

EMANUEL KULCZYCKI

**PROCEDURY EWALUACJI JEDNOSTEK
PODSTAWOWYCH I INSTYTUCJI**

RAPORT VII



EMANUEL KULCZYCKI

**PROCEDURY EWALUACJI JEDNOSTEK
PODSTAWOWYCH I INSTYTUCJI**

RAPORT VII

SERIA RAPORTÓW CENTRUM STUDIÓW NAD POLITYKĄ PUBLICZNĄ UAM

POZNAŃ 2019

Procedury ewaluacji jednostek podstawowych i instytucji

Emanuel Kulczycki

Raport VII

Redakcja merytoryczna: prof. Marek Kwiek

Publikacja finansowana w ramach programu Ministra Nauki i Szkolnictwa Wyższego pod nazwą „DIALOG” w latach 2016-2019 na podstawie umowy nr 0021/DLG/2016/10 z dnia 20 grudnia 2016 r.

Centrum Studiów nad Polityką Publiczną UAM
Poznań 2019



SPIS TREŚCI

1.	Streszczenie	5
2.	Wprowadzenie do problematyki raportu	6
3.	Kontekst – instytucjonalna ewaluacja nauki w Polsce	8
4.	Systemy instytucjonalnej ewaluacji nauki w innych krajach	13
5.	Badania nad systemami instytucjonalnej ewaluacji nauki	17
6.	Zalecenia i kierunki dla polityki publicznej	22
7.	Literatura	34

1. STRESZCZENIE

Celem niniejszego raportu jest wskazanie wyzwań polityki naukowej w zakresie instytucjonalnej ewaluacji nauki w Polsce. W związku z realizacją tego zadania definiowana jest ewaluacja nauki, charakteryzowane są podstawowe sposoby oraz momenty oceny. Raport pokazuje istotne problemy i wyzwania, przed którymi stoją twórcy systemu ewaluacji nauki w Polsce oraz prezentuje doświadczenia innych krajów europejskich mogące służyć za inspirację lub punkt wyjścia do doskonalenia instrumentów polskiej polityki naukowej.

W raporcie zostały zaprezentowane istotne informacje niezbędne do zrozumienia kontekstu oceny nauki w Polsce. Instytucjonalna ewaluacja została wdrożona w Polsce na początku lat 90. ubiegłego wieku. W 1991 roku Komitet Badań Naukowych zaproponował pierwszą wersję instytucjonalnej ewaluacji nauki, która przez ćwierć wieku uległa wielu zmianom, aby obecną formę przyjąć w 2017 roku.

Przez blisko trzy dekady funkcjonowanie ewaluacji nauki w Polsce sprawiło, że naukowcy i instytucje przyzwyczały się do sytuacji bycia ewaluowanym, co natomiast przyczyniło się do wielu pożądanых (m.in. wzrost produktywności) i niepożądanых (m.in. pojawienie się punktozy) efektów. Udoskonalenie tego systemu wymaga zatem nie tylko wykorzystania najlepszych i najbardziej adekwatnych metod oceny, ale również uwzględnienia obecnej sytuacji i statusu nauki w Polsce. Należy pamiętać, że rekomendowane w raporcie rozwiązania nie mogą być proponowane tylko dlatego, że sprawdziły się w innych krajach. Ważne jest również to, aby zaistniały przesłanki do tego, że proponowany system będzie implementowany w podobnych warunkach początkowych.

W raporcie zostało zidentyfikowanych osiem obszarów (wyzwań), w których konieczna jest interwencja polityki publicznej, aby nowy model instytucjonalnej oceny nauki mógł służyć wzmacnianiu pozycji nauki i Polski na arenie międzynarodowej. Te wyzwania to:

1. Cel instytucjonalnej oceny nauki – nowy model ewaluacji wymaga jasnego zdefiniowania celu ewaluacji, który nie będzie ograniczony jedynie do dystrybucji funduszy.
2. Transparentność procesu i wyników ewaluacji – jednym z kluczowych problemów obecnie funkcjonujących rozwiązań jest niski poziom zaufania

- do samego procesu i ich wyników. Zmiana w tym obszarze może zostać osiągnięta poprzez zwiększenie transparentności całego procesu i wyników ewaluacji.
3. Gromadzenie i analiza danych naukometrycznych – dotychczas stosowany system teleinformatyczny nie pozwala na wykorzystanie wcześniej zgromadzonych danych, a dane ostatecznie przetwarzane nie są agregowane na poziomie pojedynczego naukowca, co uniemożliwia prowadzenie rzetelnych analiz naukometrycznych.
 4. Respektowanie różnorodności dyscyplinarnej – wyzwaniem nowego modelu jest przejście od ewaluacji podstawowych jednostek organizacyjnych uczelni do ewaluacji dyscyplin naukowych w uczelni, co pozwoli na respektowanie różnic dyscyplinarnych w zakresie produkowania i upowszechniania wiedzy naukowej.
 5. Ocena ekspercka a ocena parametryczna – nowy model powinien udoskonalać i zmniejszać przeregulowanie rozwiązania parametrycznego przy jednoczesnym zwiększeniu wagi oceny eksperckiej.
 6. Ocena wszystkich a ocena liderów – ewaluacja uczelni w ramach dyscyplin powinna brać pod uwagę wszystkich pracowników, a nie tylko liderów.
 7. Ocena wpływu społecznego – obecny model w sposób niewystarczający umożliwia ewaluację pozaakademickiego wpływu badań naukowych, dlatego niezbędne jest rozszerzenie oceny eksperckiej o ocenę wpływu społecznego badań.
 8. Ocena publikacji naukowych – obecny model w różny sposób ocenia podstawowe kanały publikacji (tj. książki i czasopisma), co uniemożliwia utrzymywanie balansu w ocenie różnych dyscyplin i deprecjonuje punktowo nauki humanistyczne.

W raporcie znaczenie każdego z tych obszarów zostało omówione, a następnie – na podstawie dotychczasowych badań i doświadczeń z innych krajów wykorzystujących podobne krajowe rozwiązania – zostały zaproponowane zalecenia i kierunki dla polityki naukowej w Polsce. W ostatniej części raportu zostały zaproponowane sposoby monitorowania i ewaluowania wprowadzanych zmian.

Słowa kluczowe: ewaluacja nauki, dyscypliny naukowe, naukometria, ocena parametryczna

2. WPROWADZENIE DO PROBLEMATYKI RAPORTU

Ewaluacja nauki to ocena działalności naukowej oraz jej efektów. Ewaluację nauki można skategoryzować ze względu na przedmiot ewaluacji (np. pojedyncza książka, indywidualny naukowiec, instytucja naukowa, kraj) nazywanym również poziomem agregacji, ze względu

na moment ewaluowania (ewaluacja *ex ante* oraz *ex post*) oraz metodę ewaluacji (ocena koleżeńską, ocena ekspercka, ocena bibliometryczna). Wynik ewaluacji może być zaprezentowany w formie opisowej lub liczbowej, która umożliwi stworzenie rankingu.

Podstawowym poziomem agregacji w ewaluacji nauki jest pojedynczy naukowiec, którego działalność i efekty prac są podstawą tego, co określamy mianem efektów prac działalności naukowej. Takie efekty wypracowywane przez grupę naukowców zatrudnionych w jednej instytucji będziemy nazywać efektami prac działalności naukowej danej instytucji. Chociaż w wielu przypadkach ewaluacja efektów prac danej instytucji jest prostą sumą efektów wypracowanych przez naukowców zatrudnionych w tej instytucji, to jednakże każdy poziom agregacji w ewaluacji nauki ma swoje specyficzne metody i narzędzia, które nie powinny być przenoszone na inne poziomy (np. w ewaluacji w ramach awansów naukowych nie powinny być wykorzystywane metody służące ewaluacji instytucji naukowych).

