

Jerzy Marian Brzeziński\*  
Instytut Psychologii  
Uniwersytet Adama Mickiewicza, Poznań

## Przeciwko depersonalizacji i nadmiernej standaryzacji procesu ewaluacji w nauce

**Abstrakt.** W artykule autor bardzo krytycznie odnosi się do zjawiska nadmiernej standaryzacji ocen stosowanych w polskiej nauce. Zwłaszcza negatywnie ocenia bezkrytyczne posługiwanie się wskaźnikiem *Journal Impact Factor* (IF) z bazy WoS. Jego zdaniem takie wskaźniki jak: IF, Hirsch-h index, liczba cytowań powinny być stosowane tylko pomocniczo i bardzo ostrożnie. I pod tym względem zgadza się ze stanowiskiem DORA oraz EASE. Rzetelna ocena, to taka, która – przede wszystkim – odwołuje się do trybu *peer review* i wskaźników bibliometrycznych, ale stosowanych z dużą ostrożnością.

**Słowa kluczowe:** DORA, Journal Impact Factor (IF), peer review, wskaźniki bibliometryczne, indeks h-Hirscha, liczba cytowań, SCImago, WoS, Google Scholar

## Against depersonalization and over-standardization of the evaluation process in science

**Abstract.** In the article the author presents a very critical stance on over-standardisation of evaluation measures used in Polish science. His view concerning the indiscriminate application of the *Journal Impact Factor* (IF) from WoS database is particularly negative. In his opinion such indices as IF, Hirsch-h index, or the number of citations should be used only as a subsidiary measure in a very cautious way. Thus, he subscribes to the position represented by DORA and EASE. Reliable evaluation means, above all, resorting to such tools as *peer review* and bibliometric indices, although they should still be used with due care.

**Keywords:** DORA, Journal Impact Factor (IF), peer review, bibliometric indices, Hirsch-index, number of citations, SCImago, WoS, Google Scholar

Musimy przyjąć, jako pewien wyznacznik czasów w jakich żyjemy, że obracamy się w świecie ocen i to ocen w coraz większym stopniu odzieranych z eksperckiego odniesienia do jakości wytworów, czy to pojedynczych badaczy, czy to jednostek naukowych (uczelni, wydziałów uczelnianych, instytutów badawczych, laboratoriów). Jeśli chodzi o badaczy, to są oni oceniani w związku z ich awansami naukowymi (doktoraty, habilitacje, profesury), awansami zawodowymi (stanowiska na uczelniach), przystępowanie do konkursów o miejsca pracy w jednostkach naukowych, przystępowanie do konkursów o granty (NCN, NPRH, NCBiR itp.), stypendia, subsydia, nagrody naukowe. Z kolei jednostki naukowe są – wedle

---

\* Adres do korespondencji: Instytut Psychologii UAM, ul. Szamarzewskiego 89, 60-568 Poznań, e-mail: brzezuam@amu.edu.pl. Prof. Jerzy M. Brzeziński jest członkiem rzeczywistym PAN.

obowiązującego w Polsce prawa – oceniane, co cztery lata, przez Komitet Ewaluacji Jednostek Naukowych KEJN. I jest to największa akcja ewaluacyjna jaka prowadzona jest w Polsce. Budzi też ona liczne kontrowersje. Przede wszystkim dlatego, że nie jest „do końca” transparentna, a zasady wedle których jest prowadzona wprowadzane są nie na początku drogi, ale na jej końcu. No bo (a piszę ten tekst w połowie listopada 2015 r.) jeszcze nie jest znane kluczowe rozporządzenie ministra ds. nauki, wedle którego będzie poprowadzona ocena parametryczna jednostek naukowych w 2017 r. (obejmująca lata: 2013–2016). Brak aktualnych list czasopism naukowych, a to jest kluczowe dla oceny parametrycznej, gdyż aż 65–70% ocen (kryterium  $O_1$  w ankiecie oceny jednostki naukowej) bierze się z publikacji oraz patentów wytworzonych przez pracowników jednostek naukowych. I znowu prawo będzie działało wstecz – podobnie jak to było w latach poprzednich (por. krytyczną ocenę poprzedniej parametryzacji, przeprowadzonej w 2013 r. (Antonowicz, Brzeziński 2013)).

