

Recenze ke studii Radima Bělohlávka „Hodnocení inforatického výzkumu“

Jiří Rákosník

Studie přináší kritický pohled na problematiku hodnocení kvality výzkumu v oblasti informatiky, ukazuje na úskalí a oborová specifika takového hodnocení obecně a v podmínkách probíhající reformy systému VaV v ČR zvláště. Zajímavé a podnětné jsou především vybrané konkrétní příklady dokumentující nevhodnost přístupu uplatňovaného v rámci tzv. metodiky hodnocení prosazované minulou Radou pro výzkum, vývoj a inovace, řada odkazů na literaturu včetně tzv. Campellova a Goodhartova zákona.

Dovolím si však také několik kritických připomínek a návrhů, kterými nechci užitečnost studie nijak zpochybňovat.

Obecně spatřuji určitý metodologický problém v tom, že autor své úvahy opírá o pojmy „lepší“ a „kvalita“ ve vztahu k výsledkům výzkumu a tyto pojmy zřejmě považuje za dané. V této souvislosti pak hovoří o tom, že systém hodnocení musí být „přesný“ a jeho výsledky musí „odpovídat skutečnosti“. Domnívám se, že měřit „přesnost“ pomocí subjektivního pojmu „lepší“ nelze a že právě to je jednou z příčin bouřlivých diskusí o způsobu hodnocení výsledků výzkumu. S tím souvisí i druhá námitka, totiž že autorovy úvahy naznačují, že stávající metodiku hodnocení lze úpravou parametrů a zpřesněním předmětových kategorií vylepšit tak, aby vyhovovala i specifickým podmínkám inforatického výzkumu. Domnívám se, že by ve studii mělo být výslovně řečeno, že je sice důležité hledat co neobjektivnější způsoby metrického vyhodnocování bibliografických výstupů, ale že to samo k odpovědnému hodnocení výsledků výzkumu nestačí.

Další připomínky a návrhy se týkají jednotlivostí.

Domnívám se, že výčet situací na str. 1, ve kterých se používá hodnocení kvality výzkumu, by bylo vhodné rozšířit o stanovení celkové částky na VaV (tj. stanovení velikosti investice do VaV).

Autor správně poukazuje na neblahý trend přenášení postupů hodnocení vyšších celků na nižší nebo i na jednotlivce (str. 1). Současná praxe v ČR však ukazuje i tendenci vyjadřovat hodnocení větších celků pomocí pouhého součtu hodnocení jejich jednotlivých součástí.

Tvrzení na str. 2, že potřeba kvalitního hodnocení výsledků výzkumu je obzvláště markantní ve společnostech, kde převážná část prostředků na výzkum pochází ze státního rozpočtu, což je údajně případ ČR, nepovažuji za správné. Kvalitně hodnotit výsledky výzkumu musí každý, kdo chce efektivně do výzkumu investovat, tedy i soukromník. Kromě toho oficiální statistiky ukazují, že podíl veřejných prostředků vkládaných do VaV v ČR je nižší než podíl prostředků soukromých.

V části „Databáze bibliometrických údajů“ na str. 4 autor navrhuje, že z pohledu informatiky je vhodné používat i další databáze kromě WoS. Není však jasné, jakým způsobem by se měly používat značně různorodé databáze nabízející různé podklady a služby. V této souvislosti by zde bylo vhodné zmínit i další problém, totiž zda vůbec daný časopis je v určité databázi sledován a jakým způsobem.

Zajímavé příklady v části „Různorodost informatiky, problém vymezení oblasti...“ se patrně vztahují k datům získaným z WoS. To by mělo být explicitně uvedeno. Ve větě uvozující první tabulku na str. 5 není zřejmé, zda se slovo „relevantní“ vztahuje k třídění danému v uvažované databázi nebo zda uvedené třídění považuje za relevantní autor.

Výraz „přesnost“ v části „Systém hodnocení komisí odborníků“ na str. 7 nepovažuji za zcela vhodný. Domnívám se, že ho zpochybňuje zmínka o možné zaujatosti hodnotitelů v první odrážce. (Obdobně k výrazu „nepřesnost“ v následující části.)

Nejsem si jist, že automatický bodový systém hodnocení je skutečně snadno realizovatelný a tedy levný (str. 7). Trpká zkušenost ze současných pokusů tento systém uplatnit v ČR ukazuje, že zajistit spolehlivost obrovského množství údajů v databázi výsledků vyžaduje časově velmi náročné úsilí na všech úrovních – od autora, přes jeho zaměstnavatele až po správce databáze.

Příklady na str. 9 jsou instruktivní. U druhého z nich jako příčinu prudkého nárůstu IF autor uvádí chování editora, který údajně vybízí autory k citování prací z téhož časopisu. To je závažné tvrzení a autor by měl jasně říci, zda jde o jeho dedukci nebo o prokázanou skutečnost (na což by se dalo soudit z komentáře ke třetímu příkladu).

Několik tvrzení v části 4 by si zasloužilo upřesnění: Metodika není dílem dalších poradních orgánů RVVI. Podle postupů RVVI podíl výzkumné organizace na celkovém počtu bodů dosažených výzkumnými organizacemi v ČR určuje podíl této organizace na celkovém objemu institucionálních prostředků přímo úměrně, což je důležité, protože jiná závislost by nemusela mít tak fatální důsledky. (Ostatně hodnocení výsledků a způsob rozdělování peněz jsou do velké míry nezávislé úlohy. Pouhá „revize“ metodiky hodnocení při zachování trojčlenkového způsobu rozdělování peněz nemůže vést k rozumným výsledkům.) Vystoupení T. Opatrného byla velmi významná, na nebezpečí nadprodukce „měkkých“ výsledků již několik let před tím upozorňovali jiní, v oficiálních dokumentech pro RVVI např. Akademie věd ČR.

Na str. 12 autor uvádí, že pro výpočty použil hodnoty z informačního systému MŠMT. Domnívám se, že jde o omyl, protože patrně má na mysli databázi RIV, kterou spravuje RVVI.

Nejsem si jist, že – jinak zajímavé a poučné – srovnání hodnocení tří zahraničních odborníků s jedním českým je zcela korektní. Není totiž zřejmé, zda zahraniční odborníci také nepublikovali nějaké podprůměrné nebo průměrné výsledky ve sbornících konferencí, o kterých nevíme proto, že se jejich výsledky nezachycují v databázi typu RIV podle české metodiky.

Není mi jasné, jaký závěr by měl vyplynout z autorova konstatování za otázkou č. 2 na str. 12, že „na vině je nevhodné bodové hodnocení článků ve sbornících konferencí“. Jsem přesvědčen, že žádné vylepšení bodového hodnocení neumožní jedním číselným údajem srovnat kvalitu odborníků v informatice a fyzice. To ostatně dobře dokumentuje tragická zkušenost z pokusu „vylepšit“ vadnou metodiku v nové verzi 2010.

Přestože se autor ve své studii zaměřuje na informatiku, jeho šest závěrů uvedených na konci má platnost i pro jiné disciplíny – závěry 1–3 zčásti, závěry 4–6 zcela obecně.

Přes uvedené připomínky považuji studii za užitečný příspěvek k diskusi o hodnocení výsledků (nejen) inženýrského výzkumu, který po upřesnění a doplnění může být podkladem další odborné práce na návrhu objektivní metodiky hodnocení. Je však třeba mít na paměti, že by se toto úsilí nemělo omezit na pouhou revizi současného zcela nevhodného přístupu.