

Komentář k 1. fázi implementace Modulu-1 Metodiky17+ v průběhu roku 2018 v 6. vědní oblasti (foridy 6.1 až 6.5)

V průběhu roku 2018 (únor až červen 2018) byla v rámci prvního roku implementace Metodiky17+ [dále jen „M17+“] vedle části¹ Modulu-2 provedena také implementace části² Modulu-1.

Toto hodnocení v rámci Modulu-1 bylo založeno výhradně na posuzování „peer-review“ vzdálenými hodnotiteli, což je v případě vědních oborů 6. vědní oblasti (Humanities) vzhledem k odlišné typologické struktuře³ výsledků VaV obecně vnímáno jako pozitivní prvek celého systému. Pro 6. vědní oblast však představovala zásadní problém skutečnost, že vůbec nebyl zohledněn obsah znění M17+, která předpokládala, že pro obory společenské a humanitní bude (na základě zkušeností z projektu IPN Metodika a na základě zkušeností z předchozí Metodiky13+) postup hodnocení upraven s ohledem na oborová specifika.⁴ Dle znění M17+ měl být návrh na upřesnění metodický postup pro hodnocení oborů společenských a humanitních v rámci M17+ předložen vládě do 30. 6. 2017, což se nestalo.

V prvním roce implementace M17+ byl tedy i pro obory 6. vědní oblasti použit stejný postup hodnocení, jako pro všechny ostatní vědní oblasti, i když už samotné znění M17+ předjímá, že se tímto způsobem postupovat nemá.

Přesto se domnívám, že hodnotící proces, realizovaný v rámci 6. vědní oblasti v roce 2018, nebyl úplně zbytečný, neboť se podařilo exaktně prokázat, že pro humanitní obory zvolený postup nemá smysl a nelze jej použít k hodnocení kvality výsledků vědy a výzkumu, natožpak k hodnocení výzkumných organizací jako celku. Ani v jedné z fází letošního hodnocení totiž vůbec nebyly do hodnocení pojaty výsledky toho typu, které jsou v 6. vědní oblasti považovány (v rámci celkové typologie oborově validních výsledků vědy a výzkumu) za nejkvalitnější z hlediska domácího i mezinárodního srovnání:

- 1) **Modul-2:** Byla provedena jen bibliometrická analýza (vyhodnocení v březnu 2018). Ta vůbec nepostihuje nejkvalitnější oborové výsledky vědy a výzkumu 6. vědní oblasti (především knihy), neboť ty nejsou mechanicky posuzovatelné bibliometrickými postupy. Nebibliometrizable výsledky, tvořící integrální součást Modulu-2 M17+, vůbec nebyly v rámci prvního roku implementace M17+ brány v potaz. Modul-2 Metodiky17+, jehož hlavním cílem je zhodnocení celkové výkonnosti výzkumu, s jejich zohledněním počítá (Modul-2 zahrnuje „...kompletní produkci výsledků VO v jednotlivých oborech za hodnocené období...), nikoli však s využitím dat z mezinárodních databází (jak je tomu u výsledků bibliometrizable), ale výslovně se odkazuje na zdroj dat z Is VaVaI. Předložená bibliometrická analýza tedy v rámci 6. vědní oblasti vůbec nemohla postihnout celkovou

¹ V rámci Modulu-2 M17+ byla analyzována pouze bibliometrizable část všech výsledků VaV, vykázaných za rok 2016 do systému RIV.

² V rámci 10% vybraných výsledků byly hodnoceny pouze výsledky nebibliometrizable a to pouze podle kritéria společenské relevance.

³ V 6. vědní oblasti jsou dominantně nejvýznamnějšími výsledky odborné knihy, jejichž vědeckou kvalitu nelze hodnotit mechanicky bibliometrickými postupy.

⁴ „Metodika hodnocení výzkumných organizací a hodnocení programů účelové podpory výzkumu, vývoje“ (viz <http://www.vyzkum.cz/FrontClanek.aspx?idsekce=799796>; dále v textu jen „M17+“); zde platná verze „M17+“, schválená usnesením vlády ČR ze dne 8. února 2017 č. 107 (Úřad vlády ČR, č. j.: 21805/2016-OMP), s. 28.

výkonnost výzkumu. Výsledný podrobný komentář za 6. vědní oblast byl předložen jako samostatný materiál.⁵ Potvrzuje již dávno známou skutečnost, že humanitní vědy jsou „...segmentem vědy a výzkumu, pro které je bibliometrie nedokonalým nebo zcela nepoužitelným nástrojem hodnocení...“ a kde „mezinárodní databáze postihují jen malou část vědeckých výsledků registrovaných v IS VaVaI...“ (citováno z textu M17+, ukládající vypracování specifického postupu hodnocení společenských a humanitních oborů).

