

## Hodnotíme vědu profesionálně?

P. Ráb<sup>1</sup>, I. Kadlecová<sup>2</sup>

<sup>1</sup>Akademická rada AV ČR, <sup>2</sup>KNAV AV ČR v.v.i.

[rab@iapg.cas.cz](mailto:rab@iapg.cas.cz)

INFORUM 2008: 14. konference o profesionálních informačních zdrojích  
Praha. 28. – 30.5.2008

### Abstrakt

Věda je nejmocnější nástroj lidské civilizace a prostředky do ní vložené jsou jednou z nejvýhodnějších investic vůbec. Proto je nesmírně důležité správně alokovat dostupné zdroje prostředků a identifikovat ty oblasti vědy, organizace, skupiny a jednotlivce ve vědě a výzkumu, které jsou v systému nejvíce přínosné. Hodnocení výkonu ve vědě a výzkumu je proto klíčový moment. Takové hodnocení je však vzhledem k povaze vědy metodicky velmi složitá záležitost. Výkon ve vědě a výzkumu má vždy dvě složky (zastoupené v různém poměru): kvalitativní (= co bylo vykonáno) a kvantitativní (= kolik bylo vykonáno). Zatímco hodnocení první složky je nutno provádět peer review systémem, pro druhou složku jsou široce využívány databáze vědeckých informací a indikátory, které z nich lze odvodit. Tyto údaje (indikátory) jsou vždy kostrou hodnotících systémů a navíc jsou často používány bez zhodnocení kvalitativní složky výkonu. Správnost a úplnost údajů v takových databázích jsou tudíž nezbytným předpokladem pro hodnocení ve vědě a výzkumu. Příspěvek se zabývá některými problémy použití databází a indikátorů v hodnocení vědy a výzkumu.

### Proč hodnotit vědu a výzkum?<sup>1</sup>

Věda je nejmocnější nástroj lidské civilizace a prostředky do ní vložené jsou jednou z nejvýhodnějších investic vůbec. Vyspělé společnosti očekávají, že veřejné prostředky vynakládané na V&V přinesou jednak celospolečenský zisk („social and societal return“<sup>2</sup>) ve formě rozvoje vzdělanostní společnosti („tacit knowledge“), jednak v materiálním rozvoji společnosti a ve vyspělých zemích je V&V oblastí výhodných a strategických investic, zejména proto, že v moderní společnosti se stírají hranice mezi V&V a pokročilými

---

<sup>1</sup> Poznámka k termínům **věda a výzkum** – v současně platném zák. 130/2002 i jeho připravované novele se slovo věda nevyskytuje, zkratka v těchto zákonných normách VaV znamená **výzkum a vývoj**, v novele zák. 130 se navíc objevuje ještě termín inovace, tedy VVaI – **výzkum, vývoj, inovace**. Je pozoruhodné, že se neobjevuje oficiálně termín **věda** (a to ani v názvu vrcholného státního orgánu řídicího celou oblast – Rada pro výzkum, vývoj a inovace) a není pochyb, že tento symptomatický jev je projevem zcela základního nepochopení úlože vědy v moderní demokratické společnosti. Srov. např. komentář o stavu britské vědy v Nature 447, 7142 ze 17.5. 2007 na str. 231!

<sup>2</sup> Pro jistotu uvádíme (snad) odpovídající anglické termíny z evaluační teorie.

technologemi. Je dobře dokumentována skutečnost, že existuje korelace mezi výší podpory V&V a produktivitou ekonomik<sup>3</sup> Proto je nesmírně důležité správně alokovat dostupné zdroje prostředků a identifikovat ty oblasti vědy, výzkumu, organizace, projekty, programy, skupiny a jednotlivce ve vědě a výzkumu, které jsou rozvoji společnosti nejvíce přínosné. Hodnocení výkonu ve vědě a výzkumu je proto klíčový moment. Cílem hodnocení V&V je potom stimulace kvalitního V&V tak, aby společnosti přinášel zejména ekonomický rozvoj (na který klade důraz např. britská evaluační škola) a vzdělanostní rozvoj (na který klade důraz např. skandinávská evaluační škola). Lze však říci, že při kvalitním hodnocení V&V dochází ve vzájemných složitých souvislostech k rozvoji obou těchto aspektů. Bez kvalitního a metodicky správného hodnocení V&V jsou potom vložené prostředky neúplně využity či přímo ztraceny.

