Oborové zprávy

Popis statistik pro analýzu výsledků dle databáze Web of Science

# Úvod

V průběhu práce odborných panelů jsou jejich členům a členkám zasílány bibliometrické podklady zpracované *Odborem Rady pro výzkum, vývoj a inovace*, které budou v panelech okomentovány. Smyslem komentářů je zasadit statistiky do příslušného oborového a národního kontextu a poskytnout zpětnou vazbu pro další rozvoj nástrojů bibliometrické analýzy v následujících letech.[[1]](#footnote-1)

Je třeba vést v patrnosti, že pro obory, pro které není publikování v impaktovaných časopisech relevantním měřítkem kvality, jsou poskytnuté údaje pouze doplňkovým kritériem hodnocení, neboť mohou být zavádějící a je třeba tuto skutečnost v komentářích uvést.

# Použité ukazatele

**Article Influence Score** (AIS) reflektuje průměrný vliv článků v určitém časopisu v průběhu prvních pěti let po jejich zveřejnění v rámci databáze Web of Science (WoS). Odvozuje se z ukazatele Eigenfactor, vycházejícího z toho, kolikrát byly články z daného časopisu zveřejněné v posledních pěti letech citovány v databázi časopisů Journal Citation Reports (JCR), a odpovídá tedy zhruba pětiletému impakt faktoru časopisu (5-Year Journal Impact Factor). Eigenfactor ale navíc zohledňuje kvalitu časopisů, ze kterých tyto citace pocházejí – vysoce citované časopisy mají větší váhu než časopisy méně citované. Odkazy z jednoho článku v časopisu na jiný článek v témže časopisu se nezapočítávají. Hodnota AIS pak udává vliv časopisu normalizovaný v rámci databáze JCR. Hodnoty Eigenfactor (a tedy i AIS) jsou úměrné počtu článků publikovaných v daném časopise – pokud dva časopisy publikují stejně kvalitní články a prvním jich vydává dvakrát více, pak je i jeho Eigenfactor dvakrát větší než u druhého časopisu.[[2]](#footnote-2)

Výše ukazatele AIS slouží k rozdělení časopisů (respektive výsledků) do pásem. Pásma jsou definována pomocí **prvního decilu** a **kvartilů**, dalším ukazatelem je **medián**. Pásma jsou vytvářena vždy pro každý obor zvlášť.

**Kvartily** rozdělují pořadí časopisů na čtvrtiny (tj. první kvartil zahrnuje 25 % nejlepších časopisů dle AIS, poslední čtvrtý kvartil naopak zahrnuje spodních 25 % časopisů). **První decil** (P10) AIS tvoří 10 % nejlepších časopisů v oboru. Protože je první decil součástí horního kvartilu, jsou vždy výsledky prvního decilu i součástí prvního kvartilu.

Pro účely mezinárodního srovnání se dále určuje nejen pořadí oborových časopisů, ale také pořadí článků z daného oboru a jejich členění v  kvartilech a v horním decilu. Pořadí článků je vytvořeno pouze z citovatelných dokumentů (citable items) a každému z nich je přiřazeno AIS příslušného časopisu.

Dalším použitým ukazatelem je **medián.** Medián udává „střední“ hodnotu dané veličiny, ale na rozdíl od průměru není citlivý na extrémní minimální a maximální hodnoty. Je určen jako hodnota AIS, která se nalézá uprostřed pořadí časopisů daného oboru, respektive oborových článků.[[3]](#footnote-3) Jinými slovy medián lze též chápat jako hranici mezi druhým a čtvrtým kvartilem. Analýza pracuje s mediány podle pořadí AIS časopisů, využívá však i mediány podle pořadí AIS článků (viz níže).

# Charakteristika dat

## Oborová struktura

Oborový pohled je založený na klasifikaci podle materiálu *Struktura oborů OECD (Frascati Manual) – převodník M17+*, který byl schválen na 324. zasedání Rady pro výzkum, vývoj a inovace dne 31. března 2017 a který určuje i strukturu odborných panelů.

