

Warszawa, 2 lipca 2019

Jacek Koronacki

Komentarz do „Analizy aktywności publikacyjnej i grantowej PAN”

Słowo wprowadzenia

Zacząć należy od słów podziękowania pod adresem Autorów Analizy. Prowadzona przez Prezesa PAN wykonawczyni tego zadania, Pani Dr Katarzyna Stec, nie tylko wykonała pracę tytaniczną, ale przede wszystkim bardzo potrzebną [por. zakończenie niniejszego tekstu].

Pan Prezes Jerzy Duszyński jest doskonale świadomy, iż instytuty naukowe PAN (dalej IPAN, odpowiednio np. IPANy) muszą się stale doskonalić, a niektóre mogą wymagać nie tyle nawet poprawy jakości ich pracy co głębokiej naprawy strukturalnej. To Prezes odpowiada za osiągnięcie wymienionego celu przez społeczność wszystkich IPANów. Dzieła takiego doskonalenia nie można zacząć inaczej, niż dokonując oceny obecnej kondycji IPANów.

Taką ocenę właśnie otrzymaliśmy i jest to fakt o wielkiej wadze dla przyszłości IPANów i całej Akademii. Analiza jest pierwszym krokiem na drodze do zaproponowania sposobów dalszego doskonalenia albo wyraźnej poprawy jakości pracy IPANów. Zdaniem piszącego te słowa, drugim krokiem na tej drodze musi być pogłębienie już otrzymanej Analizy, wyciągnięcie z pogłębionej analizy właściwych wniosków i tym sposobem stworzenie warunków do dokonania kroku ostatniego i najważniejszego, czyli rozpoczęcia procesu stałego doskonalenia jakości pracy IPANów.

Piszący ten komentarz jest jego jedynym autorem. Nie konsultowałem mojego komentarza z nikim. Piszę go jako osoba znająca IPANy, nota bene wieloletni były dyrektor jednego z IPANów (to nieistotne, którego), osoba prywatna siebie tylko reprezentująca, która lepiej czy gorzej całe swoje naukowe życie przepracowała w IPANach i dla której niewiele jest wartości ważniejszych niż pomyślna przyszłość polskiej nauki i w jej służbie IPANów. Piszę ten komentarz także jako osoba, która większość swojego zawodowego życia spędziła badając naukowe aspekty analizy danych oraz uprawiając analizę danych na potrzeby naukowe, ale w pewnych okresach prowadziła również praktyczne analizy dla takich podmiotów jak wyższa Szkoła Oficerska Sił Powietrznych w Dęblinie, PKN Orlen, General Motors, czy Bank of America.

W dalszym ciągu nie stronię od mocnych sformułowań, ale na pewno pisanych z szacunkiem dla dzieła już wykonanego. Mam nadzieję, że ostrość sformułowań czyni je jaśniejszymi i przez to ułatwia ich ocenę oraz możliwość skorzystania z ich treści w przyszłości.

Komentarze ogólne

1. Analiza dotyczy zaledwie 69 instytutów, powinna zatem być dogłębna, np. znacznie głębsza niż oceny ewaluacyjne przygotowywane przez MNiSW (do tej pory KEJN, odtąd KEN). Tak niestety jeszcze w pełni nie jest. Jest to na razie analiza dość powierzchowna, niewolna od błędów metodologicznych.

2. Nawet jeśli dany wynik szczegółowej oceny jest poprawny i niesie informację, to nie jest poddany żadnej analizie interpretacyjnej, a zatem nie jest jasne, co z niego wynika. Szczególnie drastycznym przykładem jest brak analizy przyczyn niewykorzystania przez polską naukę sum wykładanych przez państwo na finansowanie projektów UE – wszyscy znają sytuację, bez znalezienia jej przyczyn informacja jest prawie bezwartościowa (ciekawe jest tylko odnotowanie stosunkowo dobrych wyników IPANów, ale i w tym przypadku brak głębszej analizy tych wyników, m.in. zbadania, jakiego typu projekty pozostają w zasięgu IPANów, a może i całej polskiej nauki, i dlaczego). Inna sprawa, że ten problem jest daleko ogólniejszej natury niż specyficzne problemy IPANów.