Przedmiotem niniejszego raportu jest instytucjonalna ewaluacja nauki, która najczęściej dotyczy uczelni oraz instytutów badawczych, wydziałów, katedr czy grup badawczych. Od ponad trzech dekad instytucjonalna ewaluacja nauki jest jednym z kluczowych instrumentów polityki publicznej w większości krajów europejskich. Ewaluacja służy wielu celom, spośród których na plan pierwszy wysuwają się przede wszystkim: (1) Diagnoza, tj. ocena działalności naukowej i umiejscowienie tych wyników na tle innych instytucji, regionów czy krajów oraz (2) Dostarczanie dowodów dla polityki publicznej, czyli ewaluacja działalności naukowej często połączona z rangowaniem produktywności naukowej, której wyniki służą organizowaniu, planowaniu i finansowaniu działalności naukowej.

Podstawą ewaluacji instytucjonalnej są zrealizowane efekty działalności naukowej, np. publikacje naukowe, patenty, wdrożenia (produktywność naukowa danej instytucji). Jest to zatem ewaluacja typu *ex post* – w odróżnieniu od ewaluacji związanej z oceną zamiarów czy planów naukowych (np. opisanych we wniosku grantowym), która określana jest mianem ewaluacji typu *ex ante*.

Instrumenty do ewaluacji działalności naukowej można podzielić na dwie główne kategorie: wskaźniki naukometryczne bazujące przede wszystkim na metodach ilościowych oraz ocenę ekspercką, która oparta jest na jakościowej ocenie danego wycinka działalności naukowej przez pojedynczego eksperta lub panel ekspertów.

Wskaźniki naukometryczne można podzielić na dwie grupy. Pierwszą grupą są wskaźniki bibliometryczne (np. Impact Factor, indeks Hirscha), czyli wskaźniki mówiące o relacjach pomiędzy piśmiennictwem naukowym (np. jak często dany artykuł jest cytowany w innych artykułach naukowych). Drugą grupę stanowią wskaźniki odnoszące się do niepublikacyjnych aspektów działalności naukowej i efektów jej prac (np. nakłady na działalność badawczo-rozwojową, współpraca lokalna i międzynarodowa, struktura zatrudnienia w jednostkach naukowych).

Ocena ekspercka dokonywana jest przez ekspertów, którzy oceniają dokonania kraju, regionu, instytucji czy czasopisma naukowego. W ocenie eksperckiej mogą być również wykorzystywane wskaźniki naukometryczne, lecz pełnią one jedynie funkcję pomocniczą. Szczególnym przypadkiem oceny eksperckiej jest ocena koleżeńska (ang. *peer review*), w ramach której naukowiec ocenia dokonania, publikacje czy aktywność innego naukowca.

Ewaluacja nauki jest jednym z podstawowych instrumentów systemów finansowania nauki opartych o produktywność (ang. *performance-based research funding system* [PRFS]). W ciągu ostatnich trzech dekad w wielu krajach europejskich wprowadzono PRFS-y, a wraz z nimi krajowe systemy ewaluacji nauki. Pierwszy system – oparty przede wszystkim na ocenie eksperckiej – został uruchomiony w Wielkiej Brytanii w 1986 roku. Również Polska, wprowadzając ocenę instytucji naukowych w latach 1990–1991, stała się jednym z pierwszych krajów, w których system podziału dotacji i funduszy na badania związany był z wynikami naukowymi finansowanych jednostek naukowych.

Polski PRFS, zwany potocznie parametryzacją, w swoim pierwotnym kształcie był oparty na ocenie eksperckiej, aby w kolejnych latach ewoluować w stronę niezwykle mocno sparametryzowanego systemu opartego na wielu wskaźnikach naukometrycznych. Przemiany tego systemu wynikały z wielu powodów (dewaluacja ocen, brak zaufania do ekspertów, informatyzacja procesu oceny) i przyczyniły się do częściowego braku zaufania do samego procesu ewaluowania działalności i wyników działalności naukowej w Polsce.

3. KONTEKST – INSTYTUCJONALNA EWALUACJA NAUKI W POLSCE

Polski krajowy system ewaluacji nauki narodził się wraz z początkiem głębokich reform strukturalnych w sektorze nauki i szkolnictwa wyższego rozpoczętych na początku lat 90. ubiegłego wieku. Przemiany te przyczyniły się m.in. do gwałtownego umasowienia edukacji wyższej, rozwoju sektora prywatnego czy też utraty przez część nauk (głównie humanistyki i nauk społecznych) międzynarodowej widoczności mierzonej prestiżowymi publikacjami. Równolegle do tych transformacji przemianom podlegał sam system instytucjonalnej ewaluacji nauki, który zapoczątkowano w 1990 roku. Już rok później Komitet Badań Naukowych zaproponował pierwszą wersję modelu (zob. Tabela 1) oceny instytucjonalnej opartej na ocenie eksperckiej, która miała służyć kategoryzacji jednostek naukowych w oparciu o poziom ich doskonałości naukowej. Doskonałość naukową w różnych jej odmianach można rozumieć jako nieustanne dążenie do jak najlepszych wyników naukowych, nieustanne umiędzynarodowienie sektora nauki oraz komunikowanie efektów prac naukowych w najważniejszych kanałach publikacyjnych.

Tabela 1. Wersje modelu instytucjonalnej oceny nauki w Polsce

Wersja modelu	Rok wdrożenia	Rok przeprowadzenia oceny	Główna metoda oceny
Wersja pierwsza	1991	1991–1992	Ocena ekspercka
Wersja druga	1998–1999	1999, 2001, 2005, 2010	System parametryczny oparty na publikacjach
Wersja trzecia	2012	2013, 2017	System parametryczny oparty na publikacjach

Ocenę przeprowadzono w 1991 i 1992 roku. W kolejnych latach – ze względu na dewaluację przyznanych kategorii naukowych – zaproponowano drugą wersję systemu opartą na systemie parametrycznym, który miał uczynić ocenę bardziej obiektywną i niezależną.

Trzecia wersja najbardziej rozbudowanego systemu parametrycznego została wdrożona w 2012 roku i nazywa się Kompleksową Ewaluacją Jednostek Naukowych. Według tego systemu została przeprowadzona ocena w 2013 i w 2017 roku. Szczegółowy opis aktualnej wersji systemu oraz jego zasad funkcjonowania można odnaleźć w kilku publikacjach (Antonowicz i Brzeziński 2013; Kulczycki 2017a; Kulczycki, Drabek i Rozkosz 2015; Kulczycki, Korzeń i Korytkowski 2017). Dyskusja nad sposobem oceniania i konsekwencjami samej parametryzacji toczyła się głównie w organach prasowych środowiska akademickiego, tj. w „Forum Akademickim” i „PAUzie”, lecz liczba tekstów naukowych poruszających tę tematykę wciąż pozostaje ograniczona.

Obecna Kompleksowa Ewaluacja Jednostek Naukowych jest przeregulowana i niedopasowana do oceny różnych grup nauk czy też respektowania różnic w prezentacji wyników w różnych dyscyplinach. W niniejszym podrozdziale opisane zostały najważniejsze problemy i wyzwania, które muszą zostać rozwiązane, aby instytucjonalna ewaluacja nauki w Polsce została usprawniona oraz zyskała akceptację większej części środowiska naukowego. Diagnoza prezentowana jest na podstawie aktualnych badań prowadzonych w obszarze ewaluacji nauki i polityki naukowej w Polsce.

WYZWANIE 1: CEL INSTYTUCJONALNEJ OCENY NAUKI

Fundamentalnym problemem instytucjonalnej oceny nauki w Polsce jako instrumentu polityki naukowej jest brak jasno zdefiniowanego celu takiej oceny. Rekonstrukcja aktów prawnych pozwala stwierdzić jedynie, że podstawowym celem jest kategoryzacja podstawowych jednostek naukowych ze względu na ich

aktywność naukową oraz na tej podstawie dystrybucja strumieni finansowania. Jednakże uznanie ewaluacji nauki jedynie jako instrumentu dystrybucji środków nie pozwala na uzasadnienie ogromnego nakładu pracy i czasu, które musi poświęcić środowisko akademickie na przygotowanie dokumentacji, raportowanie i ocenianie aktywności ponad stu tysięcy polskich naukowców. Oznacza to, że doskonalenie istniejącego rozwiązania musi być rozpoczęte od wskazania celu instytucjonalnej oceny nauki wykraczającego poza prostą dystrybucję funduszy, którą można byłoby przeprowadzać na podstawie zdecydowanie mniej czasochłonnych algorytmów.