### Jak hodnotit vědu a výzkum?

Hodnocení V&V je však vzhledem k povaze vědy a výzkumu (ve skutečnosti se jedná o jednu z nejpestřejších oblastí lidských aktivit vůbec!) metodicky velmi složitá záležitost a hodnocení V&V se v poslední době stává samostatným sociologickým vědeckým oborem<sup>4</sup>. V zásadě a stručně však lze říci, že výkon ve V&V má vždy dvě složky (zastoupené v různém poměru): kvalitativní (= co bylo vykonáno) a kvantitativní (= kolik bylo vykonáno). V kvalitních evaluačních systémech a s ohledem na stanovené cíle a objekty (tj. instituce, organizace, projekty, programy, skupiny, jednotlivci) hodnocení se pak tyto dvě základní složky vědeckého výkonu navzájem prolínají a mají různé zastoupení. Evaluační systémy mají zejména následující tři typy metodického postupu<sup>5</sup>.

*Ranking* – hodnocení se soustředěním na hodnocení výstupů podle stejných měřítek, hodnocení neobsahuje žádná doporučení, není orientováno na postupy v organizaci, nehodnotí programy podpory, nebere v úvahu rozdílnost objektů (které bývají seskupeny náhodně, např. VVŠ a v.v.i. v ČR), kritéria výkonnosti spočívají v systému známek a výsledky měření jsou shrnuty do seznamu pořadí. Komprimace informací je hlavní funkcí rankingu, slouží ke zpřístupnění informací o větším počtu heterogenních organizací a jejich seskupení podle přijatých kritérií, zpravidla jde o organizace, instituce které si navzájem konkurují. Hlavní výhodou je transparentnost, nevýhodou je problém nastavení známek

---

<sup>3</sup> OECD National accounts (1980-1998), 2001

<sup>4</sup> Jako příklad jmenujeme specializované konference New Frontiers in Evaluation (Víděň, 2006 - <http://www.fteval.at/conference06/>) nebo Peer Review (Praha, 2006 - <http://www.pragueforscience.cz/>)

<sup>5</sup> Následující termíny nemají český ekvivalent

*Benchmarking* – se liší od předešlého kompletním podchycením různých kritérií a informací (nepřesně: vědecká a výzkumná výkonnost = články, přednášky, knihy, citace, aplikace, postavení - „viditelnost“ v oboru a ambice = světová, evropská, národní, lokální úroveň, koherence aktivit = koncentrace na problematiku a jejich výběr, dynamika – schopnost přizpůsobovat se změnám v odborných tématech a v jejich řízení) a porovnáním s cíleně vybranou, výkonnou a vysoce kvalitní referenční organizací (případně institucí, projektem, programem, pracovní skupinou nebo i jednotlivcem)

*Rating* – určuje míru rizikovosti nesplnění závazků, ve VaV se jedná poměr tzv. bezpečného (u nás typicky s výsledky dodatečnými do IS VaV) a rizikového (s možností nesplnění programu nebo projektu) VaV; u vědy musí být podíl rizikového výzkumu vyšší. Celková koncepce ratingu ve VaV má následující komponenty: řídicí skupina ustavuje hodnotící skupiny, které hodnotí organizace podle příslušného oboru v základních dimenzích (vědecký výkon, transfer poznatků, výchova dorostu, atd.) podle oborově konkretizovaných kritérií. Hodnocení rizikových (tj. s možností nedosažení očekávaných výsledků a tedy ztráty vložených prostředků) oblastí a aktivit VaV je jednou z nejkomplicovanějších oblastí evaluační nauky<sup>6</sup>

