

Metody hodnocení VaV

Jiří Vaněček

Technologické centrum AV ČR

Hodnocení - důležitý strategický nástroj používaný při řízení výzkumu a vývoje (VaV).

Výsledky hodnocení slouží jako podklad pro řadu úkonů a kroků:

- Alokace veřejných finančních prostředků na podporu výzkumu
- Východiska pro tvorbu vědní politiky
- Návrhy prioritních směrů výzkumu
- Reformy systému VaV
- Reorganizace institucí VaV

Význam hodnocení v celém světě roste:

- Stupňuje se tlak společnosti na sociální zdůvodnitelnost výzkumu
- Omezenost disponibilních finančních zdrojů
- Sílí odpovědnost za účelně a hospodárně vynaložené veřejné prostředky

Typy hodnocení podle předmětu:

- Hodnocení projektů a programů (ex ante, interim a ex post)
- Hodnocení výzkumných jednotek:
 - jednotlivců a laboratoří (úroveň mikro)
 - výzkumných institucí (meso)
 - celých oborů či národních vědních systémů (makro)

Z hlediska obsahu:

- Hodnocení vstupů (finanční zdroje na VaV, lidské zdroje)
- Hodnocení procesů a struktury (výzkumná strategie, spolupráce, vědecká výchova, transfer znalostí, infrastruktura)
- Hodnocení výsledků (publikační a aplikační výsledky)
- Hodnocení dopadů na vědeckou komunitu a na společnost (citovanost publikovaných výsledků, jejich využití v různých oblastech: ekonomika, zdravotnictví, životní prostředí, řízení státu, kultura)

Podle účelu ke kterému má hodnocení sloužit:

- Hodnocení **sumativní** – důraz na analytické aspekty, hlavně popis současného stavu a kvality VaV, nejdůležitější je hodnocení výsledků,
- Hodnocení **formativní** - zaměřeno více na proces zlepšování a učení, zahrnuje zpětnou vazbu mezi hodnotitelem a hodnoceným, nejdůležitější součástí je hodnocení procesů a struktury

Podle formy hodnocení lze přístupy rozlišit na kvalitativní a kvantitativní:

- **Kvalitativní** – hodnocení oborovými panely složenými z expertů (***peer review***) případně doplněný o názor manažerů a uživatelů VaV, průmyslníků, ekonomů, politiků (**modifikovaná *peer review***)
- **Kvantitativní** – hodnocení pomocí indikátorů, které charakterizují **strukturu a procesy, vstupy** či **výstupy** instituce v různých oblastech činnosti
- **Informovaná *peer review*** - kombinuje expertní a indikátorové hodnocení. Panely hodnotí na základě informací získaných ze souboru indikátorů, které charakterizují všechny důležité oblasti a funkce hodnocené instituce.

Kvalitativní hodnocení - zásady

- **Odbornost**
 - založena na znalostech jednotlivých odborníků (hodnotitelů)
 - rozsah současného vědění neumožňuje hluboké znalosti jednotlivce v širokém oborovém spektru
 - nutný značný počet oborově specializovaných panelů
 - obsahující odborníky na všechny hlavní výzkumné směry zastoupené v hodnocených institucích
 - příklad RAE 2008: více než 60 panelů po 10-15 členech
- **Objektivita**
 - alespoň dva nezávislí hodnotitelé pro každý výsledek
 - bez vztahu k hodnocené instituci
 - s dobrou znalostí hodnocené problematiky
 - malé země používají vesměs zahraniční odborníky
- **Dostatek informačních zdrojů** o hodnocených institucích
 - výkonný pomocný tým
 - dobrá příprava hodnocení
- **Dostatek času** pro zpracování hodnocení
- **Infomovaná *peer review***
 - nevyžaduje tolik oborově specializovaných panelů
 - snižuje náklady i časovou zátěž hodnocení

Kvantitativní hodnocení

- Využívá statistická data tj. indikátory, které charakterizují především velikost vstupů a počet výsledků či aktivit
- V některých případech mohou charakterizovat i jejich kvalitu

Indikátory lze rozlišit na jednoduché (prvního řádu) a komplexní (druhého řádu):