- 2) **Modul-1:** Vyběr kvalitních výsledků (10% výsledků, evidovaných v systému RIV) byl již při samotném zadání pro první rok implementace M17+ omezen jen na výsledky nebibliometrizable s tím, že v prvním roce hodnocení budou hodnoceny jen ty výsledky, které mají být posuzovány jen podle kritéria společenské relevance (tedy primárně výsledky aplikovaného výzkumu). Tato druhá podmínka z hodnocení v prvním roce prakticky vyřazovala nejkvalitnější výsledky VaV 6. vědní oblasti, ve které dosud není obecně definováno, jaké výsledky vlastně mají charakter výsledků aplikovaného výzkumu (se dvěma výjimkami, zmíněnými níže v komentáři), resp. čím je u těchto výsledků dána jejich společenská relevance.

V rámci Modulu-1 (10% vybraných výsledků) lze zkušenosti z hodnotícího procesu v prvním roce implementace M17+ v rámci Modulu-1 (10% vybraných výsledků) u 6. vědní oblasti (Humanities) rozdělit do dvou částí. Jedna se týká věcného obsahu, druhá technické realizace.

A) Věcný obsah

Pro 6. vědní oblast (Humanities) dosud nebylo přesněji definováno, jaké výstupy VaV lze akceptovat jako výsledky aplikovaného výzkumu, resp. výstupy, u nichž je hlavním kvalitativním kritériem společenská relevance, ani není přesněji specifikováno, jakým způsobem by měla být míra společenské relevance u těchto výsledků evidována či hodnocena. Samotný text M17+ je v této otázce formulován jen zcela obecně. Vzhledem k formální potřebě i v těchto oborech nějaké výsledky aplikovaného výzkumu vykazovat jsou zatím z blíže nevysvětlených důvodů v různých projektech (např. NAKI) akceptovány jen dva přesněji definované typy výstupů, které jsou pro humanitní obory uznávány jako výsledky aplikovaného výzkumu: historická rekonstrukční mapa a výstava (resp. muzejní expozice).

Tato situace se odrazila ve značné heterogenitě vykazovaných výsledků, které měly nesourodý charakter. Vzhledem k nedostatečnému vysvětlení změny, která meziročně v systému hodnocení nastala, byly v prvním roce implementace M17+ do Modulu-1 za 6. vědní oblast vykazovány ve značné míře i kvalitní publikační výsledky základního výzkumu, u kterých VO zřejmě jen z nepochopení uvedly, že tyto výsledky mají být hodnoceny podle kritéria „společenské relevance“, neboť v prováděcí dokumentaci ke sběru dat je jasně uvedeno, že v prvním roce hodnocení se bude v rámci Modulu-1 hodnotit pouze podle kritéria „společenské relevance“. Z tohoto materiálu jako celku jasně vyplývá, že VO měly i v prvním roce sběru dat pro Modul-1 (tj. výsledky za rok 2016) možnost rozhodnout, podle jakého kritéria má být vybraný výsledek hodnocen. Informace, že vybrané výsledky, které mají být hodnoceny podle kritéria "přínosu k poznání", budou standardně hodnoceny až v dalším roce (a že se tedy „neztratí“), však nebyla vysvětlena dostatečně srozumitelně.

Samotné důvody rozdělení Modulu-1, do kterého má být dle znění M17+ vykázáno 10% vybraných kvalitních výsledků (bez ohledu na jejich formu a kritérium posuzování) nebylo výzkumným organizacím dostatečně vysvětleno, proto se nelze divit, že při tak zásadní změně hodnotícího systému, který M17+ přináší, došlo ze strany VO k interpretačním zmatkům. Pro řadu oborů mohlo být matoucí, proč do Modulu-1 nemohou v prvním roce sběru dat vykazovat mezi 10%

⁵ Viz <http://hodnoceni17.rvvi.cz/www/>, zde složka „Přílohy“, položka „6 Humanities and the arts.pdf“, text „6 Souhrnný komentář k bibliometrické analýze výsledků vědní oblasti 6. Humanities and the Arts“.

svých nejkvalitnějších výstupů právě výsledky bibliometrizovatelné (které jsou pro většinu oborů 1.-4. vědní oblasti nejvýznamnější). Stejně tak pro humanitní a společensko-vědní obory mohlo být matoucí, proč se má u jejich vybraných nejkvalitnějších výsledků posuzovat jen společenská relevance, když oborově nejkvalitnější výsledky vznikají v kategorii „přínos k poznání“.