Zatímco hodnocení kvalitativní složky výkonu je nutno provádět peer review systémem, pro kvantitativní složku jsou široce využívány databáze vědeckých informací a indikátory, které z nich lze odvodit. Peer review systém, tedy panely odborníků a nezávislých posuzovatelů, bývá nejčastěji, a to nejen u nás, kritizován pro právě pro nezastupitelnou úlohu „lidského faktoru“ v průběhu hodnocení, který může být nejslabším (nechtěná či záměrná favorizace, rozhodnutí v rozporu s dobrými mravy a morálními standardy, atd.) článkem evaluačního systému. Toto slabé místo však i) souvisí s celkovou morální vyspělostí dané společnosti a ii) existují evaluační postupy, jež kontrolují práci odborníků a nezávislých posuzovatelů a odstraňují případné problémy. V každém případě má peer review systém svoje omezení, které je nejlépe vystiženo citátem „Peer review assumes honesty and is based on trust. It can help establish scientific validity and rigor but it offers no guarantees“<sup>7</sup>. Naproti tomu databáze vědeckých informací a případné z nich odvoditelné indikátory nejsou (nebo by neměly být pokud neuvažujeme „lidský faktor“, tj. chyb učiněné vkladateli dat) ovlivněny subjektivně a proto tyto soubory vědeckých a odborných informací jsou téměř vždy kostrou a základem hodnotících systémů. Jejich předností je přehlednost, snadnost komprimace informací, číselná

<sup>6</sup> Např. Peer Review Workshop (Haague, 2008 - <http://www.esf.org/activities/mo-fora/peer-review/peer-review-workshop-in-the-hague.html>)

<sup>7</sup> Peer Review (Praha, 2006 - <http://www.pragueforscience.cz/>) z přednášky editora Science „Limits of peer review“

vyjádření. Závažným nedostatkem je to, že jsou často používány bez zhodnocení kvalitativní složky vědeckého a odborného výkonu a navíc výsledky analýz podávají v absolutních škálách. Dále jsou celé skupiny oborů nebo vědeckých a odborných aktivit, u kterých lze kvantitativní složky hodnocení uplatnit je velmi obtížně. Velkým nebezpečím toho „indikátorového“ přístupu je snadné záměrné zneužití, ale zejména nezáměrné (zejména pro nepochopení, jak vlastně tvůrčí procesy ve vědě a výzkumu probíhají a jak se výsledky zveřejňují, čili předávají odborné komunitě) aplikace při tvorbě vědní politiky, jež vede k závažným deformacím v celé oblasti VaV. Např. problém s inflační „nadprodukcí“ publikačních výsledků pro dosažení lepších „skóre“ při získávání pracovních míst, projektů, lepšího umístění institucí, atd. a tedy lepšího financování je všeobecně znám nejen u nás, ale v celém světě. Je nutné ovšem dodat, že evaluační nauka se dokázala vypořádat i s tímto jevem – navrhuje hodnotit pouze ty výstupy (publikace), které organizace, instituce, pracovní skupina, jednotlivec označí sama za nejkvalitnější.<sup>8</sup> Přes všechna nebezpečí z nesprávného použití při hodnocení je však správnost a úplnost údajů v databázích vědeckých informací nezbytným předpokladem pro hodnocení ve vědě a výzkumu.

### Jaká je situace v ČR?

Na rozdíl od mnoha jiných států je u nás povinnost hodnocení ve VaV dána legislativně zák. č. 130/2002 Sb., který jasně vymezuje povinnosti poskytovatelů veřejných prostředků v oblasti hodnocení věcné i formální náplně činnosti institucí a při řešení programů a projektů. Rovněž existence nadresortního orgánu – RVVI, v čele s předsedou vlády jako předsedou je oproti jiným systémům rozdělování veřejných prostředků pro VaV velkou systémovou a neocenitelnou výhodou, v praktické rovině je pak nevýhodou možnost politických ovlivnění řady jinak čistě odborných procesů, třeba např. v oblasti hodnocení. Výše jmenovaný zákon zároveň v navazujících vládních materiálech<sup>9</sup> přímo definuje procesy a metodiku hodnocení VaV (Předkládací zpráva II.1 str.20) „*Systém hodnocení v ČR bude respektovat světové trendy a využívat nové poznatky a nejlepší zkušenosti z hodnocení výzkumu v jednotlivých členských zemích EU, resp. OECD. Základními principy hodnocení budou: multikriteriální přístup, prokazatelná odborná kompetentnost, konkrétnost, transparentnost, nezávislost a objektivita*“ (Předkládací zpráva II.1 str.28) „*S cílem dosažení celkově vyšší kvality hodnocení výzkumu budou analyzovány používané postupy ve vyspělých zemích a vybrané metodiky a indikátory budou aplikovány v našich podmínkách. ČR se více zapojí do*

<sup>8</sup> Teoretický návrh omezit podle oboru počet publikací odborníka na předem určený počet je však v přímém rozporu s demokratickými principy svobodných společností.