- **Jednoduché indikátory** měří jednotlivé vstupy, procesy, či výsledky VaV
- **Komplexní indikátory** jsou z nich odvozeny a charakterizují účinky výsledků v podobě sumarizujícího indexu (např. Impakt faktor časopisu, Relativní citační index, nebo H-index)
- V odborné literatuře jsou někdy považována pořadí či bodová hodnocení udělená na základě hodnocení oborovými panely expertů za **indikátory třetího řádu**

Indikátory používané pro hodnocení institucí VaV

- **Indikátory vstupů:**

- podíl externího financování (z grantů, zakázek atd.)
- počet a finanční objem projektových grantů
- počet zaměstnanců či výzkumných pracovníků
- počet studentů v různých typech studia (magisterském, doktorském)

- **Indikátory procesu či struktury (výzkumné prostředí)**

- vybavení a infrastruktura instituce
- struktura lidských zdrojů
- mezinárodní mobilita a spolupráce
- členství ve vědeckých společnostech
- zvané přednášky

- **Indikátory výsledků VaV**

- počet publikací
- počet patentů
- další nepublikační (aplikační) výsledky
- počet absolventů
- obhájené magisterské a PhD tituly

- **Indikátory dopadů na vědeckou komunitu a na společnost (socio-ekonomické dopady)**

- citovanost publikací
- vědecká prestiž instituce
- udělené ceny a uznání
- transfer technologií
- příjmy z licencí
- vznik nových firem
- zaměstnanost či úspěšnost absolventů

Komplexní indikátory

Hodnocení časopisů

- **Impakt faktor časopisu (IF)** – počet citací v běžném roce na všechny články časopisu otištěné v předchozích 2 letech dělené počtem těchto publikací – **Průměrná citovanost publikací daného časopisu**
- **Source Normalised Impact per Paper (SNIP)** – počet citací v běžném roce na všechny články časopisu otištěné v předchozích 3 letech dělené počtem těchto publikací. Vážení citací podle průměrného počtu citací uváděných v citujícím časopisu - **Korekce na citační potenciál**
- **SCImago Journal Rank (SJR)** – počet citací v běžném roce na všechny články časopisu otištěné v předchozích 3 letech dělené počtem těchto publikací. Vážení citací podle významu citujícího časopisu - **Korekce na kvalitu citace**

Hodnocení souborů publikací

- **Relativní citační index (RCI)** – průměrná citovanost publikací hodnoceného subjektu dělená průměrnou citovaností všech publikací v databázi (ve světě)
- **Relativní citační index oboru (RCIO)** – průměrná citovanost publikací hodnoceného subjektu normalizovaná (dělená) podle oborových standardů
- **H-index:** h = počet publikací hodnoceného subjektu, které jsou každá citovány alespoň h -krát.

Kvantitativní hodnocení institucí VaV

- Mělo by vycházet ze všech 4 typů indikátorů
 - vstupy, prostředí, výsledky, dopady
 - váha jednotlivých indikátorů v celkovém hodnocení je různá (sumativní a formativní hodnocení)
 - váhy se liší i podle typu hodnocené instituce a oboru
- Institucionální specifika:
 - instituce základního výzkumu (publikace, citovanost)
 - instituce průmyslového výzkumu (aplikace, použití)
 - infrastrukturní a servisní pracoviště (využití)
- Oborová specifika:
 - druhy výsledků
 - počty výsledků

Oborové rozdíly: Druhy výsledků v RIV 2008 (procent ČR celkem)

	článek	sborník	kniha	patent
Společenské vědy	23,8	30,9	71,0	0,0
Informatika	1,0	3,6	1,6	0,0
Průmysl	11,6	32,6	9,1	36,5
Fyzika, matematika, chemie, vědy o zemi	27,3	23,3	6,4	32,9
Biovědy, Lékařské vědy, Zemědělství	36,1	9,2	11,0	30,6