Procedurom ewaluacyjnym stosowanym w nauce – a odnoszę się do nauki uprawianej w Polsce, na polskich uniwersytetach i w polskich instytutach badawczych – nadają ton zwolennicy ujęć biblio- i naukometrycznych. Głoszą oni prymat podejścia bibliometrycznego nad podejściem *peer review*. Ich argument w tej dyskusji jest jeden – obiektywność, transparentność, intersubiektywność ocen ilościowych. Wygrywa ona z nieuchronnym subiektywizmem ocen eksperckich (modne hasło – „Polacy nie ufają ekspertom”). I drugi argument – zwłaszcza podnoszony w kontekście ocen jednostek naukowych stosowanych przez KEJN – ich relatywnie niska kosztochłonność. W dalszej części niniejszego artykułu spróbuję odnieść się do owej „obiektywności” ocen odwołujących się do popularnych wskaźników biblio- i naukometrycznych. Moim zdaniem – zwłaszcza, gdy są stosowane bez należytego umiaru i w oderwaniu od ocen *peer review* – prowadzą do obciążenia (*bias*) ocen. Jestem też przekonany – i nie jestem w tym odosobniony – że to, co dobre dla ocen jednostek naukowych nie musi być (i nie jest) dobre dla ocen poszczególnych osób.

### Przeciwko totalnej standaryzacji ocen

W zależności od przyjętej w danej procedurze ewaluacyjnej bazy bibliograficznej będziemy mieli różne wartości wskaźników zastosowanych do oceny dorobku naukowego zawężonego tylko do publikacji. To nie powinno dziwić, gdyż najbardziej popularna – ale tylko w Polsce! – baza **JCR**: firmy **Thomson Reuters** uwzględnia wyłącznie artykuły pomieszczone w czasopismach ujętych tylko w tej bazie. I w ten sposób zgodzono się na wyeliminowanie wielu czasopism o profilu humanistycznym i społecznym oraz tych, które nie ukazują się w języku angielskim. Co prawda, w tej bazie znajdują się też czasopisma polskojęzyczne, ale jest ich niewiele i zajmują niską pozycję punktową. Jest tak, gdyż ważność czasopi-

sma – a zatem i danego artykułu w tym czasopiśmie opublikowanego – zależy od wartości wskaźnika IF (*impact factor*) jaki to czasopismo uzyskało za ostatnie dwa lata. Przy czym nie powinno się zapominać, że ocenie podlegają tylko te czasopisma, które znalazły się w tej komercyjnej bazie. W świecie uznaniem (nawet większym) cieszy się niejako konkurencyjna baza **SCImago Scopus** firmy **Elsevier**. Niestety w Polsce nie może się przebić mimo usilnych starań różnych osób (także niżej podpisanego). Osobiście uważam, że powinno korzystać się z dwóch, a nawet trzech baz. Tą trzecią mogłaby być baza: **Google Scholar** (wysoko oceniana przez badaczy z obszaru informatyki). Jest ona też przydatna dla grupy nauk humanistycznych i społecznych (rejestruje artykuły, książki i rozdziały w pracach zbiorowych opublikowane w języku polskim). Także SCImago jest przyjazna dla nauk humanistycznych.

To, że wyróżnioną pozycję w naukometrycznie pojmowanej ocenie dorobku naukowego przypisano wskaźnikowi IF (pełna nazwa: *Journal Impact Factor*) doprowadziło do tego, że jednostki naukowe z grupy „nauk humanistycznych i społecznych” uzyskują w ocenach wystawianych przez KEJN zaniżone wartości punktowe w kryterium  $O_1$ ). W obowiązującym Rozporządzeniu Ministra Nauki i Szkolnictwa Wyższego z dnia 13 lipca 2012 r. w sprawie kryteriów i trybu przyznawania kategorii naukowej jednostkom naukowym. (*Dziennik Ustaw*, Poz. 877), w §14. 3 zapisano:

§14. 3. Wykaz czasopism naukowych składa się z:

- 1) części A – zawierającej liczbę punktów za publikacje w czasopismach naukowych posiadających współczynnik wpływu *Impact Factor* (IF), znajdujących się w bazie *Journal Citation Reports* (JCR);
- 2) części B – zawierającej liczbę punktów za publikacje w czasopismach naukowych nieposiadających współczynnika wpływu *Impact Factor* (IF);
- 3) części C – zawierającej liczbę punktów za publikacje w czasopismach naukowych znajdujących się w bazie *European Reference Index for the Humanities* (ERIH).