WYZWANIE 2: TRANSPARENTNOŚĆ PROCESU I WYNIKÓW EWALUACJI

Wyniki przeprowadzanych do tej pory cykli ewaluacji były publikowane w formie ministerialnych komunikatów i aktów prawnych. Jednakże główni zainteresowani – tj. dziekani, rektorzy, naukowcy – dowiadywali się jedynie, jaką kategorię naukową otrzymała poszczególne instytucja lub – jak miało to miejsce w 2013 i 2017 roku – jaki wynik uzyskała w każdym z czterech głównych kryteriów. Środowisko naukowe i obywatele nie wiedzą, jakie osiągnięcia przedstawiły polskie instytucje do oceny, jak te aktywności i wyniki naukowe zostały ocenione, ilu polskich naukowców nie wykonało żadnej pracy w ciągu czterech lat itd. To sprawia, że transparentność samego procesu jest bardzo niska i z zewnątrz przypomina wpadanie danych do „czarnej skrzynki” i wyskakiwanie z niej wyników. Taka konstrukcja instytucjonalnej ewaluacji nauki zdecydowanie obniża zaufanie do samego procesu. Dlatego wyzwaniem stojącym przed nowym systemem jest takie komunikowanie i informowanie o całym procesie (sposobie prowadzenia go oraz publikowaniu wyników ewaluacji), aby zwiększyć jego transparentność i zaufanie naukowców do samej ewaluacji.

WYZWANIE 3: GROMADZENIE I ANALIZA DANYCH NAUKOMETRYCZNYCH

Polska jest jednym z pierwszych krajów, który wdrożył PRFS. Jednocześnie praktycznie nie gromadzi się na jej gruncie danych naukometrycznych w taki sposób, który pozwoliłby – przede wszystkim głównym aktorom polityki naukowej (w tym ministerstwu odpowiedzialnemu za sprawy nauki) – prowadzenie analiz naukometrycznych. Bez takich analiz nie można prowadzić polityki naukowej opartej na dowodach, jak również projektować lepszych wersji systemów ewaluacji nauki, ponieważ nie jest możliwe modelowanie i przewidywanie potencjalnych konsekwencji zmian. Oznacza to, że dotychczasowe decyzje o zmianach były i są podejmowane w głównej mierze w oparciu o „intuicję decydentów” bądź szczątkowe dane pozyskane z samej kategoryzacji jednostek naukowych. Dlatego jednym z wyzwań udoskonalenia procesu ewaluacji nauki jest stworzenie i finansowanie

instytucji (niezależnej od organów finansujących naukę), która będzie dostarczała obywatelom i aktorom polityki naukowej wszechstronnej wiedzy o polskiej nauce, jej wynikach, wpływie i widoczności w świecie.

WYZWANIE 4: RESPEKTOWANIE RÓŻNORODNOŚCI DYSCYPLINARNEJ

Obecne rozwiązanie, w ramach którego ocenia się podstawowe jednostki organizacyjne (np. wydziały, instytuty Polskiej Akademii Nauk), sprawia, że różnice dyscyplinarne w zakresie sposobów ogłaszania wyników naukowych (waga różnych typów publikacji w różnych kontekstach dyscyplinarnych, produktywność, odrębne kultury cytowań) nie są respektowane. Wprowadzone w 2017 roku rozwiązanie w kategoryzacji, tzw. jednostki niejednorodne, w żaden sposób nie poprawiły tej sytuacji. Oznacza to, że wyzwaniem stojących przed nowym rozwiązaniem jest takie zaprojektowanie podstawowej jednostki oceny (ang. *unit of assessment*), które nie będzie opierało się na strukturze organizacyjnej uczelni, lecz reprezentowanej przez daną grupę uczonych dyscyplinie naukowej.

WYZWANIA 5: OCENA EKSPERCKA A OCENA PARAMETRYCZNA

Każdy system instytucjonalnej oceny nauki wykorzystuje jedno lub dwa podejścia, tj. ocenę ekspercką lub ocenę parametryczną. Obecne rozwiązanie w zakresie trzech głównych kryteriów oceny odnoszących się do dorobku publikacyjnego, potencjału naukowego i zewnętrznego finansowania opiera się na parametrach ilościowych. Natomiast kryterium czwarte jest oparte na ocenie eksperckiej. Wyzwaniem jest odpowiednie zbalansowanie tych dwóch podejść. Celem zbalansowania jest wdrożenie bardziej adekwatnych kryteriów oceny, co pozwoli na wzmocnienie obiektywności oceny, jak również zwiększenie obszaru, który podlega ocenie eksperckiej przy założeniu większej transparentności tego procesu i włączenia w niego ekspertów z międzynarodowego grona badaczy. Jednocześnie kluczowym zadaniem stojącym przed nowym rozwiązaniem jest zmniejszenie liczby parametrów, gdyż dotychczasowe rozwiązanie jest przeregulowane, a wiele parametrów jest ze sobą skorelowanych. Oznacza to, że środowisko naukowe poświęca czas na raportowanie zbędnych i niepotrzebnych informacji z perspektywy celu ewaluacji. Nowe rozwiązanie natomiast nie ma służyć raportowaniu i ocenianiu wszelkiej aktywności naukowej, lecz ma ewaluować doskonałość naukową w zgłoszonych aktywnościach i wynikach.

WYZWANIA 6: OCENA WSZYSTKICH A OCENA LIDERÓW

W każdym krajowym systemie nauki pojawia się jedno z dwóch rozwiązań: albo (1) oceniamy wszystkich pracowników wchodzących w skład jednostki oceny (w przypadku polskiego rozwiązania mowa zatem o wydziale czy też instytucji), czyli od każdego

pracownika oczekujemy odpowiedniej liczby dorobku, albo (2) oceniamy całą jednostkę, wyznaczając górną granicę dorobku, który może zostać przez nią zgłoszony – liczba pracowników wypracowujących ten dorobek nie ma znaczenia. Polskie rozwiązanie oparte na wskaźniku $3N-2No$ oznacza w uproszczeniu, że 100-osobowy wydział może zgłosić 300 publikacji naukowych (nie jest ważne to, czy każdy ze 100 pracowników zgłosi 3 publikacje, czy też 10 pracowników – tzw. liderów – zgłosi po 30 publikacji). Takie rozwiązanie ma wiele zalet (np. docenienie liderów), jak i wad (np. nie mobilizuje wszystkich pracowników do pracy). Dlatego nowe rozwiązanie wyznaczając nowy cel instytucjonalnej oceny nauki, stoi również przed wyzwaniem określenia, które rozwiązania rozkładające „ciężar wypracowania dorobku” zostanie przyjęte.

WYZWANIE 7: OCENA WPŁYWU SPOŁECZNEGO

Ewaluacja nauki w wielu krajach zaczyna obejmować również ewaluację wpływu społecznego (ang. *impact* lub *societal impact*), kulturowego, ekonomicznego i naukowego. Systematyczna ocena wpływu została wdrożona na szeroką skalę na poziomie krajowym w 2014 roku w Wielkiej Brytanii w ramach Research Excellence Framework (REF 2014) – obecnie do jej implementacji przygotowują się kolejne kraje (m.in. Holandia i Norwegia). Doskonałość naukowa mierzona była dotychczas przez pryzmat oddziaływania wyłącznie w obszarze badań naukowych poprzez wskaźniki prestiżu i wpływu (np. Impact Factor czy wskaźnik Hirscha). Wyzwaniem, przed którym stoi nowy system w Polsce, jest zaproponowanie takiego sposobu ewaluacji wpływu społecznego, które pozwoli adekwatnie ukazać rolę i sposób oddziaływania poszczególnych dyscyplin. Jest to szczególnie ważne dla nauk humanistycznych i społecznych, których tzw. kultura cytowań sprawia, że ich wpływ nie jest łatwo mierzalny wskaźnikami bibliometrycznymi.