Data z Web of Science, oborově strukturovaná do *WoS categories fields* (WC), jsou tedy převedena do **oborové struktury dle OECD**: *Fields of Research and Development* (FORD). Oborová struktura OECD je užší (41 FORD), oborová struktura Web of Science je zhruba šestkrát větší (252 WC). Klastrování oborů do širších FORD kategorií poskytuje oborový pohled na obecnější úrovni a dovoluje zpracovávat rozumný a přehlednější počet vstupů/výstupů hodnocení a také ustavit racionálnější strukturu panelů i počet panelistů. V některých oborech však může dojít k systematickému znevýhodnění. Za tímto účelem je proto připraven doplňující datový podklad (viz níže) umožňující tuto okolnost v důležitých případech okomentovat.

Ukazatel AIS je v případě **multioborových časopisů** zahrnut do všech oborů, ke kterým jsou publikace afiliovány.[[4]](#footnote-4)

## Propojení databází IS VaVaI a Web of Science

**Propojení údajů** z RIV s daty Web of Science probíhalo primárně na základě unikátního kódu každé publikace (*accession number*). Protože toto pole bylo do konce roku 2017 v RIV nepovinné, byly výsledky dále napojovány pomocí ISSN/e-ISSN/ISBN a názvů.[[5]](#footnote-5) Tento postup je ovlivněný nekorektními záznamy v RIV. Při tvorbě podkladů byla proto řada chybných záznamů v matici opravena, avšak objem výsledků nedovoluje provést úplnou korekci dat. Chyba má náhodný charakter a pravděpodobně se týká zlomku analyzovaných dat.

Protože databáze Web of Science i Scopus jsou živými, průběžně aktualizovanými databázemi, část výsledků v RIV nemusela být propojena nebo doplněna o ukazatel AIS, přestože lze tyto údaje v současné chvíli individuálně dohledat. Platí proto, že data, která bude **možno propojit zpětně**, budou do analýzy zahrnuta v roce 2018.

## Spolupráce výzkumných organizací

Metodika 17+ u výsledků vzniklých ve spolupráci již nezohledňuje jednotlivé podíly. Pokud je analýza prováděna pro úroveň výzkumných organizací, pak je daný výsledek započten každé organizaci, která se podílela na tvorbě výsledku. Pro úroveň oborových analýz jsou naopak data deduplikována a každý výsledek do analýzy vstupuje pouze jednou.

## O čem předložené statistiky informují a o čem neinformují

Předložená data striktně vzato vypovídají pouze o distribuci národních článků v impaktovaných časopisech rozdělených podle jejich významu z hlediska citovanosti do zvolených pásem. O kvalitě jednotlivých článků nevypovídají. Nevypovídají ani o produktivitě oborů a výzkumných organizací, neboť to není předmětem této analýzy.

# Popis výstupů

## Oborové zprávy

**Tabulka O1a: doplňující údaje**

Z důvodů rozsahu je tento podklad dodáván v odděleném souboru ve formátu MS Excel a obsahuje čtyři listy.

První list s názvem **„WOS časopisy“** obsahuje seznam časopisů v daném oboru seřazených sestupně podle výše AIS. Doplňková tabulka ukazuje celosvětové počty článků a jejich podíly v kvartilech a v horním decilu.

Druhý list s názvem **„Národní výsledky“** obsahuje seznam národních výsledků příslušného oboru seřazených sestupně podle výše AIS. Výsledky, na nichž se podílelo více výzkumných organizací, jsou v seznamu zavedeny opakovaně, tedy vždy pro každou podílející se výzkumnou organizaci zvlášť a jsou označeny červeně.[[6]](#footnote-6) Pro účely výpočtů oborových statistik byly duplikace odstraněny.

Třetí list s názvem **„Neanalyzované“** je doplňkový a obsahuje souhrn výsledků, které byly propojeny s databází WoS, avšak nemají přidělen AIS, respektive daný časopis tímto ukazatelem nedisponoval.[[7]](#footnote-7) Další množinu tvoří výsledky, které nebyly propojeny například z důvodu změny ISSN časopisu nebo chybného záznamu v RIV. Všechny tyto výsledky budou znovu propojovány v následujícím roce.