Oborové rozdíly: Počet publikací ČR v roce 2008

	Počet publikací 2008		
	ČR - SCOPUS	Svět - SCOPUS	Podíl ČR (%)
Engineering	2172	441565	0,49
Mathematics	1175	138127	0,85
Medicine	3103	614609	0,50
Neuroscience	224	50918	0,44
Biochemistry, genetics, molecular biology	1690	238127	0,71

Oborové rozdíly: Citovanost publikací ČR z roku 2008

	Počet citací na publikaci		RCIO
	ČR - SCOPUS	Svět - SCOPUS	Podíl ČR (%)
Engineering	4,29	4,7	91,28
Mathematics	5,24	5,45	96,15
Medicine	11,16	11,06	100,90
Neuroscience	14,94	18,54	80,58
Biochemistry, genetics, molecular biology	16,25	18,78	86,53

Výhody a nevýhody obou forem hodnocení:

- **Kvalitativní**

- Nevýhody – hodnocení oborovými panely je drahé, pomalé, náročné na čas hodnocených institucí i hodnotitelů. Hodnotitelé mohou být předem zaujatí a ne zcela objektivní (některé panely hodnotí přísněji než jiné). Škála hodnocení je často redukována. Nelze zpětně kontrolovat celý proces (audit).
- Výhody - oborové panely přirozeně dodržují oborová specifika, dobře hodnotí kvalitu výzkumu i dopady, lehce odhalí „gaming“ (podvádění), mají větší důvěru vědecké veřejnosti
- Důsledky – hodnocení probíhá jednou za 4 až 7 let, hodnocena kvalita pouze u vybraných výsledků, dvou-stupňové řízení (kontinuita a jednotné měřítko). Některé nevýhody lze odstranit pomocí informované peer review. Země používající kvalitativní hodnocení často rozdělují podle jeho výsledků významný podíl podpory VaV.

- **Kvantitativní**

- Nevýhody – indikátory mají omezenou schopnost ohodnotit kvalitu výsledků a jejich využitelnost (impakt), zvláště v některých oborech. Citlivé na podvádění („gaming“).
- Výhody – dobře hodnotí kvantitativní aspekty, využívá plnou škálu hodnocení, je laciné, objektivní a konzistentní. Lze zpětně kontrolovat celý proces (audit).
- Důsledky – hodnocení probíhá každý rok, hodnoceny bývají všechny výsledky. Je nezbytné předem zohlednit oborová specifika. Na základě hodnocení se zpravidla rozděluje pouze malý podíl podpory VaV.

Zásady platné pro oba typy hodnocení

Transparentnost:

- v postupech hodnocení - předem zveřejnit způsob sestavování hodnocení, váhy jednotlivých faktorů
- využití výsledků hodnocení - zveřejnit výsledky hodnocení, způsob, jak se k nim došlo a důsledky, které výsledky hodnocení budou mít

Volba optimální metody podle předmětu hodnocení:

- instituce základního výzkumu - podle kritérií vědecké kvality, tj. z hlediska originality, inovativnosti a impaktu na vědní obor. V některých vědách je hodnocen též kulturní, sociální a ekonomický dopad.
- instituce aplikovaného výzkumu - z hlediska významu pro společnost (relevance). souhlas mezi plánovanými cíli a dosaženými výsledky (*effectiveness*) a poměr mezi dosaženými cíli a vynaloženými prostředky (*efficiency*).

Volba metody podle úrovně agregace hodnoceného systému:

- jednotlivci či malé výzkumné skupiny (úroveň „mikro“) – spíše kvalitativní
- jednotlivé instituce VaV (úroveň „meso“) - značné mezinárodní rozdíly
- celé sektory VaV (vysoké školy, vládní sektor a podnikový sektor) či celé státy (úroveň „makro“) – indikátory, bibliometrie a benchmarking

Mezinárodní trendy v přístupu k hodnocení VaV:

- Rychlý růst počtu a významu hodnocení, důvody (financování, organizační rozhodnutí)
- Požadavky na snížení finanční a časové náročnosti hodnocení jak na straně hodnocených, tak na straně hodnotitelů
 - široké využívání indikátorů
- Vysoké nároky na profesionalitu (odbornost) a důvěryhodnost hodnocení ve všech oborech a typech institucí
 - důraz na peer-review hodnocení