WYZWANIE 8: OCENA PUBLIKACJI NAUKOWYCH

Kryterium ewaluacji posiadającym największą wagę oraz znaczenie są publikacje naukowe. W dotychczasowych zasadach ewaluacji obowiązywały dwa niespójne rozwiązania: artykuły w czasopismach oceniano na podstawie Wykazu czasopism punktowanych (artykuł jest wart tyle, ile czasopismo, w którym jest opublikowany), natomiast książki oceniano na podstawie kryteriów formalnych (książka warta jest zawsze tyle samo, jeśli spełnia kryteria formalne). Problemy z odpowiednią wyceną publikacji naukowych połączone są z kluczowym – z perspektywy wpływania polityki naukowej na produktywność polskich uczelni – zagadnieniem podziału punktów za publikacje. Wszystkie te trzy elementy, tj. ocena artykułów, ocena książek i podział punktów za publikacje wymagają udoskonaleń w nowym systemie.

4. SYSTEMY INSTYTUCJONALNEJ EWALUACJI NAUKI W INNYCH KRAJACH

Instytucjonalna ocena nauki realizowana jest na różnych poziomach (np. krajowym, regionalnym, lokalnym) wynikających z uwarunkowań danego systemu polityki naukowej oraz organizacji nauki i szkolnictwa wyższa. Sytuacja polska sprawia, że warto przyglądać się przede wszystkim takim rozwiązaniom, które realizowane są na poziomie krajowym dla wszystkich instytucji sektora szkolnictwa wyższego i nauki. Natomiast rozwiązania lokalne, które – choć funkcjonować mogą dobrze i skutecznie – nie są skalowalne na poziom krajowy (dotyczy to na przykład rozwiązań stosowanych w landach w Niemczech czy wręcz systemów tworzonych na potrzeby pojedynczych uczelni w Szwajcarii).

Poniżej zostały zaprezentowane rozwiązania z trzech krajów europejskich. Kryterium doboru było następujące. *Research Excellence Framework*, czyli rozwiązanie brytyjskie jest nie tylko najstarszym i najlepiej opisanym rozwiązaniem, ale również jest jedynym systemem w pełni opartym na ocenie eksperckiej. Z tego powodu w wielu debatach publicznych w Polsce przywołuje się rozwiązanie brytyjskie jako wzór do naśladowania. Drugie rozwiązanie, tj. Model Norweski, jest również jednym z najlepiej przebadanych i, co istotne, stanowiło inspirację dla innych krajowych rozwiązań (m.in. w Danii i Finlandii). Rozwiązanie trzecie, czyli dotychczasowy model ewaluacji zaimplementowany w Czechach jest jedynym – oprócz systemu polskiego – rozwiązaniem z Europy Środkowo-Wschodniej, które doczekało się badań i analiz opublikowanych w głównych czasopismach poświęconych naukometrii oraz badaniom nad nauką i szkolnictwem wyższym.

RESEARCH EXCELLENCE FRAMEWORK

W dyskusjach nad polskim systemem ewaluacji nauki często przywołuje się model brytyjski jako punkt odniesienia lub wręcz jako niedościgniony wzór ewaluacji opartej na ocenie koleżeńskiej (*peer review*). W 1986 roku Wielka Brytania jako pierwsza wdrożyła model instytucjonalnej oceny nauki, który w 1992 roku uzyskał swoją dojrzałą formę w ramach *Research Assessment Exercise*. Ostatnia edycja oceny instytucjonalnej prowadzona była w 2014 roku pod szyldem *Research Excellence Framework* (REF 2014). Brało w niej udział 154 instytucji, które wysłały informacje o dorobku 52 061 naukowców i ich 191 150 publikacjach¹. Kolejna edycja planowana jest w 2021 roku.

Celem prowadzenia ewaluacji w Wielkiej Brytanii jest zapewnienie przejrzystości publicznej inwestycji w naukę i pokazywanie płynących z niej korzyści. Dlatego od ewaluacji w 2014 roku ocenia się nie tylko wyniki naukowe, ale również ich wpływ społeczny na świat pozaakademicki.

¹ REF 2014, <http://www.ref.ac.uk/2014/> [31.10.2017].

Ocena w ramach REF 2014 odbywała się w ramach 36 paneli dyscyplinarnych (*unit of assessment*), podzielonych na cztery panele główne (A, B, C, D) reprezentujące cztery główne obszary nauki. Ewaluacja dyscyplinarna w Wielkiej Brytanii w sposób istotny odróżnia się od ewaluacji prowadzonej w Polsce, gdzie ową jednostką oceny (*unit of assessment*) nie jest dyscyplina w obrębie uczelni, ale sama jednostka administracyjna, czyli podstawowa jednostka organizacyjna (np. wydział czy instytut Polskiej Akademii Nauk).

Każda jednostka naukowa przygotowuje raport składający się z pięciu głównych elementów (Wróblewska 2017): (1) informacji o pracownikach zgłoszonych do ewaluacji, (2) informacji o publikacjach pracowników zgłoszonych do ewaluacji – na każdego pracownika można było zgłosić do czterech publikacji z okresu 2008–2013, (3) informacje potwierdzające „wpływ społeczny” osiągnięty przez daną jednostkę, (4) informacje o doktoratach przyznanych w danej jednostce oraz zewnętrznym finansowaniu, (5) opis środowiska badawczego w jednostce.

Ocena dorobku danej jednostki była oceniana przez panele eksperckie złożone głównie przez brytyjskich naukowców oraz – w mniejszości – naukowców spoza Wielkiej Brytanii i osoby spoza środowiska akademickiego. Ocena jakości publikacji stanowiła 65% oceny końcowej jednostki. Ocena wpływu społecznego stanowiła 20%, a 15% przeznaczono na ocenę środowiska badawczego w kategoriach jego aktywności i trwałości. Ostateczny wynik oceny wyrażany jest w pięciostopniowej skali.

MODEL NORWESKI

Obok rozwiązania brytyjskiego jednym z najpopularniejszych modeli instytucjonalnej ewaluacji nauki jest Model Norweski, który został wprowadzony na poziomie krajowym w Norwegii w 2004 roku, z efektami finansowymi dla norweskich instytucji badawczych w 2005 roku. W następnych latach model ten zaaplikowany został w Belgii (Flandrii), Danii i Finlandii, a jego implementacja rozważana jest również przez inne kraje (m.in. Portugalię). Model Norweski składa się z trzech podstawowych elementów (Sivertsen 2017):

1. Krajowej bazy danych zawierającej kompletną reprezentację sprawdzalnych i zgodnych z kryteriami indeksowania informacji o publikacjach naukowych ze wszystkich dyscyplin.
2. Wskaźnika publikacji o różnych wagach punktowych, który ma umożliwić porównywanie różnych praktyk publikacyjnych.
3. Modelu finansowania instytucji opartego na produktywności, który przyznaje niewielką część bezpośredniej dorocznej subwencji instytucjonalnej na podstawie udziału instytucji w ogólnej liczbie punktów za publikacje.

Główną zaletą Modelu Norweskiego, którą podnosi się w literaturze przedmiotu, jest adekwatne i umożliwiające porównywanie reprezentowanie w bazie publikacji różnych obszarów nauki. Jest to szczególnie istotne dla reprezentantów nauk humanistycznych i społecznych, których publikacje (szczególnie w krajach nieanglojęzycznych) nie są odpowiednio reprezentowane w największych bazach takich jak Web of Science Core Collection czy Scopus. Jest to możliwe dzięki funkcjonowaniu bazy CRISTin (Current Research Information System in Norway), w której zbierane są informacje o publikacjach dostarczanych przez same instytucje naukowe.

Istotnym rozwiązaniem w kontekście polskiego systemu ewaluacji nauki jest wprowadzenie listy czasopism oraz listy wydawców książek naukowych. W Norwegii każdy z tych głównych kanałów publikacji jest podzielony na dwa poziomy (w zależności od jakości naukowej czasopisma/wydawcy i jego zasięgu oddziaływania). Każdy z poziomów przyporządkowuje inną liczbę punktów za publikację (zob. Tabela 2).