W przeliczeniu na skalę punktową, zaakceptowaną przez ministerstwo, mamy (listopad 2015 r.) następujące zakresy punktów, które można uzyskać za artykuł opublikowany w czasopiśmie mieszczącym się w jednym z wyróżnionych przez ministerstwo wykazów. I tak, wykaz A: 15–50 pkt., wykaz B: 1–10 pkt., wykaz C: 10 pkt.

Gdyby za tymi punktami nie był ukryty mechanizm „rywalizacyjny”, to można by powiedzieć, że nic złego się nie dzieje. Oto bowiem artykuł Kowalskiego ukazał się w czasopiśmie X i uzyskał ocenę 40 pkt., które znalazło się w „prestizowym” wykazie A na wysokiej pozycji (z uwagi na wysoką wartość IF tego czasopisma), a artykuł Nowaka został opublikowany w czasopiśmie Y i uzyskał ocenę 10 pkt., bo ukazał się w bardzo dobrym czasopiśmie polskim, ale... z „gorszego” wykazu B. Problem jednak w tym, że czasopismo Y, mimo wysokiej oceny w środowisku, nie

mogło uzyskać wyższej oceny, gdyż nie mieściło się w bazie JCR. Mało tego, jego ocena znajduje się poniżej najniższej oceny czasopism z tej bazy – wynoszącej 15 pkt. Rodzi się – i to zupełnie naturalnie – pytanie: *kto* i na podstawie jakiej wiedzy przesądza o poziomie publikacji w nauce po prostu, a nie tylko w tym jej podzakresie, który, wychodząc z anglosaskiej tradycji, nazywa się *science* w opozycji do *arts*?

Jeżeli jednostka naukowa (np. wydział uczelniany) skupia, w danej grupie wspólnej oceny, badaczy publikujących w czasopismach z każdej z tych baz, to rzecz jasna lepiej wypadnie w finalnej ocenie ta jednostka, która zatrudnia więcej osób publikujących w czasopismach z bazy JCR. Trudno nie zauważyć, że humaniści nie publikują w czasopismach z bazy JCR, tak jak fizycy nie publikują w czasopismach z bazy C: ERIH. Wydziały wielu uczelni nie są *de facto* jednorodnymi grupami wspólnej oceny. W wysokim stopniu jednorodne są instytuty PAN i instytuty badawcze. W „lepsz” sytuacji są na uczelniach te wydziały społeczne, które mają w swoich składach subjednostki o profilu psychologicznym (instytuty, katedry), gdyż psychologowie relatywnie częściej od np. filozofów czy pedagogów publikują w czasopismach z bazy JCR i podnoszą wartość kryterium  $O_1$ .

Problem można rozwiązać dwojako. Po pierwsze zmienić zasady oceny. Nie ograniczać się tylko do oceny jednostek uczelnianych uwzględnionych w ich statutach jako „jednostki podstawowe” i zejść „w dół” (to było możliwe w czasach funkcjonowania KBN). Zatem, chcąc uniknąć obciążenia ocen można by postąpić tak, jak to już wcześniej sugerowali Antonowicz i Brzeziński (2013):

Naszym zdaniem konieczne będzie rozbicie składankowych wydziałów i zejście na niższy poziom organizacji, wewnątrz tych wydziałów. Niech, przykładowo, psychologia będzie bezpośrednio porównywana z psychologią, socjologia z socjologią, a polonistka z polonistyką. Nowe zdefiniowanie jednostki naukowej jako takiej, która jest jednorodna pod względem reprezentowanej dziedziny, subdziedziny czy dyscypliny naukowej uważamy za punkt wyjścia przyszłej (2017 r.) parametryzacji. Bez tego podstawowego zabiegu dalsze korekty będą miały charakter jedynie kosmetyczny. Jako pracownicy instytutów socjologii i psychologii [...] (a „nasze” wydziały mają charakter składankowy) chcielibyśmy wiedzieć, jak nasze instytuty sytuują się na tle pozostałych instytutów w kraju, a – co też jest możliwe do przeprowadzenia, gdy pomyślimy o naukach społecznych – także poza jego granicami. Nie interesują nas oceny hybrydowych wydziałów. One nie są rzetelne i budzą zrozumiałe negatywne emocje.

