikátorů se velmi liší s ohledem na zaměření analyzované jednotky, obor, účel analýzy a sledované jevy, velikost datové základny a mnoha jiných proměnných. Nelze proto stanovit set vhodných indikátorů pro určité účely. V následujícím přehledu proto přinášíme pouze příklady nejčastěji se vyskytujících indikátorů. Indikátory je vždy potřeba interpretovat ve vzájemné souvislosti.

Pro získání přehledu o **aktivitě** analyzované jednotky obvykle poslouží jednoduché statistiky počtu publikací v databázích, struktura typů výsledků a jejich zastoupení v databázích. Pokud sledujeme **publikační výkon**, je již potřeba tyto hodnoty vztáhnout například k počtu FTE pracovníků, množství vynaložených prostředků, apod.

Další skupinou nejčastěji sledovaných informací jsou indikátory **spolupráce**, tedy například míra mezinárodní spolupráce, přehled spolupracujících zemí, institucí, případně osob. Z analýzy sítí spolupráce mohou vyvstat i další informace, jako například nové trendy ve výzkumu.

**Metriky dopadu** založené na citovanosti publikací je potřeba normalizovat a nahlížet v kontextu oboru. Vypovídajícími indikátory jsou v tomto případě různé varianty citačního dopadu, např. normalizovaný citační dopad (poměr získaných citací vůči očekávaným v rámci oboru, časopisu, země, atd.). Ukazatelem kvalitní produkce v oboru je zařazení článku v prvním 1% a 10% nejcitovanějších článků v rámci oboru a odvozené hodnoty percentilu, který udává, kolik % článků je více citováno.

Vhodným doplněním je analýza viditelnosti vyjádřená indikátory, vztahujícími se k používaným zdrojům – např. struktura časopisů, ve kterých jednotka publikuje, popularita těchto časopisů v rámci oboru, zde lze využít také impakt faktor a indikátory z něj odvozené.

---

<sup>8</sup> Variantou je počítat h-index za zvolené časové období, např. posledních 10 let.

## Kontakt a nabídka zpracování bibliometrických analýz na vyžádání

Volba sledovaných indikátorů se velmi liší s ohledem na zaměření analyzované jednotky, obor, účel analýzy a sledované jevy, velikost datové základny a mnoha jiných proměnných. Indikátory sledujeme z pohledu aktivity, publikačního výkonu, spolupráce a dopadu.

Nelze obvykle stanovit set univerzálně vhodných indikátorů pro všechny účely, Odbor výzkumu RMU proto nabízí **zpracování bibliometrické zprávy s využitím nástroje InCites** pro fakulty, pracoviště i jednotlivce na základě individuálních požadavků. Analýze vždy předchází zjištění specifických potřeb. Výsledkem analýzy není hodnocení, ale detailní informace, které lze využít pro strategické řízení, průběžný monitoring, jako podklad pro zdokonalování publikační strategie, apod.

V případě zájmu o konzultaci, nebo bibliometrickou analýzu kontaktujte Odbor výzkumu RMU.

### Kontakt:

**Mgr. Michal Petr**

e-mail: [petr@rect.muni.cz](mailto:petr@rect.muni.cz)

tel: 549 49 5887

<http://vyzkum.rect.muni.cz/>

Zpracoval Mgr. Michal Petr, dne 3.11.2015

Odbor výzkumu RMU  
Rektorát Masarykovy univerzity  
Žerotínovo náměstí 9  
601 77 Brno