<sup>9</sup> Usnesení vlády ze dne 23. června 2004 č. 644 k hodnocení výzkumu a vývoje a jeho výsledků

*mezinárodních systémů hodnocení a ve větší míře bude uplatňován benchmarking. Všeobecně bude velmi žádoucí zvyšovat celkovou hodnotitelskou kulturu. Rovněž bude podporován odborného zázemí pro oblast hodnocení a další relevantní otázky výzkumné politiky.“*

Bohužel však v praktické rovině nejsou tyto metodicky správné přístupy zcela uplatňovány. Již několik let se snaží RVVI vypracovat tzv. Metodiku hodnocení VaV, jejíž hlavní princip spočívá v tom, že vědecká a odborná produkce je definována podle „produktů“, čili výsledků (Jimp - článek v časopise s indikátorem IF, J článek v recenzovaném časopise, B – odborná kniha, atd.), „produktům“ je přidělena „váha -body“, tj. arbitrárně a dohodou stanovený koeficient, který „produkt“ podle významnosti „zvýhodní“ (např. výsledek typu Jimp, patent) nebo „nivelezuje“ (např. článek ve sborníku). Suma vah výsledků se stanoví u příjemců prostředků z veřejných prostředků na VaV a následně se stanoví suma vah/bodů příjemců jednoho poskytovatele. Do Metodiky hodnocení 2007 se vyhodnocovala tzv. „efektivnost“ = suma vah výsledků se dělila množstvím prostředků SR na VaV spotřebovaných na „výrobu“ daného váženého/bodovaného množství výsledků. Tento postup byl však pro nemožnost správně stanovit vstupy prostředků a nákladovost oborů pro „produkci“ výsledků opuštěn. Cílem takového hodnocení institucí, jež dostávají prostředky z veřejných zdrojů, je rozdělování institucionálních prostředků. Protože cílem tohoto příspěvku je spíše popis tohoto přístupu z hlediska evaluačních přístupů a metodik, omezíme se jen na stručné shrnutí kladů a záporů celého postupu.

#### Klady

1. Je nutno vysoce ocenit vůli k hodnocení oblasti VaV, včetně snahy o zlepšování evaluačního systému, tedy vůli k získání objektivních poznatků o výkonnosti systému VaV.
2. Každoroční hodnocení oblasti VaV zavádí systémový prvek interim evaluace. Poznání tendencí a trendů je zřetelně cennější než poznatky o okamžitém (= v minulém roce) stavu.
3. Zařazení zástupců vědecké komunity do odborných komisí RVV má daleko hlubší smysl než jen přizvání k řešení zadaného úkolu (např. stanovení vah „produktů“) a je systémovým pozitivem
4. Vysoce pozitivní je úsilí o shromažďování dat o výsledcích do Informačního systému RIV elektronicky bez přílišného zatěžování vlastních „producentů“ výsledků