Po drugie, obok bazy JCR Thomson Reuters wprowadzić, niejako równolegle, bazę *SCImago* Elseviera. W ostatnich konkursach NCN (panele: „Nauki Humanistyczne, Społeczne i o Sztuce”, 6 paneli: HS1 – HS6), przy ocenie pozycji naukowej kandydata dopuszczono możliwość podania dorobku mieszczącego się także w innych, aniżeli JCR bazach<sup>1</sup>. Przykładowo przytaczam cytaty z zasad oceny dorobku naukowego osób przystępujących do konkursu OPUS 10:

<sup>1</sup> Por. Zasady oceny wniosków w konkursie na finansowanie projektów badawczych, w tym finansowanie zakupu lub wytworzenia aparatury naukowo-badawczej niezbędnej do realizacji tych projektów – „OPUS”: [https://www.ncn.gov.pl/userfiles/file/konkursy\\_ogloszone\\_2015-09-15/opus10-zal5.pdf](https://www.ncn.gov.pl/userfiles/file/konkursy_ogloszone_2015-09-15/opus10-zal5.pdf), dostęp: 17.11.2015.

[...] publikuje [kierownik projektu – JB] w [...] wydawnictwach/czasopismach, w tym również w czasopismach znajdujących się na liście Journal Citation Reports (JCR) lub, w przypadku HS, na porównywalnej (o ile mają zastosowanie), jego prace są średnio cytowane (zgodnie z bazą Web of Science™ Core Collection, w przypadku HS, tam gdzie istnieją odpowiednie bazy danych).

Optymalne rozwiązanie, to takie, które łączy oba podejścia.

Problem preferowania jednej tylko bazy czasopism odnosi się nie tylko do oceny pozycji jednostek naukowych, ale dotyczy także ocen formułowanych przez recenzentów opiniujących dorobek naukowy osób ubiegających się o stopień naukowy doktora habilitowanego.

Nie tylko moim zdaniem (zob. Zakrzewski 2012) należy – najlepiej „od zaraz” – znowelizować *Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 1 września 2011 r. w sprawie kryteriów oceny osiągnięć osoby ubiegającej się o nadanie stopnia doktora habilitowanego* (Dziennik Ustaw, Poz. 1165). Czytamy w nim bowiem – także w odniesieniu do badaczy z obszaru czystej humanistyki, np. polonistyki, że:

§ 4. Kryteria oceny w zakresie osiągnięć naukowo-badawczych habilitanta we **wszystkich obszarach wiedzy** obejmują: [...]

- 3) sumaryczny *impact factor* publikacji naukowych według listy Journal Citation Reports (JCR), zgodnie z rokiem opublikowania;
- 4) liczbę cytowań publikacji według bazy Web of Science (WoS);
- 5) indeks Hirscha opublikowanych publikacji według bazy Web of Science (WoS).

Ktoś jednak nie pomyślał! Nie mam zamiaru „znęcać się” nad tym – tak po prostu – złym i kompromitującym ministerstwo dokumentem; zresztą zrobił to dostatecznie dotkliwie cytowany wyżej Zakrzewski.

Trzeba jednak zapytać:

1. Dlaczego – w odniesieniu do wszystkich dyscyplin naukowych – tylko baza WoS jest brana pod uwagę?
2. Dlaczego liczba cytowań jest brana tylko z bazy WoS, a nie – co szczególnie ważne dla humanistów – z bazy Google Scholar?
3. Dlaczego wartość wskaźnika h-Hirscha obliczana jest tylko z bazy WoS, a nie też z baz: Scopus i Google Scholar?

Kolejny obyczaj ewaluacyjny – zwłaszcza pielęgnowany na uczelniach medycznych i Akademii Wychowania Fizycznego (te zapewne pod wpływem tych pierwszych) – polega na charakteryzowaniu osób ubiegających się o awans naukowy (doktorat, habilitacja, profesura) za pomocą dwóch wskaźników będących prostym zsumowaniem (za pewien okres czasu ubiegający od ostatniego awansu) wartości liczbowych: (a) IF czasopism w których publikował kandydat – *sumIF* i (b) zsumowaniu punktów zebranych za publikacje: monografie, rozdziały w pracach zbiorowych, artykuły w czasopismach naukowych z ministerialnych baz: A, B, C; co urzędnicy uczelniani oznaczają jako: „suma punktów KBN/MNiSzW/