Tabela 2. Dwa poziomy kanałów publikacji w Modelu Norweskim

Typ publikacji	Poziom 1	Poziom 2
Artykuł w czasopiśmie	1	3
Rozdział w monografii	0.7	1
Monografia	5	8

Liczba punktów na danym poziomie (np. za artykuł w czasopiśmie na poziomie 1) jest taka sama dla każdego z czasopism czy wydawców znajdujących się na danym poziomie. W Polsce w ramach „Wykazu czasopism punktowanych” obowiązuje zróżnicowana skala w obrębie każdego z poziomów – jeśli by tak nazwać części A, B, C tego wykazu. Kolejną istotną różnicą Modelu Norweskiego w stosunku do polskiego rozwiązania jest relacja liczby punktów za monografię do liczby punktów za artykuł. Za artykuł w czasopiśmie na poziomie 1 otrzymuje się pięć razy mniej punktów niż za monografię wydaną u wydawcy z poziomu 1. W Polsce za artykuł można otrzymać dwa razy więcej punktów niż za monografię (wyluczając wąską grupę tzw. monografii wybitnych, których punktacja jest równa punktacji artykułu opublikowanego w najlepszym czasopiśmie).

CZESKA METODA INSTYTUCJONALNEJ EWALUACJI NAUKI

Czeska instytucjonalna ewaluacja nauki (*Metodika hodnocení*) w wielu aspektach przypomina model polskiej ewaluacji nauki. Chociaż wciąż nie jest pewne, w którą

stronę pójdzie nowe czeskie rozwiązanie, to jednak ewaluacja czeskiej ewaluacji przeprowadzona w 2015 roku pokazuje, jakie konsekwencje może mieć parametryczny system oceny nauki w krajach Europy Środkowo-Wschodniej (Good, Vermeulen, Tiefenthaler i Arnold 2015). Co istotne, Czechy są jedynym krajem z naszego regionu, który posiada bardzo podobny PRFS do polskiego rozwiązania. Dlatego warto przyjrzeć się, jakie konsekwencje mogą przynieść rozwiązania stosowane w podobnym systemie, aby uniknąć powtórzeń podobnych błędów.

Czeski model ewaluacji uwzględnia różne typy dorobku badawczego i przypisuje im konkretną liczbę punktów, które się sumują i nie są ważone. Jest to istotna różnica w stosunku do rozwiązania polskiego: w Polsce bowiem 5 pkt przyznane za opublikowanie rozdziału w monografii nie jest równe 5 pkt przyznanym w kryterium eksperckim za osiągnięcia naukowe. Owe punkty są ważne w obrębie kryterium, a następnie dodatkowe wagi przykłada się ze względu na rodzaj jednostki naukowej (np. wydział lub instytut Polskiej Akademii Nauk). W Czechach natomiast każdy punkt waży tyle samo. Co więcej, punkty są „tłumaczone” bezpośrednio na korony czeskie. Z powodu sprowadzania wszelkiej aktywności naukowej do wspólnego mianownika czeski system ewaluacji nauk nazywany jest potocznie „młynkiem do kawy”, do którego wpadają różne typy aktywności (różne wielkości i jakości ziarna kawy), a wypadają te same korony (zmielona kawa). W ten sposób model czeski pełni dwie funkcje: jest to system ewaluacji działalności naukowej oraz system dystrybuowania funduszy do instytucji sektora nauki i szkolnictwa wyższego.

W ocenie uwzględniane są publikacje (artykuły, książki, prace konferencyjne), jak i wyniki badań stosowanych (m.in. patenty). Tak jak w Polsce rozwiązanie to stosowane jest do wszystkich rodzajów jednostek naukowców oraz do wszelkich dyscyplin. Rozpiętość punktowa waha się od 4 do 500 punktów. Najmniej punktów można uzyskać za artykuł opublikowany w czeskim czasopiśmie naukowym (4 pkt), a najwięcej za publikację w *Nature* lub *Science* oraz za patent europejski (EPO), amerykański (USPTO) lub japoński (500 pkt).

Interesujące z perspektywy polskich rozwiązań są konsekwencje czeskiego modelu i sposób adaptowania się kadry czeskich naukowców do ewaluacji. Część środowiska wypracowała strategię zwiększania liczby osiągniętych przez siebie punktów poprzez uczynienie dorobku „niepunktowanego” „punktowanym”. Typowym przykładem jest publikowanie podręczników jako monografii naukowych czy też ponowne publikowanie książek i niewskazywanie, że jest to wznowienie. Co więcej, praktyki te odnoszą się również do czasopism, pod które podszywają się inne formy publikacji (np. serie *working papers*). System czeski był i jest przedmiotem krytyki wielu czeskich naukowców głównie ze względu na sposób oceny (ściśle punktowy) i automatyczne przekładanie liczby punktów na finanse.

5. BADANIA NAD SYSTEMAMI INSTYTUCJONALNEJ EWALUACJI NAUKI

Wiele krajów (np. Czechy, Finlandia, Włochy, Norwegia, Polska) wprowadziło różne wersje PRFS-ów i uczyniło je częścią systemów ewaluacji nauki (Hicks 2012; Schneider, Aagaard i Bloch 2016). Systemy ewaluacji nauki są przedmiotem badania przede wszystkim w tych krajach, w których takie rozwiązania funkcjonują.

Ewaluacja produktywności naukowej opiera się głównie na ocenie publikacji, a inne aspekty działalności naukowej są traktowane jako działania i efekty towarzyszące powstaniu i publikowaniu publikacji. Hicks (2012) podkreśla, że każdy PRFS z czasem robi się coraz bardziej skomplikowany, ponieważ są to systemy dynamiczne, których rozwój przyspieszany jest przez implementację nowych sposobów oceny takich jak indeks Hirscha czy alternatywne metryki (Sugimoto, Work, Larivière i Haustein 2017). Im bardziej takie systemy stają się skomplikowane, tym bardziej wskaźniki i metryki w nim używane mogą być ogrywane, czyli przestać pełnić rolę mierników doskonałości naukowej (Kulczycki 2017b; de Rijcke, Wouters, Rushforth, Franssen i Hammarfelt 2016). To sprawia natomiast, że model ewaluacji nauki w danym kraju zaczyna mieć problemy z jego legitymizacją w środowisku akademickim. Dlatego też, jak sugerują autorzy „Manifestu Lejdejskiego” (Hicks, Wouters, Waltman, de Rijcke i Rafols 2015), wskaźniki wykorzystywane w ocenie nauki powinny być regularnie aktualizowane i badane pod kątem możliwego ich ogrywania. W ten sposób zmienianie i doskonalenie PRFSów w sposób nieustanny mierzy się z takimi wyzwaniami jak: utrzymywanie równowagi pomiędzy oceną ekspercką a oceną parametryczną, tworzenie jednego rozwiązania dla wszystkich grup nauk lub rozwiązania „szytego na miarę” dla każdej dyscypliny naukowej czy też decyzja o tym, jakie aktywności naukowe uwzględniać w ocenie (Kulczycki, Korzeń i Korytkowski 2017). Jednocześnie decyzje te należy przekuć na konkretne rozwiązania, które opierają się na różnych sposobach liczenia i punktowania publikacji naukowych (Schneider, Aagaard i Bloch 2014) czy też tworzenia rankingów kanałów publikacji w oparciu o ocenę ekspercką lub liczenie cytowań (Saarela, Kärkkäinen, Lahtonen i Rossi 2016).