Postup při výběru externích hodnotitelů nebyl příliš efektivní, takže 6. vědní oblast se potýkala i s nedostatkem vhodných hodnotitelů a doplňující navrhování proces hodnocení zdržovalo. To je však také relativně běžná komplikace na počátku zavádění každého nového systému. Během 2-3 let je možné oborově stabilizovat okruh kvalitních a „funkčních“ externích hodnotitelů.

B) Technická realizace

Technická realizace vykazovala běžné prvky chybovosti, typické při zavádění nového, nedostatečně odzkoušeného systému, jehož „ostrému“ zpuštění nepředcházela vůbec žádná praktická zkušební fáze, byť jsme o ni na úrovni předsedů hodnotících panelů (po zkušenostech s obdobnými problémy při zavádění Metodiky13+) s dostatečným předstihem žádali. Hodnotící databáze nakonec byla funkční, byť v některých detailech zbytečně těžkopádná, ale to lze pro další fáze hodnocení snadno upravit.

Za větší problém považuji spíše nedomyšlenou logistiku celého hodnotícího procesu, který se zcela zbytečně „zadržával“ na triviálních komunikačních chybách, a to především ve vztahu k externím hodnotitelům. V roce 2018 se jednalo o tyto čtyři hlavní nedostatky:

1) Externí hodnotitelé nebyli včas dopředu informováni o náročnosti, rozsahu a časových požadavcích na hodnotící proces. Před zahájením hodnotícího procesu jim nebyla poskytnuta žádná přesnější specifikace, jak mají při hodnocení postupovat a jaké aspekty posuzovaného výsledku by měli brát prioritně v potaz, mají-li jej posuzovat podle kritéria „společenské relevance“. V tom byl podle mého názoru u 6. vědní oblasti hlavní „zádrhel“, neboť výsledky VaV v oblasti humanitních věd dosud v žádném předchozím systému centrálního hodnocení výslovně podle tohoto kritéria posuzovány nebyly. Významná část externích hodnotitelů vůbec nebrala v potaz rozdíly mezi „přínosem k poznání“ a „společenskou relevancí“, a výsledky hodnotila podle jejich přínosu k poznání. Také nebylo jasně specifikováno, jakým způsobem mají být výsledky tohoto typu věrohodně vykazovány. To se ukázalo být například problémem v případě výstavy či expozice, u které je skutečně otázkou diskuse, jak lze vlastně prokázat míru společenské relevance takové akce. Je jí pouze počet prokazatelných návštěvníků (tj. počet prodaných vstupenek), nebo i jiná kritéria?

2) Hodnotící škála („1“ až „5“) dle kritéria společenské relevance nebyla pro humanitní obory dostatečně definována; hodnotitelé pouze obdrželi odkaz na text „M17+“, kde je tato škála definována jen zcela obecně (a spíše pro technické obory a uplatnění jejich výsledků v aplikační sféře). Pro 6. vědní oblast tedy neproběhla žádná „kalibrace“; každý hodnotitel si jednotlivé stupně hodnocení definoval sám podle svých zkušeností, což se pak odráželo i na vyšší četnosti rozdílů v hodnocení i o dva a více stupňů hodnotové škály.

3) Externí hodnotitelé nebyli na počátku hodnotícího procesu vůbec informováni o tom, že za svou práci budou honorováni. Tato okolnost, navíc umocněná okolností, zmíněnou výše v bodu 1), podle mého názoru zvyšovala četnost negativních reakcí na výzvy k hodnocení či ignorování těchto výzev či následných urgencí.