Przykład dotyczący ocen szczegółowych związanych z aktywnością IPANów: Wskazuje się IPANy publikujące dużo i często w wysoko punktowanych czasopismach; wskazuje się też instytuty publikujące dużo ale rzadko w wysoko punktowanych czasopismach, przy czym jednocześnie niemało publikacji w wysoko punktowanych czasopismach przypada w tych IPANach na jednego pracownika naukowego. Co to oznacza, jak te fakty łącznie ocenić, jak wytłumaczyć różnice między obydwoma typami IPANów? Brakuje też rozróżnienia między instytutami bardzo dobrymi i dobrymi dzięki wyjątkowo dużej aktywności niewielu najlepszych pracowników, którzy z nawiązką rekompensują „lenistwo” większości, a instytutami bardzo dobrymi jako całość (tej wady nie ma planowana ocena ewaluacyjna MNiSW: po pierwsze pod uwagę bierze się tylko 3N publikacji i po drugie każdy pracownik będzie miał jedynie 4 „sloty” publikacyjne; inna sprawa, że spłaszczy to oceny wyjątkowo dużo i dobrze publikujących jednostek naukowych).

Jeszcze jeden przykład: Fakt niskich wskaźników dotyczących patentowania wynalazków IPANów woła o identyfikację przyczyn, zresztą przynajmniej częściowo łatwych do znalezienia.

3. Bardzo poważnym błędem metodologicznym Analizy jest oparcie się na publikacjach indeksowanych jedynie przez bazę Scopus. Dlaczego akurat ta i tylko ta baza, gdy wiadomo, że jest niekompletna (np. ma bardzo poważne braki w obszarze nauk humanistycznych i poważne w obszarze nauk fizycznych). Poważnym błędem metodologicznym jest także oparcie się na kryterium FWCI dla jednego ustalonego okresu aktywności IPANów i wzięcie pod uwagę jedynie średniej z wartości FWCI dla kolejnych lat z tego okresu (p. uwagi szczegółowe).

Oparcie się jedynie na bazie Scopus jest tym bardziej niezrozumiałe, że polskie jednostki naukowe swój dorobek publikacyjny (i inny) przedstawiają w PBN, a zatem to baza PBN

powinna być punktem wyjścia, następnie rewizji jej danych i przejścia do właściwej analizy. Łatwiejszym rozwiązaniem, zresztą przyjętym przez MNiSW byłoby uwzględnienie innych baz publikacji i cytowań. Najbardziej właściwe byłoby wstrzymanie się z nawet ograniczonym rozpowszechnieniem Analizy do chwili ukazania się nowego rankingu czasopism i prac konferencyjnych przygotowywanego przez MNiSW (KEN) i (częściowe) przeanalizowanie zgodności między bazami użytymi i bazą MNiSW.

4. Rankingi instytutów w ramach Wydziałów (podział na ponadprzeciętne itd.) nie mają potrzebnego uzasadnienia merytorycznego - nie jest podane żadne ściśle określone kryterium, wedle którego owe podziały zostały dokonane.

5. W pewnych przypadkach stwierdzono wyraźną rozbieżność między ostatnimi ocenami ewaluacyjnymi MNiSW a wynikami sugerowanymi przez Analizę. To powinno było zaalarmować Autorów Analizy. Ocena MNiSW jest wolna od wyraźnych błędów metodologicznych, a zatem wyniki sugerowane przez Analizę powinny skłonić Autorów do sprawdzenia celności Ich metod analitycznych.

6. Nie wolno ferować wyroków, gdy nie ma się pełnych danych o aktywności niektórych IPANów.

7. Zaczynają napływać z IPANów komunikaty o błędach w danych, na których oparto analizy.

8. Niedoskonałość Analizy skłania do uznania jej rozpowszechnienia poza gronem ekspertów i może dyrektorów IPANów za przedwczesne.

Uwagi szczególne:

A.

Field-Weighted Citation Impact (FWCI) jest zdecydowanie gorszym wskaźnikiem niż odpowiednio wyskalowany wskaźnik cytowalności prac danego instytutu w ostatnich latach. FWCI bierze pod uwagę cytowania w ciągu 3 następnych lat po roku ukazania się danej pracy. A co jeśli np. opublikowana 6 lat temu praca zaczęła być masowo cytowana począwszy od czwartego roku po jej opublikowaniu? Przynosząc temu instytutowi zasłużoną sławę dzisiaj?