#### Zápory

1. Současná Metodika stanovení výkonnosti VaV má zjednodušený charakter administrativní input-output analýzy

2. Z velmi rozdílné povahy oborů VaV je zřejmé, že jednotlivé častníky/segmenty/organizace nelze hodnotit jednotným způsobem. Je nutno daleko přesněji formulovat otázky, které mají přinést objektivní údaje.
3. Vypovídací/informační hodnota databáze RIV je nedostačující, přináší zkreslené údaje, protože úroveň vkladatelů dat do systému u jednotlivých poskytovatelů se někdy diametrálně liší. K hodnocení pak dochází ne na základě skutečného výkonu, ale podle dat vložených do RIVu. Na druhé straně je to však nesmírně cenný soubor údajů, pokud by se správně (tj. nemechanicky) využíval.
4. Odborná komunita se pragmaticky přizpůsobuje požadavkům toho kterého znění Metodiky hodnocení, uvádějí se jakékoliv výsledky „jen, když přinesou body“. Navíc tlak na „produkcii“ co největšího počtu výsledků je v přímém rozporu s etickými principy VaV. Důsledkem je Metodikou přijatý princip, že rozhodující je množství výsledků uplatnitelných v RIVu bez ohledu na jejich kvalitu.
5. Výkonnost, efektivitu, potenci, excelenci a další parametry účastníků VaV nelze hodnotit jediným, byť sebesložitěji sestaveným indikátorem, navíc na absolutní škále.
6. Složení veřejných prostředků použitých pro produkci výsledků vkládaných do RIVu u jednotlivých poskytovatelů je odlišné, měření „efektivnosti“ je potom více než problematické
7. Sestavování hodnotícího indikátoru (=váhy, body) bylo primárně odvozeno od indikátoru IF společnosti Thomson Scientific bez hlubší analýzy jeho použitelnosti v různých oborech VaV a bez srovnání s obdobnými databázemi např. Scopus, ERIH, apod. Navíc nebyla vůbec uplatněna úvaha, že impakt (vliv) určité jedné publikace se statisticky významně liší od indikátoru IF časopisu samého. Hodnota vah, bodů byla stanovena arbitrárně bez hlubší analýzy a modelování, co vlastně ta která navržená hodnota způsobí s výslednou strukturou nejen produkce VaV, ale i „producentů“ výsledků.
8. Pracovní skupina stanovující tyto váhy je složena z odborníků různých zaměření, nicméně v oblasti hodnocení a jeho metodik jsou to jen poučení laici. Z hlediska evaluačních metodik je tento postup především *ranking* (viz definice výše) a poskytuje pouze omezený soubor výsledných komprimovaných informací. Použití a případné následky tohoto použití z takto získaných data o výkonnosti ve VaV není tématem tohoto příspěvku.

### Hodnotíme vědu profesionálně?

Z velmi stručného přehledu výše vyplývá, že u nás věda a výzkum dosud na profesionální úrovni hodnocen není, ač je to i při současné legislativně možné. Co však chybí? Profesionální experti, hlubší dialog s odpovědnými místy v tomto případě s RVVI ale také veřejná publicita hodnocení VaV. Řešením je tedy změna přístupu a uplatnění základních evaluačních zásad, jak je to obvyklé v rozvinutých společnostech:

- hodnocení je multikriteriální, včetně indikátorových údajů a rozsáhlého použití databází vědeckých a odborných informací
- hodnocení je periodické
- výsledky by měly být hodnoceny bez ohledu na zdroje získané k produkci výsledků

- důležitější je hodnocení trendů, než okamžitý stav
- *peer review* se vždy provádí v mezinárodním srovnávacím kontextu
- pro zajištění objektivnosti *peers* systém zahrnuje hodnocení kvality posudků, výroků panelů a způsob řešení rozporů
- pro hodnocení je vždy nutno vytvořit taxonomii oborů
- kvalita „produktů“ je hodnocena ne ABSOLUTNÍ ale relativní škále (např. E – excellent, G dobrá – Good, A přijatelná – Acceptable, L podprůměrná – Limited), kde potom index excellence lze stanovit následovně  $I_E = [(E \times 1.0) + (G \times 0.8) + (A \times 0.6) + (L \times 0.2)]/T$

Závěrem lze říci, že komplexní návrh na hodnocení institucí VaV v ČR financovaných z veřejných prostředků zatím podala AV ČR a jeho plné znění je přístupné na [http://www.avcr.cz/stanoviska\\_avcr.php](http://www.avcr.cz/stanoviska_avcr.php). Jedná se „Research Assessment Exercise<sup>10</sup>“ na národní úrovni a je kombinací evaluačních postupů z Itálie, Německa a Holandska. Pokud lze odhadnout vývoj v této oblasti, je možné s vysokou pravděpodobností konstatovat, že zrychlená a zrychlující se internacionalizace vědy a výzkumu povede i v oblasti hodnocení VaV k metodicky správným a profesionálním postupům.

---

<sup>10</sup> Neznáme český ekvivalent