KEJN”. Zdefiniowane liczbowo wartości progowe tych dwóch wskaźników traktowane są – orientacyjnie (mam nadzieję, że wyłącznie orientacyjnie) – jako wskaźniki dojrzałości kandydatki/kandydata do awansu naukowego. Pozostawiam to bez komentarza. Czyż nie modeluje to, zwłaszcza wśród młodych – jeszcze na początku drogi naukowej, badaczy – recepty na życie polegającej na ciułaniu punktów? Młodzi nie mają wyjścia w silnie zhierarchizowanym społeczeństwie nauki muszą się podporządkować. Tylko dlatego dojrzałe uczeni wpadli w pułapkę „punktomanii”? Dlaczego „dodają” i „dzielą” te punkty sądząc, że wyłoni się z nich trafny obraz dokonań naukowych młodego badacza i instytucji naukowych? Mało tego, są przekonani, że w ten oto prosty, perswazyjny sposób wydatnie poprawią stan polskiej nauki. Od samego mieszania łyżeczką herbata nie staje się słodsza.

Skupianie się jedynie na kolekcjonowaniu punktów (inna sprawa czy są one trafnie „zważone”?) prowadzi do sytuacji bardzo trafnie opisanej przez Davida L. Parnas (2007):

Measuring productivity by counting the number of published papers slows scientific progress; to increase their score, researchers must avoid tackling the tough problems and problems that will require years of dedicated work and instead work on easier ones. Evaluation by counting the number of published papers corrupts our scientists; they learn to „play the game by the rules”.

I dlatego, jak zaleca Parnas (ibidem):

[...] If you get a letter of recommendation that counts numbers of publications, rather than commenting substantively on a candidate’s contributions, ignore it; it states only what anyone can see.

Przed swoistą „punktomanią” przestrzegał też, w liście do redakcji *PAUzy Akademickiej*, wybitny fizyk Andrzej Kajetan Wróblewski (2014), który opisuje jedną z takich, wręcz modelowych, sytuacji w jakiej się przed laty znalazł, gdy oczekiwało się od komisji konkursowej, że dojdzie ona do trafnych rozstrzygnięć na podstawie bogactwa zebranych punktów. Komisja jednak chciała widzieć badacza jako twórcę ciekawych prac, a nie jako „nośnika” punktów (przytaczam list *in extenso*):

W ubiegłym roku byłem członkiem jury dość prestiżowej nagrody fizycznej. Z tego aż sześciu kandydatów do nagrody podało w swych dokumentach tylko wartości „impact factor” i punkty ministerialne czasopism, w których opublikowali swoje prace, bez wyjaśnienia czego dotyczyły badania i jakie wyniki zostały osiągnięte! To dowodzi spustoszenia, jakie w ciągu ostatnich lat dokonało się w umysłach wielu członków środowiska naukowego w Polsce.

Oczywiście nasze jury odrzuciło od razu te 6 wniosków i wyłoniło laureata spośród pozostałych dwóch zgłoszonych kandydatów.

Od kilkunastu lat walczę o wyeliminowanie z ocen osiągnięć naukowych takich pseudonaukowych substytutów jak punkty i „impakty”. Z wielką satysfakcją przy-

jąłem mocny tekst przemówienia profesora Piotra Sztompki na Kongresie Kultury Akademickiej (PAUza Akademicka 247), jak i obecne Oświadczenie władz Fundacji na rzecz Nauki Polskiej. Jest nadzieja, że może uda się odrobić straty i oprzeć ocenę osiągnięć nauki na racjonalnych zasadach.

Rzecz jasna, można by przytaczać jeszcze dalsze przykłady absurdalnego, mechanicznego odwoływania się wskaźnika *IF* czy *sumIF* jako miarodajnego kryterium jakości dorobku naukowego danego badawcza.

Jednym zdaniem. Rzetelna i trafna ocena – nie stroniąca od danych naukometrycznych, ale bardzo ostrożnie i pomocniczo wykorzystywanych – musi wyjść poza konfigurację punktów i musi odwoływać się do wnikliwej oceny jakościowej tego, co badacz dokonał; musi być przeprowadzona w trybie *peer review*.

### Deklaracja DORA – czy nastąpi przełom?

Myślę, że już doszliśmy do kresu fascynacji pozorną li tylko ścisłością i obiektywnością ocen odwołujących się wyłącznie do bibliometrii (albo jej dominacji). Pierwszym, na skalę światową, krokiem na tej drodze przywracania normalności w ocenach stosowanych w nauce była *Deklaracja DORA*, czyli *San Francisco Declaration on Research Assessment. Putting science into the assessment of research* z 16 grudnia 2012 r. (DORA, 2012)<sup>2</sup>, która została przygotowana przez grupę redaktorów i wydawców czasopism naukowych w trakcie corocznego spotkania *American Society for Cell Biology* (ASCB) w San Francisco.