Informacje zawarte w artykułach bibliometrycznych (np. w *Journal of Infometrics*) oraz w materiałach SciVal pozwalają dokładnie określić sposób obliczania wartości wskaźnika FWCI dla konkretnej publikacji. Z informacji tych należy chyba też wnosić, że wartość tego wskaźnika dla zbioru publikacji wydanych w danym roku (wartość FWCI dla danego roku) jest średnią z wartości FWCI dla wszystkich publikacji z tego roku. Z kolei sumaryczna wartość wskaźnika FWCI dla okresu paroletniego (wieloletniego) obliczana jest jako średnia z wartości FWCI dla kolejnych lat z tego okresu.

W takiej sytuacji nie dziwi, że wartości wskaźnika FWCI dla kolejnych lat zachowują się bardzo niestabilnie. SciVal w jednym ze swoich przewodników podkreśla fakt owej niestabilności i radzi nie posługiwać się jedynie tym wskaźnikiem. Ale jeśli tak, to oparcie się w Analizie jedynie na średniej wartości wskaźnika FWCI dla ustalonego, paroletniego okresu nie ma sensu, ponieważ składowe wartości tej średniej „skaczą” z roku na rok. Sama wartość średnia nie daje wiarygodnej informacji o analizowanym procesie owych skoków.

Średnia wartość ciągu składowych, corocznych wartości w żaden sposób nie odzwierciedla zmienności wartości tych składowych. Nie tylko nie mówi, jak duże są „odchylenia” corocznych pomiarów FWCI od ich średniej (w górę czy w dół), ale czy te odchylenia są np. mniej więcej takie same co do ich wielkości (obojętne, w górę czy w dół – losowo, raz w górę, raz w dół). A bez tych informacji właściwe zinterpretowanie otrzymywanych wyników zbiorczych nie jest możliwe (zwłaszcza wtedy, gdy losowe odchylenia od średniej nie zachowują się z roku na rok podobnie, na przykład dlatego, że coroczne wartości FWCI wykazują – co powinno być uznane za pożądane – rosnący trend).

Co gorsze, wybranie właściwego okresu pomiarów corocznych wartości FWCI jest z praktycznych względów ... niemożliwe. Jak wspomniano, wskaźnik FWCI mierzy liczbę cytowań w ciągu 3 następujących lat po roku ukazania się danej pracy. Ale to znaczy, że jeśli się weźmie pod uwagę okres 2008 – 2017 albo 2013 – 2017 (jak w Analizie), to FWCI dla roku 2017 (znowu jak w Analizie) będzie obliczany dla minimum 8 (jeśli praca ukazała się 31.12.2017, a pamiętajmy, że wartości FWCI zostały podane dla 24.08.2018) i maksimum 20 miesięcy (jeśli praca ukazała się 1.01.2017).¹ Dopiero FWCI dla roku 2014 będzie liczony dla pełnych trzech kolejnych lat po ukazaniu się pracy. Z kolei wcześniejsze wartości FWCI nie uwzględniają cytowań z lat ostatnich.

Podsumowując, rzeczywiście trudno uznać pojedynczy (okresowy) wskaźnik FWCI za niosący wiele informacji. Powyższe nie ma być argumentem za zupełnym odrzuceniem obliczania wartości FWCI, ale ma pokazać, że wyciąganie właściwych wniosków na podstawie tego i każdego innego kryterium oceny nie jest sprawą trywialną.

W swoim liście do Pana Prezesa Jerzego Duszyńskiego, Prof. Włodzisław Duch (UMK) tak m.in. napisał poruszając problem interpretacji wskaźnika FWCI oraz podziału obszarów badawczych na dyscypliny:

„Jak wiadomo nie ma idealnych sposobów oceny, ale martwi mnie powszechne używanie field-weighted citation index (FWCI). W teorii to sensowne - normalizacja w danej dziedzinie. W praktyce oceny stabilności takiego indeksu dla małych próbek rzędu 100

¹ Uwzględniany w Analizie wskaźnik FWCI za rok 2017 nie niesie żadnej informacji o cytowalności prac danego instytutu. Jeśli dany instytut zatrudniałby tylko jednego pracownika naukowego, taki FWCI wynosiłby z dużym prawdopodobieństwem 0. Jeśli kilku, to jego wartość losowo rozkładałaby się w przedziale od 0 do 1. A jeśli parudziesięciu, to z dużym prawdopodobieństwem jego wartość byłaby bliska jedności. Praktycznie bez względu na jakość instytutu.