4) Systém fungování panelů nebyl dobře promyšlen, natožpak předem odzkoušen. Panelisté v prvním roce implementace fungovali vlastně jen jako „distributoři“ výsledků k externím hodnotitelům. Neexistovala ani možnost vyřadit z hodnocení výsledek, který byl zjevně chybně vykázan (např. špatná oborová atribuce; výsledek nebyl hodnotitelům poskytnut on-line; výsledek byl zjevně pod oborově uznávaným kvalitativním prahem apod.). Výlučná a nezastupitelná role garanta komplikovala průběh hodnocení, pokud garant z nějakého důvodu (např. nemoc, zahraniční pobyt apod.) nemohl do hodnotícího procesu aktivně vstupovat. Na druhou stranu (v zájmu urychlení

hodnotícího procesu) byli panelisté ze strany administrátorů žádáni o přímou komunikaci s externími hodnotiteli, což do budoucna v zájmu zachování objektivit hodnotícího procesu akceptovat nelze.

V prvním roce implementace bylo v rámci 6. vědní oblasti předloženo k hodnocení podle kritéria společenské relevance celkem 179 výsledků. Ve fordech 6.1, 6.3 a 6.5 se podařilo v zadaném čase vyhodnotit všechny výsledky. U fordu 6.2 zůstal nevyhodnocen jeden výsledek, neboť v důsledku jeho jazykové specifiky (kniha v jazyce tureckém) nebyl k dispozici žádný vhodný hodnotitel, který by byl schopen tento výsledek posoudit podle kritéria společenské relevance. Relativně velké množství výsledků se nepodařilo v úplnosti vyhodnotit (tj. nejsou k dispozici dvě platná hodnocení) v případě fordu 6.4 „Arts and History of Arts“ (10 výsledků z 57 předložených k hodnocení), neboť vzhledem k relativně velkému počtu výsledků a jejich specifickému charakteru se nepodařilo zajistit adekvátní externí hodnotitele.

Struktura výsledné kvalitativní stratifikace hodnoceného souboru se jeví na první pohled jako více-méně standardní s mírně převažujícím podílem kvalitnějších výsledků. V systému jsou tyto výsledky evidovány následovně:

ford	Výsledné hodnocení					Ne-hod.	Celkem počet
	1	2	3	4	5		
6.1	4	16	25	10	3	-	58
6.2	1	14	8	8	2	1	34
6.3	-	3	6	1	1	-	11
6.4	1	9	11	22	4	10	57
6.5	-	13	5	1	-	-	19
celkem	6	55	55	42	10	11	179

Výše nastíněný (z hlediska formálně statistického) vcelku příznivý obraz je však klamný, neboť v důsledku výše popsaných věcných i technických komplikací, které hodnotící proces provázely, jsou konkrétní výsledná hodnocení nesouměřitelná. Panelisté tuto situaci nemohli ovlivnit, takže k externím hodnotitelům byly opakovaně zasílány k hodnocení i výsledky, které podle našeho názoru hodnoceny vůbec být neměly. Systémem byly akceptovány i externí posudky zjevně matoucí, které vůbec nebraly v potaz, že výsledky mají být hodnoceny podle kritéria společenské relevance.

Největší počet výsledků (4) s výsledným hodnocením „1“ generoval ford 6.1 *History and Archeology*. Z těchto čtyř jsou však tři výsledky jednoznačně publikacemi základního výzkumu a také z komentářů externích hodnotitelů je zjevné, že je posuzovali podle kritéria „přínosu k poznání“ (tedy jako výsledky základního výzkumu), nikoli podle kritéria společenské relevance. Jen čtvrtý z těchto nejlépe hodnocených výsledků ve fordu 6.1 [Afghanistan. Rescued Treasures of Buddhism] podle mého názoru splňuje kritérium vysoké míry společenské relevance a patří do příslušné kvalitativní kategorie (jde o výsledek typu „B“, ale jeho vazba na výstavu světového formátu je dostatečně vysvětlena ve zdůvodnění i v komentářích externích hodnotitelů). V případě fordu 6.2 je otázka míry společenské relevance u jediného výsledku, hodnoceného celkově „1“ [Black Feminist Literary Criticism: Past and Present], dobře vysvětlena ve zdůvodnění VO i v komentářích posudků; stejně tak u jediného výsledku, který takto vysoko hodnotil ford 6.4 [Bedřich Smetana. Korespondence I. (1840-1862)]. Z šesti výsledků, které systém generoval pro 6. vědní oblast jako excelentní (hodnocení „1“) lze tedy z ryze formálního hlediska postup hodnocení akceptovat jako správný jen u tří (50%).