Skupieni wokół tej deklaracji jej bezpośredni sygnatariusze oraz liczne grono uczonych i znaczących instytucji naukowych zdecydowanie wypowiedzieli się przeciwko używaniu wskaźnika *IF* wytworzonego przez firmę Thomson Reuters jak wiarygodnej miary jakości dorobku naukowego poszczególnych badaczy oraz całych jednostek naukowych. Uzasadnienie dla tak jednoznacznie negatywnej opinii sprowadzono zostało do następujących punktów:

*Journal Impact Factor* jako narzędzie oceny badań ma wiele dobrze udokumentowanych niedostatków. Te ograniczenia to:

- A) w znacznym stopniu nierównomierne zamieszczanie cytowań w obrębie rozmaitych czasopism, a więc interpretacja – wypaczona (Adler i in. 2008; Seglen 1997; Edytorial 2005);
- B) właściwości *Journal Impact Factor* różniące się w rozmaitych obszarach (specjalnościach) wiedzy; jest to wypadkowa mieszanki wielu, znacznie różniących się między sobą, rodzajów publikacji, w tym zarówno prac prezentujących wyniki badań oryginalnych, jak i prac o charakterze przeglądowym (Adler i in. 2008; Seglen 1997; Edytorial 2005; Vanclay 2012);
- C) *Journal Impact Factors* mogą być przedmiotem manipulacji (czy „rozgrywania”) poprzez dobór odpowiedniej polityki redakcyjnej (The PLoS Medicine Editors 2006);

<sup>2</sup> Por. także przekład na j. polski opatrzony komentarzem redaktora naczelnego czasopisma: *Nowotwory Journal of Oncology* (Towpik 2013). Przytoczone w tym artykule cytaty z *Deklaracji DORA* pochodzą z tego przekładu.

- D) brak przejrzystości oraz pełnej dostępności dla opinii publicznej danych, które wykorzystuje się do obliczania *Journal Impact Factors* (Vanclay 2012; Rossner i in. 2007; Rossner i in. 2008).

Wymienione ograniczenia oraz inne – nie mniej istotne – wady wskaźnika IF zostały bardzo wnikliwie przeanalizowane w bardzo dobrze udokumentowanym (staranna analiza obszernej literatury przedmiotu) opracowaniu profesora medycyny Edwarda Towpika (2015).

Wedle Deklaracji DORA należy, w ocenie osób i instytucji, kierować się ogólnym zaleceniem:

1. Nie należy stosować danych bibliometrycznych czasopism (*journal-based metrics*), takich jak *Journal Impact Factors*, jako zastępczej miary jakości poszczególnych artykułów naukowych, do oceniania wkładu konkretnych badaczy czy podejmowania decyzji o ich zatrudnieniu, promowaniu lub finansowaniu.

Autorzy *Deklaracji DORA* ogłosili – poza wyżej przytoczonym ogólnym zaleceniem – 17 zaleceń szczegółowych. Można je sprowadzić do nakłaniania badaczy (instytucji, komisji konkursowych, redakcji) do tego, aby oceny pracy naukowej opierali na opiniach odnoszących się bezpośrednio do treści analizowanych prac, a nie do oderwanych od tych treści wskaźników bibliometrycznych, aby, jeżeli już odwołują się do wskaźników bibliometrycznych, aby czynili to bardzo ostrożnie i zawsze w powiązaniu z treścią ocenianych prac.

Fundacja na Rzecz Nauki Polskiej, która – bodajże jako jedyna znacząca polska instytucja naukowa – podpisała się pod *Deklaracją DORA* w wydanym oświadczeniu (FNP 2013) napisała, iż w wypadku, gdy ocenie podlegają małe jednostki (wg Żylicza o  $N < 60$ ) lub poszczególni badacze, to:

[...] należy bezwzględnie stosować system oceny *peer review*, w którym starannie wyselekcjonowani uczeni oceniają oryginalność indywidualnych osiągnięć naukowych innych badaczy, a pomocniczą rolę może spełniać liczba cytowań oraz wskaźnik *h*.

Uważam to za trafne uszczegółowienie, zwłaszcza w kontekście praktyk ewaluacyjnych stosowanych przez KEJN, uzupełnienie Deklaracji DORA.