Módní vlny v hodnocení institucí:

- **První vlna** v osmdesátých letech 20. století – kvalitativní hodnocení. Velká Británie (RAE) - první a vzorové hodnocení VŠ, ale nákladné a časově náročné .
- **Druhá vlna** asi o 10 let později – kvantitativní hodnocení VŠ pomocí indikátorů. Austrálie – po zavedení hodnocení růst počtu publikací, ale pozdější důkladná analýza ukázala pokles kvality: fragmentace publikací, méně náročné časopisy, pokles citovanosti.
- **Třetí vlna** – po roce 2000
 - kombinace indikátorů a peer review - informovaná peer review (NL) nebo kombinované hodnocení (NZ, AU)
 - hodnocení aplikačních výsledků (třetí role universit)
 - v posledních letech i hodnocení socio-ekonomických dopadů (impaktu)

Britské RAE

- organizováno centrálně pro celou Británii od roku 1986, poslední 2008
- hodnocení kvality výzkumu na univerzitách
- probíhá odděleně od hodnocení úrovně výuky
- organizuje *Higher Education Funding Council for England* (HEFCE)
- hodnocení slouží jako klíč pro rozdělení institucionální podpory na VaV univerzitám
- od počátku koncipováno jako kvalitativní, využívající oborové panely expertů
- v mnoha dalších aspektech prodělalo značný vývoj
- zpočátku hodnoceny všechny výsledky, frekvence 3 roky
- později však jejich počet omezen na maximálně 4 výsledky, interval prodloužen na 5 let a nakonec až na 7 let
- mezinárodní vzor a dobrá praxe

Kritická meta-evaluace RAE v roce 2001

- RAE shledáno velmi nákladné a časově náročné jak pro hodnocené organizace, které připravují podklady, tak i pro hodnotitelské panely, kterých je velký počet
- malé rozdíly v hodnocení:
 - 80 % hodnocených univerzit jedno ze tří nejvyšších hodnocení (tj. známka 4, 5, nebo 5*)
 - více než polovina univerzit hodnocena dokonce nejvyšší známkou (5, nebo 5*).
- nedostatek finančních prostředků na významné navýšení podpory výborným univerzitám
- navržena kontinuální škála hodnocení (profily) - těsnější vazba mezi hodnocením a financováním
- více využívat pro *peer review* zahraniční experty
- doporučeno prodloužení intervalu hodnocení na 7 let.

V roce 2006 - nová britská vláda

- Rozhodnutí o zrušení RAE
- Náhrada převážně kvantitativním hodnocením založeným na indikátorech, bibliometrii a benchmarkingu
- V důsledku nedostatku času RAE 2008 proběhlo v původní podobě, ale s dílčími úpravami
 - kvalitativní profil instituce, určující podíl výsledků instituce patřících do jednotlivých kvalitativních kategorií
 - paralelně s hodnocením oborovými komisemi i hodnocení kvantitativní
 - výsledky vzájemně porovnány
- Ve většině oborů koreluje bibliometrické hodnocení s hodnocení panelů RAE velmi dobře.
- Nedostatečná korelace
 - humanitních a některých sociálních vědách
 - inženýrských oborech
 - počítačových oborech
 - v klinické medicíně korelovala vyšší citovanost publikací pouze s nejvyšší kvalitativní kategorií hodnocení RAE
- Nedostatečná korelace hodnocení a výhrady vůči kvantitativnímu hodnocení VaV v britské vědecké komunitě
- Nové hodnocení založeno na expertních panelech

Research Evaluation Framework (REF) 2014

Hodnocení kvality výsledků

- v exaktních oborech lze využít citační analýzu a případně i další indikátory
- IF časopisů nelze použít
- 3 kritéria hodnocení:
 - originalita výsledku
 - jeho význam pro obor
 - průkaznost zjištění

Hodnocení impaktu

- pouze uskutečněné dopady uvedené v případových studiích (*case studies*)
- hodnocení dopadů v různých oblastech: zdraví, společnost, kulturu, ekonomiku, politiku a služby, produktivitu, prostředí a mezinárodní vztahy
- hodnocena 2 kritéria:
 - dosah (šíře) vlivu
 - významnost vlivu