Kraje takie jak Australia, Brazylia, Francja, Włochy czy Polska wdrożyły w ramach krajowych systemów ewaluacji rankingi czasopism naukowych, w których reprezentowane są wszystkie dyscypliny naukowe (Ferrara i Bonaccorsi 2016; Haddow i Genoni 2010; Vanclay 2011). Rankingi te służą do wskazywania, które publikacje są uwzględniane w przeprowadzanej ocenie. Chociaż publikacje stanowią najważniejszy komponent wszystkich PRFS-ów, to jednakże również inne aktywności, takie jak przyznane granty, patenty, zaangażowanie edytorskie czy udział w konferencjach naukowych, są uwzględniane. Wyniki badań pokazują, że dobór tych różnych aktywności do uwzględnienia w ocenie ma istotny wpływ

na wyniki ewaluacji (Abramo i D'Angelo 2015; Ancaiani i in. 2015; Good, Vermeulen, Tiefenthaler i Arnold 2015; Langfeldt, Bloch i Sivertsen 2015).

W literaturze przedmiotu związanej z krajowymi systemami instytucjonalnej oceny nauki istotne miejsce zajmują obecnie badania nad konsekwencjami funkcjonowania tych systemów. Poniżej prezentuję najważniejsze ustalenia i wątki podejmowane w tych badaniach. Należy również podkreślić, że osobną (rozległą) grupę zagadnień stanowią kwestie związane z oceną i liczeniem publikacji naukowych – te zagadnienia są szczegółowo omówione w Raporcie VIII pt. „Ocena publikacji jako instrument polityki naukowej” (Kulczycki 2019).

Poniżej zostały opisane badania dotyczące Australii, Norwegii oraz – w sposób rozbudowany – Polski. Inne systemy funkcjonujące na poziomie krajowym również zostały przebadane w zakresie konsekwencji i wdrażania i funkcjonowania. Jednakże celem omówienia jest skorzystanie z doświadczeń i analiz przeprowadzonych w takich warunków, które pozwolą odnieść się do nich z perspektywy polskiej sytuacji. Dlatego też analiza konsekwencji REF-u 2014 nie byłaby w tym miejscu uzasadniona, chociaż jego główne założenia zostały zaprezentowane w poprzednim rozdziale, aby pokazać, jak funkcjonuje ten – często obecny w polskiej debacie – model. Jednocześnie badania prowadzone w Finlandii i Danii pokrywają się z badaniami Modelu Norweskiego, gdyż rdzeń tych rozwiązań jest wspólny.

AUSTRALIA

Butler (2003) pokazała na przykładzie analizy australijskiego systemu ewaluacji nauki, że zasady oceny mogą mieć istotne konsekwencje dla produktywności naukowej. W badaniach opublikowanych w 2003 roku australijska badaczka pokazywała, że liczba publikacji z australijską afiliacją w ramach Science Citation Index, na podstawie którego wylicza się wskaźnik Impact Factor dla czasopism z nauk twardych, wzrósł o 25% w ciągu poprzedzającej dekady. Jednocześnie jednak wpływ Australii mierzony indeksem cytowań istotnie zmalał w stosunku do innych krajów. Wynikało to z przemian, które wymusił system ewaluacji: naukowcy publikowali więcej artykułów, ale gorszej jakości. Zdaniem Butler błędem polityki naukowej było wprowadzenie systemu finansowania nauki w oparciu o produktywność. Badania te były wielokrotnie cytowane i stały się istotnym narzędziem krytyki PRFS-ów.

Takie spojrzenie na australijską sytuację zostało skrytykowane przez van den Besselaara, Heymana i Sandströma (2017), którzy – po kilkunastu latach od pierwotnej publikacji Butler – opublikowali na łamach *Journal of Informetrics* inne spojrzenie na konsekwencje australijskiej metody. Zdaniem autorów z perspektywy czasu i gdy spojrzymy na dłuższy okres, nie jest uzasadnione twierdzenie,

że przy zwiększonej produktywności zmalał wpływ mierzony cytowaniami. Co więcej, autorzy podkreślają, że nawet używając zrekonstruowanych dowodów z wykorzystanych przez Butler danych, nie da się tego twierdzenia utrzymać. Van den Besselaar, Heyman i Sandström pokazują, że jeśli naukowcy są stymulowani przez system ewaluacji, stają się bardziej produktywni i tworzą lepsze pomysły (a w konsekwencji artykuły).

NORWEGIA

Wprowadzony w 2005 roku Model Norweski oparty na różnicowaniu kanałów publikacji na dwa poziomy został przebadany pod kątem jego wpływu na produktywność publikacyjną norweskich naukowców (Schneider, Aagaard i Bloch 2016). W odróżnieniu od sytuacji australijskich uczelni w Norwegii produktywność publikacyjna wzrosła, a wpływ mierzony cytowaniami pozostał stabilny. Co więcej, w Norwegii nie zauważono zwiększonej aktywności publikacyjnej w czasopismach niższej jakości. Jedną z przyczyn tej różnicy zachodzącej między konsekwencjami rozwiązań australijskich i norweskich jest odmienny model punktowania publikacji oraz dzielenia punktów za publikacje. W efekcie te różnice sprawiły, że w Australii aktywność publikacyjna rosła, lecz miała ona odwrotny efekt na wpływ mierzony cytowaniami oraz na zwiększenie liczby publikacji w czasopismach gorszej jakości. Model Norweski inspirowany częścią rozwiązań australijskich wdrożył jednak różnicowanie kanałów publikacji (dwa poziomy), co pozwoliło odwrócić niepożądane efekty. Analiza wieloletniego funkcjonowania modelu nie wykazała obniżonych wskaźników cytowań.

POLSKA

Badania nad polskim systemem instytucjonalnej oceny nauki koncentrują się na dwóch obszarach, tj. na kategoryzacji jednostek naukowych oraz ocenie kanałów publikacji.

W badaniach w ramach pierwszego obszaru Kulczycki (2017a) pokazał, że w ocenie instytucji prowadzonej do 2013 roku najważniejszymi wyzwaniami było: (1) utrzymanie równowagi pomiędzy oceną różnych grup nauk, (2) mieszany sposób liczenia publikacji oraz (3) używanie zasad instytucjonalnej oceny nauki do oceny indywidualnych naukowców. Analiza wyników kategoryzacji jednostek naukowych z 2013 roku pokazała, że system jest przeregulowany i składa się ze zbyt wielu skorelowanych ze sobą parametrów (Kulczycki, Korzeń i Korytkowski 2017). W szczególności zbędne – z perspektywy technicznej sprawności systemu i osiągania zbliżonych wyników oceny – jest całe Kryterium II w kategoryzacji jednostek naukowych, w ramach którego mierzony jest potencjał naukowych danej jednostki (np. liczba uprawnień doktorskich czy habilitacyjnych).

Badania nad sposobem liczenia i punktowania publikacji w instytucjonalnej ocenie nauki (Kulczycki, Drabek i Rozkosz 2015) pokazały, że wybór sposobu liczenia (całościowe lub ułamkowe) publikacji w istotny sposób wpływa na obraz produktywności polskich naukowców. Zmienia się on bowiem w zależności od tego, czy do liczenia produktywności używamy „publikacji” czy też „zgłoszenia publikacyjnego” – jedna publikacja napisana przez dwóch autorów z dwóch uniwersytetów pozostaje jedną publikacją, lecz stanowi dwa zgłoszenia publikacyjne w ramach ewaluacji. Okazuje się, że różnice w praktykach publikacyjnych w różnych dyscyplinach naukowych są na tyle istotne, że używając do analiz zgłoszeń publikacyjnych, otrzymujemy obraz zniekształcony – na ważność danej publikacji w procesie kategoryzacji wpływa w ten sposób przede wszystkim liczba autorów z różnych polskich jednostek, a przestaje mieć znaczenie całkowita liczba autorów czy też ich miejsce w szeregu. Również wykorzystanie nieadekwatnych, czyli niestworzonych jako narzędzia bibliometryczne, rozwiązań takich jak listy ERIH (European Reference Index for the Humanities) sprawia, że obraz publikacyjny jest niepoprawny. Listy ERIH miały służyć do docenienia i wskazywania międzynarodowych publikacji polskich humanistów. Jednakże okazało się, że w okresie 2009–2012 ponad 75% publikacji polskich naukowców opublikowanych w czasopiśmie z list ERIH stanowiły publikacje w polskich czasopiśmie o najniższym zasięgu, tj. krajowym oddziaływaniu (Kulczycki, Rozkosz i Drabek 2015).