W podobnym duchu wypowiedziało się European Association of Science Editors (EASE) w oświadczeniu *EASE statement on inappropriate use of impact factors*:

European Association of Science Editors zaleca, aby kryterium wskaźnika oddziaływania czasopisma posługiwać się wyłącznie – i to z wielką ostrożnością – do oceny i porównywania jedynie wpływu czasopism jako takich, a nie do oceny poszczególnych prac. Z pewnością zaś nie należy go wykorzystywać do oceniania badaczy lub programów badawczych, tak bezpośrednio (*directly*), jak i zastępczo (*as a surrogate*). (cyt. za: EASE (2012))



To, co teraz należy zrobić, to wdrożyć zalecenia zawarte w tej Deklaracji do praktyki – zwłaszcza takich instytucji jak: KEJN, NCN, Centralna Komisja do spraw Stopni i Tytułu, MNiSzW oraz Rady Naukowe Instytutów PAN i Instytutów Badawczych oraz Rady Wydziałów i Senaty uczelni.

### **W stronę normalności – tylko ocena *peer review* ostrożnie wspomagana wskaźnikami bibliometrycznymi**

W poprzednich punktach starałem się przekonać Czytelnika, iż nadmierna standaryzacja i pozorna li tylko obiektywizacja ocen stosowanych w procedurach ewaluacyjnych – czy to pojedynczych badacz czy to jednostek naukowych – nie tylko nie przynosi zamierzonych efektów, ale wikła gremia decydujące o ocenach w problemy natury etyczne. Niech wymownym przykładem będzie status wskaźnika czerpanego z bazy WoS: *Journal Impact Factor* (IF). No dobrze, jeżeli nie wskaźniki bibliometryczne, to co? Moim (ale nie tylko) zdaniem powinniśmy sięgnąć do ocen poprowadzonych w trybie *peer review*. Jest tylko jeden problem, taka ocena jest i kosztowna i bardziej pracochłonna (zwłaszcza gdy będzie stosowana w odniesieniu do jednostek naukowych, szczególnie dużych), a także (gdy będzie poprowadzona niestarannie) narażona na błąd stroniczości. Jeśli chodzi o zarzut kosztowności, to odpowiedź jest tylko jedna. Albo przeprowadzimy oceny starannie, wnikliwie, z uwzględnieniem treści prac i wówczas to musi kosztować, albo też zrobimy to powierzchownie, szybko, byle jak i wówczas będzie to tanio.

Jak jednak postąpić, aby nie dać się przytłoczyć nadmiarem prac przekazywanych do oceny? Teraz zespoły dokonujące ocen w ramach prac np. KEJN znalazły się w trudnej sytuacji. Nie tylko nie mogą oceniać treści, to jeszcze muszą się niekiedy bawić w detektywów, aby upewnić się czy dana praca należy do danej jednostki (błędy i oszustwa). Przejście na system *peer review* – a sądzę, że jest ono nieuchronne – wymaga radykalnych zmian w zakresie przygotowania materiałów przekazywanych ekspertom.

Jednym z trafnych wyjść z sytuacji patowej w jakiej znalazły się gremia opiniodawcze jest propozycja złożona przez Macieja Żylicza (Żylicz 2008):

W ocenie naukowej indywidualnej lub zbiorowej (np. ocena parametryczna) nie powinniśmy uwzględniać wszystkich prac opublikowanych, ale na przykład poddawać ocenie 25% prac, które sam autor w okresie oceny, np. nie krótszym niż 5 lat, będzie uważał za najważniejsze. Może po paru latach takiej praktyki nasze środowisko samo dojdzie do wniosku, że opta się publikować mało, ale za to dobre prace.

Można, rzecz jasna, podyskutować o tym, ile lat publikowania należy wziąć pod uwagę i ile publikacji powinien przedstawić kandydat, ale to sprawa wtórna. Ważne, aby nie brać pod uwagę „śmieci”, aby nie zamulać oceny nadmiarem i tak

nie mających znaczenia publikacji, które, w sumie podobnych im, wytworzonych przez innych niby-badaczy składają się na „naukę śmieciową” (*junk science*).