Naopak například u výsledku „Krvavé léto 1945“ (ford 6.1) hodnotitel „známkuje“ velmi nízko (4) a ve slovním komentáři toto rozhodnutí zdůvodňuje, že u této knihy „...*popisnost problematizuje zařazení do kategorie vědeckých monografií*...“. Externí hodnotitel tedy zjevně posuzoval tento výsledek podle „přínosu k poznání“, a jeho „společenskou relevanci“ vůbec nezohledňoval.

Další deformace výsledné kvalitativní stratifikace vznikly i jiným způsobem. Například ve formu 6.5 vykazala jedna VO prakticky jeden a týž výsledek pětkrát [Památky nás baví 1.-5.], což je asi ryze formálně správně (jednalo se o pět jednotlivých samostatně evidovaných metodik). I z vlastního „Zdůvodnění významu výsledku“ (jehož text je u všech pěti vykázaných výsledků stejný) však vyplývá, že těchto pět metodik vzniklo v rámci jednoho projektu a vytvářejí integrální části jednoho celkového výsledku (cit: „...*Publikace je součástí celého souboru metodik Památky nás baví 1 – 5...*“). Měly být tedy posuzovány jako jeden společný výsledek s vysokou mírou společenské relevance (což je v tomto případě oprávněné), nikoli jako pět různých výsledků. Ani v tomto případě systém panelistům nenabízel žádné rozumné řešení: Vykázané metodiky jsou samy o sobě kvalitní a jedná se o ukázkový případ aplikovaného výstupu s vysokou mírou společenské relevance (hodnoceno „2“), ale pětinasobným započtením téhož výsledku po jednotlivých částech (tvořících i podle samotného projektu jeden celek) by byla předkládající VO neoprávněně zvýhodněna.

Není účelem tohoto závěrečného komentáře podrobně analyzovat celkový obraz kvalitativní stratifikace výsledků, předložených v prvním roce implementace M17+ v rámci 6. vědní oblasti k hodnocení podle kritéria společenské relevance. Výše zmíněnými příklady jen konkrétně dokumentují, jakým způsobem je vzniklý výsledek deformován a proč ani Modul-1 (podobně jako Modul-2) nelze akceptovat jako věrohodný podklad pro hodnocení VO:

1) Výsledky vykazují vysokou míru konfuzity (u M-2 chybou zvolené metodiky; u M-1 v důsledku nedostatečné komunikace s VO a s externími hodnotiteli).

2) Vůbec neproběhlo hodnocení výsledků, které jsou oborově považovány v 6. vědní oblasti za nejvýznamnější ani v Modulu-1, ani v Modulu-2.

Souhrnně lze tedy konstatovat, že výsledky prvního roku implementace (2018) M17+ nepřinesly v rámci 6. vědní oblasti validní výsledky, neboť vzhledem k zadání byl v rámci příslušných vědních oborů hodnocen pouze okrajový segment výsledků VaV. Nejvýznamnější výsledky nebyly v rámci implementace M17+ zatím vůbec brány v potaz.

Odvedená práce však nebyla zbytečná, pokud bude zkušenost z letošního roku efektivně využita pro další vývoj hodnotícího systému. Po zkušenostech z prvního roku implementace M17+ bude jistě možné problematická místa odstranit. Je však nutné naplnit co nejdříve v plném rozsahu znění Metodiky17+_ve vztahu ke společenským s humanitním vědám. Výsledky za 6. vědní oblast by měly být vyhodnoceny znovu podle úplně upřesněné metodiky pro obory SHV.

Členové hodnotícího panelu pro 6. vědní oblast jsou díky svým zkušenostem pochopitelně schopni KHV doporučit postupy, které by zvýšily kvalitu, efektivitu a ekonomickou rentabilitu hodnotícího procesu v Modulu-1 i Modulu-2 M17+ u výsledků VaV humanitních oborů. Pokud o takový materiál, který by mohl přispět k naplnění požadavku M17+ na dopracování zvláštní metodiky hodnocení pro společenské a humanitní obory, bude ze strany KHV (resp. RVVI) zájem, můžeme jej vypracovat.

V Pardubicích 29. června 2018

Prof. PhDr. Petr Vorel, CSc.
předseda hodnotícího panelu
6. vědní oblasti