Polski model ewaluacji nauki oparty głównie na parametrach jest główną przyczyną wytworzenia się tzw. punktozy, którą Kulczycki (2017b) nazywa strategią w grze parametrycznej. Natomiast owa gra, to takie prowadzenie badań i publikowanie ich wyników (np. wybieranie „publikowalnych” problemów badawczych), które pozwala wpisywać się w priorytety i reguły funkcjonującego w danej społeczności akademickiej systemu ewaluacji nauki, któremu podlega dana społeczność akademicka. W związku z tym granie w grę oznacza funkcjonowanie w społeczności akademickiej, awansowanie, zdobywanie stanowisk poprzez odpowiednie realizowanie celów systemu ewaluacji nauki.

Jak pokazał Kulczycki (2017b), w polskim systemie ewaluacji nauki istnieją mechanizmy, które zabezpieczają przed takim negatywnym zjawiskiem, lecz nie oznacza to, że są one wystarczające. Pierwszym głównym mechanizmem jest ograniczenie liczby publikacji zgłaszanych przez jednostkę naukową do oceny – może ona zgłosić do oceny za okres czterech lat maksymalnie tyle publikacji, ile wynosi (w uproszczeniu) trzykrotność liczby jej pracowników. Określane jest to formułą $3N$, gdzie N to liczba pracowników danej jednostki. Oczywiście mechanizm ten działa tylko wówczas, gdy: (1) wszyscy pracownicy w jednostce publikują i (2) taka reguła jest „przekładana” na system oceny indywidualnego naukowca

w poszczególnych jednostkach naukowych. Siła oddziaływania tego mechanizmu jest osłabiona, gdyż chociaż suma publikacji nie może przekroczyć $3N$, to można zgłosić dowolną liczbę publikacji jednego pracownika (do limitu $3N$). Drugim mechanizmem jest ograniczenie liczby monografii i rozdziałów zgłaszanych do oceny – w ramach $3N$ publikacji jednostka może zgłosić ograniczoną liczbę monografii i rozdziałów. Najwięcej mogą zgłaszać jednostki naukowe z obszaru nauk humanistycznych i społecznych, gdzie 40% wszystkich publikacji mogą stanowić monografie.

Jednakże oprócz tych dwóch mechanizmów można wskazać takie rozwiązania, które sprzyjają wytwarzaniu punktozy. Jest to przede wszystkim definiowanie monografii poprzez kryteria formalne. Sprawia to, że każde wydawnictwo, zakład produkujący pieczątki czy punkt ksero może wydawać książki i wszystkie te książki będą punktowane w ramach polskiego systemu ewaluacji nauki. Opisane ograniczenie polegające na redukowaniu np. do 40% liczby monografii i rozdziałów zgłaszanych do oceny nie spełni swojej roli, jeśli wszystkie te monografie zostaną wydane w podrzędnych wydawnictwach. Kolejnym istotnym mechanizmem jest system podziału punktów za publikacje wieloautorskie. Jak pokazały badania nad rozwiązaniami australijskim i norweskim, wdrożenie nieodpowiedniej zasady podziału punktów może przyczynić się do zmniejszenia współpracy międzynarodowej oraz publikowania większej liczby artykułów, lecz w czasopiśmie o niewysokiej jakości.

Polski mechanizm podziału punktów za publikacje łączy ze sobą dwa podejścia: podobnie jak w Norwegii podział całościowy na poziomie jednostki naukowej oraz podział całościowy używanego w systemie flamandzkim czy szwedzkim. Oznacza to, że pojedyncza publikacja nie jest liczona wielokrotnie w jednej jednostce, lecz – z perspektywy krajowego systemu – może być liczona wielokrotnie w różnych jednostkach. To natomiast sprawia, że badaczom opłaca się publikować przede wszystkim z autorami spoza własnej jednostki. Jednocześnie przy takim podziale punktów oraz zasadach uznawania monografii najbardziej opłaca się publikować monografie wieloautorskie w podrzędnych wydawnictwach. Oczywiście autorzy pochodzą wówczas z różnych jednostek naukowych. I takie też zachęty do publikacji od quasiwydawnictw pojawiają się już powszechnie w polskim środowisku naukowym.

Dodatkowo należy wskazać na dwa dodatkowe problemy, które wynikają głównie z kondycji polskiego środowiska naukowego. Pierwszym jest brak zaufania do ekspertów skutkujący zaufaniem do obiektywności wskaźników. Transformacje społeczne i polityczne po 1989 roku pociągnęły za sobą zwiększenie braku zaufania do ekspertów (Jabłeczka i Lepori 2009). Podobną sytuację można zaobserwować w Czechach, gdzie używanie wskaźników parametrycznych

zamiast ocen eksperckich było powodowane pragnieniem „depolityzacji i depersonalizacji procesów oceny i finansowania” (Good, Vermeulen, Tiefenthaler i Arnold 2015: 102).

Warto też podkreślić, że w wielu dyscyplinach naukowych w Polsce środowisko akademickie jest zbyt małe, aby wdrożyć – bez otworzenia się na zagranicznych ekspertów – dobre praktyki i rozwiązania w zakresie systemu eksperckiego. Ten brak wiary w ekspertów staje się mechanizmem legitymizującym całkowicie parametryczny system oceny, który postrzegany jest wówczas jako obiektywny, czyli niezależny od ekspertów, ich zainteresowań czy woli politycznej. To natomiast sprawia, że system parametrów rozbudowywany ponad wszelkie zdroworozsądkowe granice pozwala na uwzględnianie wszelkich przejawów działalności pracowników naukowych: doskonałość naukowa owych przejawów nie ma znaczenia, liczy się jedynie to, że w systemie istnieją parametry pozwalające uwzględnić i zmierzyć owe przejawy.

Drugim problemem, jeszcze niezbyt głęboko przebadanym, lecz uświadamianym przez szerokie grono naukowców, jest wykorzystywanie mechanizmów instytucjonalnej oceny nauki do oceny pojedynczych naukowców (Kulczycki 2017a). W polskim systemie awansów naukowych tzw. punkty ministerialne, czyli punkty przyznawane jednostkom za publikacje ich pracowników w procesie parametrycznej oceny ich działalności, nie są uwzględnione w żadnym akcie prawnym. Jednakże rektorzy, dziekani, dyrektorzy instytutów badawczych zaczęli „tłumaczyć” reguły gry parametryzacyjnej na zasady przyznawania awansów. W zarządzeniach rektorów czy dziekanów zaczęły się pojawiać sformułowania nie tylko o rekomendowanej, lecz niekiedy wręcz o wymaganej liczbie punktów, które musi zebrać dany kandydat na stopień naukowy lub tytuł, aby mógł mieć przeprowadzone postępowanie awansowe: może to być 150, 200 czy 250 punktów.

6. ZALECENIA I KIERUNKI DLA POLITYKI PUBLICZNEJ

Polityka naukowa w Polsce może sprostać wyzwaniom wskazanym w rozdziale trzecim niniejszego raportu prezentującego kontekst instytucjonalnej ewaluacji nauki w Polsce, bazując na rozwiązaniach sprawdzonych w innych krajach oraz proponując takie zmiany, które będą dostosowane do sytuacji w krajowym systemie. Oznacza to, że rozwiązania te nie mogą pomijać kontekstu historycznego rzutującego na kształt struktury instytucjonalnej polskiej nauki oraz muszą uwzględniać profile kadry naukowej w Polsce: inny system ewaluacji stosuje się do już w większości umiędzynarodowionej kadry, a inny do kadry aspirującej do umiędzynarodowienia.

Poniżej dla każdego wskazanego wyzwania zostały zaprezentowane rekomendowane kierunki zmian dla polityki publicznej.