A moje zalecenia? Te można sprowadzić do następujących punktów:

1. W większym, aniżeli dotychczas, stopniu uwzględnić specyfikę dziedzin nauki – zwłaszcza nauk humanistycznych. To zaś oznacza wycofanie złego rozporządzenia ministra ds. nauki w sprawie kryteriów oceny dorobku naukowego badaczy ubiegających się o stopień doktora habilitowanego. Absurdalność proponowanych w nim rozwiązań szczególnie jest widoczna jeżeli weźmiemy pod uwagę specyfikę nauk humanistycznych.
2. Ograniczyć – do rozsądnych granic pojmowanych ilościowo – liczbę prac, które powinny być oceniane (akceptowany przez mnie postulat Żylicza).
3. Zrezygnować z bardzo wątpliwego wskaźnika IF (to między innymi postulat DORA).
4. Rozszerzyć liczbę baz czasopism uwzględnianych w procedurze ewaluacyjnej – koniecznie o bazę Scopus Elsewiera. Z dwóch, a może nawet trzech (jeszcze w grę może wchodzić Google Scholar) obliczać wskaźniki takie jak: liczba cytowań i h-Hirscha. Te zaś uwzględniać tylko pomocniczo.
5. Fundamentem podejmowanych decyzji powinna być ocena przeprowadzona w trybie peer review.

## Literatura

- Adler R., Ewing J., Taylor P., 2008, *Citation statistics. A report from the International Mathematical Union*, [www.mathunion.org/publications/report/citationstatistics0](http://www.mathunion.org/publications/report/citationstatistics0), dostęp: 16.11.2015.
- Antonowicz D., Brzeziński J., 2013, *Doświadczenia parametryzacji jednostek naukowych z obszaru nauk humanistycznych i społecznych 2013 – z myślą o parametryzacji 2017*, „Nauka” 4: 51–85.
- DORA, 2012, *San Francisco Declaration on Research Assessment. Putting science into the assessment of research*, <http://www.ascb.org/files/SFDeclarationFINAL.pdf>, dostęp: 16.11.2015.
- EASE, 2012, *Oświadczenie Europejskiego Stowarzyszenia Redaktorów Naukowych European Association of Science Editors – EASE, w sprawie niewłaściwego stosowania wskaźników oddziaływania IF – impact factors*, „Nowotwory. Journal of Oncology” 62.2: 139–140.
- Edytorial, 2005, *Not so deep impact*, “Nature” 435: 1003–1004.
- FNP, 2014, *Fundacja na Rzecz Nauki Polskiej: Oświadczenie władz Fundacji na rzecz Nauki Polskiej*, [www.fnp.org.pl/assets/O%C5%9Bwiadczenie-Rady-i-Zarz%C4%85du-FNP.pdf](http://www.fnp.org.pl/assets/O%C5%9Bwiadczenie-Rady-i-Zarz%C4%85du-FNP.pdf), dostęp: 16.11.2015.
- Parnas D.L., 2007, *Stop the numbers game*, „Communications of the ACM”, 50.11: 19–21. [ce.sharif.edu/~ghodsi/PaP/stop\\_the\\_number\\_game.pdf](http://ce.sharif.edu/~ghodsi/PaP/stop_the_number_game.pdf), dostęp: 16.11.2015.
- Rossner M., Van Epps H., Hill E., 2008, *Irreproducible results: A response to Thomson Scientific*, „J. Cell Biol.” 180: 254–255.
- Rossner M., Van Epps H., Hill E., 2007, *Show me the data*, „J. Cell Biol.” 179: 1091–1092.
- Seglen P.O., 1997, *Why the impact factor of journals should not be used for evaluating research*, „BMJ” 314: 498–502.
- The PLoS Medicine Editors, 2006, *The impact factor game*, „PLoS Med” 3.6: e291.

- Towpik E., 2013, *Deklaracja z San Francisco nt. Oceny Badań Naukowych DORA: czy to początek końca dotychczasowej metody oceny parametrycznej?*, „Nowotwory. Journal of Oncology”, 63.6: 501–504.
- Towpik E., 2015, *IF-mania: Journal Impact Factor nie jest właściwym wskaźnikiem oceniania wyników badań naukowych, indywidualnych uczonych ani ośrodków badawczych*, „Nowotwory. Journal of Oncology”, 65.6: 465–475.
- Vanclay J.K., 2012, *Impact Factor: Outdated artefact or stepping-stone to journal certification*, “Scientometric”, 92: 211–238.
- Wróblewski A.K., 2014, *Listy do Redakcji PAUzy Akademickiej*, „PAUza Akademicka” 249: 2.
- Zakrzewski J., 2012, *O reformie nauki*, „PAUza Akademicka”, 154: 2–3.
- Żylicz M., 2008, *Wypowiedź w dyskusji o polskich czasopismach naukowych*, „PAUza Akademicka” 14: 3.