Hodnocení výzkumného prostředí

- výzkumná strategie instituce
- lidské zdroje - odbornost zaměstnanců, jejich spolupráce a postavení v oboru (obdržené ceny, členství v redakčních radách či expertních panelech atd.) a počet PhD studentů
- infrastruktura
- finanční zabezpečení instituce - výše příjmů na VaV a příjmy z prestižních grantů
- dvě kritéria hodnocení:
 - vitalita (tj. intenzita)
 - udržitelnost do budoucnosti

REF 2014 - Organizace hodnocení

- Hodnocení je kvalitativní, provádí jej oborové panely složené z expertů
- Do hodnocení REF je možné přihlásit maximálně 4 výsledky na každého výzkumného pracovníka:
 - různé druhy publikací (články v recenzovaných časopisech, knihy, kapitoly, monografie, konferenční články atd.)
 - různé aplikace (např. patenty, užité vzory či software, ale i nové materiály, návrhy staveb či strojů nebo expozice v závislosti na daném oboru)
- Hodnocení má 5 stupňů (známek), od podprůměrných (*unclassified*) po světovou třídu (*world-leading*)
- Výsledkem hodnocení jsou profily (nikoliv průměrná známka) vyjadřující kolik procent výsledků, dopadů či prostředí v každé instituci odpovídá danému stupni (známce) hodnocení.
- Oborová kritéria jsou předem pevně určená a zveřejněná.
- Toto hodnocení bude tvořit podklad pro rozdělování institucionální podpory od roku 2015.

REF 2014 – Hodnotící oborové panely

- Hodnocení provádí 36 pod-oborových panelů pracujících pod vedením 4 hlavních oborových panelů
- Pod-oborové panely jsou složeny z 20-30 expertů
- Hlavní panely mají asi po 20 členech
- V hodnocení REF je celkem zapojeno asi 1000 expertů
- Předsedy sub-panelů jsou členové hlavního oborového panelu, čímž je zajištěna kontinuita
- Hodnocení dvoustupňové, sub-panely zajišťují specifickou oborovou expertízu, hlavní panely jednotné měřítko hodnocení
- Hlavní oborové panely
 - specifická oborová kritéria platná pro všechny pod-obory
 - dohlíží na dodržování těchto kritérií v jednotlivých sub-panelech
 - zajišťují jednotnou náročnost hodnocení
 - schvalují konečné výsledky hodnocení (tj. profily institucí)
- Sub-panely
 - provádějí vlastní hodnocení podle schválených kritérií a postupů
 - vytváří hodnotící sub-profil pro každou instituci (kvalita výsledků, impakt, výzkumné prostředí)
 - každý výsledek hodnotí alespoň 2 členové sub-panelu

Hodnocení VaV v Austrálii

1992 – Composite Index: hodnocení VŠ podle jednoduchých indikátorů

- podíl financování z projektových grantů
- podíl publikací v časopisech vedených ve WOS
- podíl absolventů PhD a MS studia

1995 – Podle výsledků hodnocení rozdělováno universitám 7% institucionální podpory VaV

2003 – L. Butler: Analýza dopadů

- rychlý růst počtu publikací
- pokles citovanosti publikací
- časopisy s nižším IF, fragmentace

Příprava Research Quality Framework (RQF) – informovaná peer review

- pokročilé indikátory
- důraz na excelenci a impakt
- značná složitost hodnocení

2008 – zrušena příprava RQF, nahrazena jednodušším hodnocením Excellence in Research for Australia (ERA)

Excellence in Research for Australia (ERA)

- Centrální řízení
 - z pověření ministerstva pro VaV - *Department of Innovations, Industry, Science and Research* (DIISR)
 - organizuje *Australian Research Council* (ARC) ve spolupráci s *National Health and Medical Research Council* (NHMRC)
- Účelem hodnocení je
 - identifikovat excelentní australský výzkum ve všech oborech
 - porovnat výzkum na australských univerzitách s mezinárodními standardy
 - vytvořit nové pobídky pro kvalitní výzkum
 - podpořit spolupráci mezi institucemi VaV navzájem
 - podpořit spolupráci mezi institucemi VaV a koncovými uživateli
- V roce 2009 - předběžné testovací kolo ve 2 oborových klastrech
- V roce 2010 a 2012- regulérní hodnocení ve všech oborech