Čtvrtý list **„FORD vs. WoS AIS“** slouží jako analytický podklad pro případné komentované zohlednění oborů, které mohou být významně znevýhodněny převodem mezi Web of Science categories a FORD. Data ve sloupcích jsou strukturována takto:

* WoS - seznam Web of Science categories spadajících do příslušného FORD;
* Max percentile – maximální pásmo ve FORD, na které dosáhne nejlepší časopis dané Web of Science category. Pásma jsou vyjádřena kvartily a horním decilem;[[8]](#footnote-8)
* Počet časopisů dané Web of Science category;
* Max AIS WC – maximální hodnota AIS;
* Median WC – medián příslušné Web of Science category;
* Median FORD – medián příslušného FORDu k porovnání;
* P 10 WC – doplňující údaj o spodní hranici horního decilu dané Web of Science category;
* P10 FORD – doplňující srovnávací údaj o spodní hranici horního decilu.

**Tabulky O1b a O1c: mezinárodní hranice prvního decilu a kvartilů podle pořadí časopisů a podle pořadí článků na základě AIS.; mezinárodní oborový medián dle pořadí časopisů a dle pořadí článků na základě AIS**

Pomocná přehledová tabulka O1b indikuje mezinárodní spodní hranice prvního decilu a kvartilů podle pořadí časopisů a podle pořadí článků. Tabulka O1c indikuje hodnoty světového mediánu oboru dle pořadí časopisů a dle pořadí článků. Standardně se pracuje se statistikami na úroveň časopisů. Avšak statistiky pracující s počtem článků zohledňují fakt, že v časopisech vychází odlišný počet textů, což dovoluje lepší srovnání s národními statistikami, které jsou založeny právě na jednotlivých článcích.[[9]](#footnote-9)

**Tabulka a graf O2a: rozčlenění národních výsledků v oboru v prvním decilu a v kvartilech podle mezinárodního pořadí časopisů**

Rozčlenění národních výsledků v oboru v prvním decilu a v kvartilech podle mezinárodního pořadí časopisů. Podíly tedy ukazují, kolik procent z národních oborových výsledků bylo publikováno v jednotlivých pásmech, pokud časopisy seřadíme podle jejich AIS. Připomínáme, že horní decil (P10) je součástí horního kvartilu a specificky vykazuje pásmo 10 % nejlepších časopisů.

**Graf O2b: srovnání podílů národních a mezinárodních výsledků (článků) oboru v prvním decilu a v kvartilech podle AIS časopisu**

Srovnání podílů národních a mezinárodních *výsledů* (tedy článků) v prvním decilu a v kvartilech pořadí časopisů. Distribuce *časopisů* je pochopitelně rovnoměrná (tj. ve všech kvartilech je 25 % časopisů, v horním decilu je 10 % časopisů). Nerovnoměrné rozložení článků souvisí s tím, že v časopisech vycházejí různé počty citovatelných textů (citable items). Graf tedy lépe vykresluje poměry na národní a mezinárodní úrovni z hlediska toho, jak se daří publikovat jednotlivé texty v časopisech rozdělených do pásem podle jejich významu, je-li hlediskem kvality citační odezva. Pro podíly národních časopisů je báze 100 % souhrn všech národních výsledků v oboru (FORD).

**Tabulka a graf O2c: porovnání národního a mezinárodního oborového mediánu podle pořadí AIS článků**

Porovnání národního a mezinárodního oborového mediánu podle pořadí článků a jejich vzájemný poměr. Údaj velmi hrubě zobrazuje „průměrnou“ výši AIS článků v daném oboru na národní a mezinárodní úrovni. Národní oborový medián je odvozen z řazení všech národních výsledků v oboru (FORD), mezinárodní oborový medián je odvozen z řazení všech citovatelných výsledků v databázi WoS v daném oboru (FORD). Poměr obou údajů je vyjádřen v procentech. Báze 100 % pro výpočet poměru je odvozen od mezinárodní úrovně mediánu.

**Tabulka a graf O3a, tabulka a graf O3b: seznam výzkumných organizací s největším počtem výsledků oboru v prvním decilu a kvartilu dle pořadí AIS časopisů**

Seznam výzkumných organizací s největším počtem výsledků v prvním decilu dle pořadí AIS časopisů. Bází pro vyjádření podílů institucí je souhrn všech národních výsledků oboru (FORD) v prvním decilu.