publikacji to +/- 25% błędu a dla 50 to 65%, więc oceny małych jednostek (nie wspominając o ocenach indywidualnych) są mało wiarygodne. Jakie specjalności brane są pod uwagę? Scopus ma 27 głównych i 334 mniejszych, ale nawet w tych bardziej szczegółowych nie ma bioinformatyki, neuroinformatyki czy medycyny komputerowej, która sama ma ze 100 specjalności. Każda z nich ma inny profil cytowań, inne średnie zależne od popularności danej dyscypliny. Dlatego wnioski z ocen szczegółowych (Ruiz-Castillo, J., & Waltman, L. (2015), Field-normalized citation impact indicators using algorithmically constructed classification systems of science, *Journal of Informetrics*, 9(1), 102–117) są takie, że potrzeba paru tysięcy specjalności by można było zrobić dobrą normalizację. Dlatego nie warto przywiązywać wielkiej wagi do FWCI, zwłaszcza przy ocenie mniejszych jednostek.”

B.

Oceniając aktywność publikacyjną oraz jej wpływ na naukę światową należy pamiętać, że wszelkie możliwe wskaźniki z trudem, jeśli w ogóle, są w stanie docenić prace pionierskie i wybitne – trudne i wyprzedzające swój czas. Co znowu nie jest argumentem za odrzuceniem obliczania wartości różnych wskaźników bibliometrycznych, ale ma pokazać, że nie jest możliwa pełna ocena aktywności publikacyjnej instytutu jedynie na ich podstawie.

C.

Wskaźniki *Publications in Top Percentiles*, *Outputs in Top Percentiles* nie mają sensu, jeśli poza nimi nie ma wyników otrzymywanych przez ich podzielenie przez liczbę pracowników naukowych danego IPANu. Te drugie wyniki, otrzymywane po podzieleniu, są podane w Analizie, ale bez wyciągnięcia z nich żadnych wniosków – wnioski są wyciągnięte jedynie na podstawie wcześniej wymienionych wskaźników.

Uwaga końcowa

Raz jeszcze podkreślam, że już otrzymana Analiza jest dokumentem wartościowym i niezbędnym, i jednocześnie łatwo zidentyfikować sposoby jej udoskonalenia, a przede wszystkim jej pogłębienia.

Przygotowanie opartego na danych raportu wymaga zespołu większego niż jednoosobowy. Przygotowanie analizy dogłębnej wymaga wyjścia poza poleganie na magii liczb. Nawet dobrze skonstruowane kryterium liczbowe, odpowiadające na zadane pytanie, nie ma wielkiej wartości bez eksperckiej analizy przyczyn, które przynoszą taką, a nie inną wartość liczbową tego kryterium. Dogłębna ocena kondycji danej jednostki naukowej, czy ich zbioru, wskazująca tym samym warunki konieczne dla poprawy tej kondycji, nie może się obejść bez udziału człowieka.

Dopisane 6 lipca

- Materiały SciVal nie mówią tego wprost, ale należy domniemywać, że FWCI bez autocytowań rzeczywiście ich nie uwzględnia, gdy podaje liczbę cytowań prac, dla którego

się go oblicza, ale TYLKO W LICZNIKU WZORU definiującego FWCI. W mianowniku tego wzoru podaje średnią liczbę cytowań dla całego świata i podobnych prac z uwzględnieniem autocytowań. Przy okazji punktu *Outputs in Top Citation Percentiles* dokumentu wydawnictwa Elsevier pt. *Research Metrics Guidebook* napisane jest explicite: “The exclusion of self-citations affects only the entity, and not the the universe.”

Ale jeśli tak, to $FWCI = 1$ nie mówi o takiej samej cytowalności prac naszych (bez autocytowań) i tych ze świata.

- Wcześniej przygotowany tekst tego dokumentu przekazałem 3 lipca Panu Prezesowi Jerzemu Duszyńskiemu. Docierające do pracowników instytutów PAN informacje o błędach Analizy każą mi wycofać słowa pochwały pod adresem jej Autorów.