ERA 2012

- Kombinované hodnocení pomocí indikátorů a expertních názorů mezinárodních i národních odborníků
- Hodnocení podle 4 hledisek:
 - intenzita výzkumu
 - kvalita výzkumu
 - excellence v aplikovaném výzkumu
 - uznání ve vědecké komunitě (vědecká reputace)
- Klasifikace institucí do stupňů 1 až 5 (nejvyšší) a nehodnocen (tj. nedostatečný objem výsledků pro hodnocení)
- V počátečních fázích nemá ERA vliv na rozdělení institucionálních prostředků
- Po zavedení systému se podle výsledků hodnocení bude rozdělovat podpora z programu Sustainable Research Excellence in Universities (SRE)

Intenzita VaV

- Počet výsledků VaV,
- Počet a objem získaných kompetitivních projektových grantů (klasifikace do 5 kategorií podle typu a míry konkurence),
- Příjmy z kontraktů a ze zakázkového výzkumu,
- Počet studentů ve vědecké výchově
- Počet absolventských titulů (Masters či PhD) udělených za poslední 3 roky
- Hodnoty vztaženy (normalizovány) na počty zaměstnanců instituce (FTE)

Kvalita výzkumu

- Analyzována kvalita publikací vydaných během posledních 6 let
- Stanovena procesem *peer review* na základě indikátorů
- Podklad citační analýza porovnání se světovými standardy v daném oboru (SCOPUS)
 - průměrná citovanost publikací
 - percentilová analýzu citovanosti publikací
- V oborech, jejichž publikační média nejsou ve SCOPUS dostatečně zastoupena nelze citační analýzu použít (např. humanitní vědy a sociálních vědy)
 - publikační profil: publikační média (časopisy, knihy, sborníky) jsou předem klasifikována do 4 kategorií
 - posouzení publikací externími experty (distanční *peer review*)

Excelence v aplikovaném výzkumu

- počet patentů
- počet registrovaných nových odrůd či plemen
- počet registrovaných užitných vzorů
- počet léčebných procedur potvrzených *National Health and Medical Research Council*
- suma příjmů z licenčních poplatků a dalších komercializací

Vědecká reputace je hodnocena podle *esteem* indikátorů

- vědecké ceny
- zvané přednášky
- členství v prestižních radách
- členství ve statutárních orgánech
- členství v učených společnostech
- editorství prestižního díla

ERA 2012 - rozčleněna poměrně hrubě do 8 rámcových oborových klastrů:

- Fyzikální, chemické a geologické vědy
- Humanitní vědy a tvůrčí umění
- Inženýrské vědy a environmentální vědy
- Vzdělávání a lidská společnost
- Ekonomika a obchod
- Matematika, informatika a komunikační vědy
- Biologické vědy a biotechnologie
- Lékařské vědy a zdravotnictví

Dodržována oborová specifika, srovnání uvnitř oborů

Hodnocení 2012

- hodnocen obor v instituci, odpovídá úrovni oddělení, katedry či fakulty
- výsledky z období 2005-2010
- citace z let 2005-2011
- příjmy, aplikace a *esteem* indikátory z let 2008-2010

Zdroj informací o publikacích

- elektronické archivy universit – úspora času

Důraz na metriku

- indikátory hodnocení jsou oborově specifické
- dovoluje sestavit omezený počet komisí
- omezení množství práce vyžadované od komisí
- úspora finančních prostředků

Expertní složka

- koriguje zjednodušené závěry vyplývající z indikátorů („gaming“)

Kombinovaný systém hodnocení

- transparentní a rychlý
- levnější a jednodušší než klasická peer review
- hodnocení institucí v kratších intervalech
- hodnocení posuzuje všechny výsledky, nejen vybrané

Děkuji vám za pozornost