Statistiky jsou ovlivněny velikostí výzkumných organizací, respektive počtem jejich výzkumných pracovníků, infrastrukturou a výší prostředků. Dosud chybějící jednotná metodika pro vykazování těchto zaměstnanců nedovoluje provést odpovídající statistické srovnání. Komentář by tedy měl uvést prezentované statistiky do příslušného institucionálního kontextu oborů.

Obdobné údaje poskytují tabulka a graf O3b, avšak pro širší pásmo prvního kvartilu, které zahrnuje i údaje za první decil.

**Tabulky O3c: seznam všech národních výsledků (bez ohledu na obor) publikovaných v časopisech Nature a Science**

Publikování v časopisech Science a Nature je pro část oborů vysoce prestižní a zároveň se jedná o časopisy, které nemají stanovenou specifickou oborovou příslušnost. Tabulky představují seznam všech národních výsledků (bez ohledu na obor), které byly publikovány v těchto periodicích. Doplňující tabulka kumuluje výskyty výsledků na úroveň jednotlivých výzkumných organizací.

## Zprávy za skupiny oborů

Společná vizualizace vybraných statistik jednotlivých oborů dané skupiny má být podnětem pro vzájemnou diskusi a společnou reflexi. Vzhledem k rozdílům mezi obory z hlediska relevance bibliometrie pro jejich hodnocení není mechanické srovnání postačující.

**Tabulka a graf OS1a: porovnání mezinárodních a národních oborových mediánů dle pořadí AIS článků**

Porovnání mezinárodních a národních oborových mediánů dle pořadí AIS článků. Jedná se o statistiky vytvořené stejným způsobem jako u oborových zpráv (tabulka a graf O2c).

**Tabulka a grafy OS1b: mezioborové srovnání podílů národních a mezinárodních výsledků (článků) v prvním decilu a v kvartilech**

Rozčlenění národních a mezinárodních výsledků v prvním decilu a v kvartilech v jednotlivých oborech. Jedná se o statistiky vytvořené stejným způsobem jako u oborových zpráv (tabulka a graf O2a a O2b).

1. Každoročně zpracovávané bibliometrické zprávy mají charakter monitoringu, proto mají relativně jednoduchou podobu. Jednou za pět let, během kompletního hodnocení, budou použity sofistikovanější postupy. [↑](#footnote-ref-1)
2. Viz [www.eigenfactor.org/about.php](http://www.eigenfactor.org/about.php). [↑](#footnote-ref-2)
3. Pokud má soubor sudý počet prvků, mediánem je průměr dvou prostředních sousedících hodnot. [↑](#footnote-ref-3)
4. Stávající struktura RIV nedovoluje jednoznačně určit oborovou afiliaci jednotlivých výsledků dle nomenklatury WoS či OECD. V současné době je iniciována změna. [↑](#footnote-ref-4)
5. Vyplňovací povinnost pro databáze Scopus a Web of Science platí od 1. 1. 2017. [↑](#footnote-ref-5)
6. Některé faktické duplikace nejsou vyznačeny z důvodu nestejného záznamu v RIV. Strojové zpracování desítek tisíc záznamů nedovoluje provést kompletní ruční opravu datamatice. [↑](#footnote-ref-6)
7. Obvyklým důvodem je, že takový časopis zanikl, nebo existuje příliš krátkou dobu. Databáze Web of Science údaje doplňuje průběžně, proto je možné, že danému časopisu byl či bude ukazatel doplněn, avšak v datových podkladech dodaných Clarivate Analytics, vlastníka Web of Science, nebyl evidován. [↑](#footnote-ref-7)
8. Tzn., pokud je maximální percentil identifikován jako P10, pak obor dle členění Web of Science dosahuje na možná maxima příslušného FORD, do kterého byl převeden. Pokud je maximální percentil IV. kvartil, pak časopisy z takového oboru sice dosáhnou do pásma horního kvartilu, ale nemohou se zařadit mezi nejlepších 10 % časopisů atd. [↑](#footnote-ref-8)
9. Článkům je přiřazeno AIS časopisu, ve kterém jsou publikovány. Počet článků je odvozen od údajů WoS o počtu citovatelných textů uveřejněných v jednotlivých časopisech. [↑](#footnote-ref-9)