PROPOZYCJE ZMIAN

WYZWANIE 1: CEL INSTYTUCJONALNEJ OCENY NAUKI

Głównym celem instytucjonalnej oceny nauki jako instrumentu polityki naukowej powinno być wspieranie doskonałości w badaniach naukowych. Taki cel możliwy jest do osiągnięcia poprzez tworzenie odpowiednich warunków prawnych oraz instytucjonalnych na poziomie systemowym, a także tworzenie czytelnego systemu wspierania najlepszych badaczy, zespołów czy uczelni. Ewaluacja nauki ma nie tylko wskazywać, kim są i gdzie pracują najlepsi badacze, ale powinna również określać, jakie osiągnięcia uznawane są za doskonałe naukowe.

Celem badań naukowych jest pogłębianie wiedzy o świecie przyrody, techniki i kultury, a także zrozumienie występujących w nim zjawisk oraz zachodzących procesów. Doskonałość naukowa powinna przyświecać wszystkim prowadzonym badaniom, jednakże nie każdemu i nie w równym stopniu udaje się ją osiągnąć. W związku z tym system ewaluacji nauki powinien: (1) definiować doskonałość naukową i wyznaczać jej miary; oraz (2) tworzyć silne impulsy do kierowania się doskonałością w pracach badawczych. Oznacza to, że w ewaluacji nauki powinny być uwzględniane i doceniane wyłącznie najważniejsze osiągnięcia naukowe, a aktywności naukowe, które jedynie są środkiem do tego celu, nie powinny być raportowane.

Proponując instrumenty polityki naukowej w Polsce, należy mieć na uwadze sposoby mierzenia doskonałości naukowej wykorzystywane w świecie. Dlatego nowy model ewaluacji nauki powinien koncentrować się wyłącznie na efektach działalności naukowej (tj. publikacjach i patentach, finansowaniu zewnętrznym oraz wpływie społecznym) i poziomie ich doskonałości przy jednoczesnej rezygnacji z pomiaru pozostałych pośrednich aktywności.

WYZWANIE 2: TRANSPARENTNOŚĆ PROCESU I WYNIKÓW EWALUACJI

Zwiększenie transparentności procesu ewaluacji powinno zostać osiągnięte poprzez: (1) szerokie informowanie środowiska naukowego o trwających pracach nad kolejnymi wersjami oceny instytucjonalnej; (2) jak najwcześniejszym publikowaniu aktów prawnych definiujących kryteria i parametry oceny; (3) dbanie o nieprzeregulowanie systemów poprzez dokładanie nieistotnych – z perspektywy celu prowadzonej ewaluacji – parametrów; (4) pełne publikowanie wyników ewaluacji oraz informacji o zgłoszonych osiągnięciach (z wyjątkiem osiągnięć, które z mocy prawa są tajne bądź poufne).

Obecny system teleinformatyczny jest dostosowany do tego, aby wskazane powyżej informacje udostępniać. Oznacza to, że zwiększenie transparentności

procesu i wyników oceny jest decyzją polityczną, która powinna zostać podjęta, aby zwiększyć zaufanie do procesu instytucjonalnej ewaluacji nauki.

WYZWANIE 3: GROMADZENIE I ANALIZA DANYCH NAUKOMETRYCZNYCH

Obecne rozwiązanie w ewaluacji nauki sprawia, że te same dane bibliometryczne i naukometryczne są wielokrotnie wprowadzane przez naukowców i gromadzone przez jednostki naukowe. Nowe rozwiązanie powinno być oparte na systemie informatycznym, który sprawi, że dane będą jednokrotnie wprowadzane i wykorzystywane w wielu zastosowaniach (ankieta jednostki, postępowania awansowe, wnioski grantowe itd.). W tym celu należy wdrożyć rozwiązanie, które w maksymalnym stopniu będzie wykorzystywać trwale identyfikatory cyfrowe takie jak DOI, ORCID, ISBN, ISSN.

Oparcie ewaluacji instytucjonalnej na dyscyplinach naukowych wymaga gromadzenia danych na innym poziomie, niż miało to miejsce do tej pory. Dotychczas dane były gromadzone na poziomie jednostki naukowej, tzn. wiedzieliśmy, ile i jakie publikacje zostały afiliowane do danego wydziału. Jednakże tak zebrane dane nie pozwalają na ocenę w obrębie dyscyplin naukowych, a w konsekwencji na efektywną ewaluację nauki na poziomie krajowym. Dlatego kluczowe jest wykorzystanie identyfikatora ORCID będącego dla naukowca tym, czym jest ISBN dla książki – czyli niezmiennym i jednoznaczny identyfikatorem. Rozwiązanie oparte na zbieraniu informacji na poziomie pojedynczego pracownika pozwoli głównym aktorom polityki naukowej w Polsce na analizowanie w czasie rzeczywistym (a nie tylko periodycznie, co cztery lata) aktywności i działalności naukowej polskich instytucji i naukowców; dziedzin i dyscyplin oraz porównywanie polskich osiągnięć z innymi krajami. To natomiast może przysłużyć się podejmowaniu rozsądnych – kluczowych dla wiarygodności systemu – decyzji personalnych, to znaczy w sprawach mianowania do ciał eksperckich lub przedstawicielskich w oparciu o intersubiektywnie weryfikowany dorobek naukowy. Tylko przy tak gromadzonych danych jest możliwe wdrażanie i proponowanie nowych rozwiązań wraz z symulacją potencjalnych efektów.

WYZWANIE 4: RESPEKTOWANIE RÓŻNORODNOŚCI DYSCYPLINARNEJ

Dotychczasowy system instytucjonalnej ewaluacji nauki oceniał dorobek naukowców poprzez przypisywanie ich do wydziałów – najczęściej – niejednorodnych pod względem praktyk publikacyjnych. Wprowadzona w 2017 roku zasada wyróżniania naukowych jednostek niejednorodnych jako tych, których co najmniej 35% pracowników (i niekiedy publikacji) należy do innego obszaru nauki (nie dziedziny czy dyscypliny, ale właśnie „najszerzego” obszaru), jedynie usiłuje wprowadzać

mechanizmy kompensacyjne, ale nie jest w stanie rozwiązać problemu odmienności wzorów publikacyjnych. Dotychczasowy system, który można określić jako „jeden system pasujący do wszystkich dyscyplin”, nie pozwalał na utrzymywanie równowagi w ocenie różnych dyscyplin.

Niejednorodne jednostki to bowiem nie tylko takie jednostki, w których jedną połowę stanowią historycy (obszar nauk humanistycznych), a drugą prawnicy (obszar nauk społecznych). W rzeczywistości niejednorodne jednostki to również takie, w którym prawnicy (obszar nauk społecznych) pracują z ekonomistami (również obszar nauk społecznych). Obie grupy dyscyplin różni nie tylko przedmiot badań, ale również częstotliwość wybierania określonego typu publikacji (monografia, artykuł) oraz ich języka.

W ocenie nauki należy pamiętać, że różne dyscypliny charakteryzują się różnymi sposobami komunikacji i publikacji swoich wyników. Te różnice określa się najczęściej poprzez dwa główne pojęcia: kultur epistemicznych (różne dyscypliny w różny sposób określają zasady wytwarzania wiedzy naukowej, co przekłada się na różne typy publikacji czy rolę konferencji naukowej w danej kulturze) oraz kultur cytowań (struktura cytowań różnicuje dyscypliny).

W krajowym systemie nauki poszanowanie różnych kultur epistemicznych może odbywać się poprzez poszanowanie różnych kultur cytowań. To natomiast będzie osiągnięte poprzez ewaluowanie publikacji w ramach dyscyplin – wówczas każdy filozof będzie rzeczywiście oceniany na tle jedynie pozostałych filozofów, a każdy historyk będzie porównywany z innymi historykami itd.

Ta zmiana polegająca na poszanowaniu różnych sposobów komunikacji przez różne dyscypliny będzie miała szczególne znaczenie dla nauk humanistycznych i społecznych, w których ocenie równowaga była w szczególności zachwiana przez dwa czynniki: (1) najlepsze czasopisma humanistyczne nie mogły osiągnąć takiej samej punktacji jak czasopisma z innych obszarów nauki, (2) monografie były oceniane punktowo niżej niż artykuł w najlepszym czasopiśmie – ta sytuacja ponownie najbardziej doskwiera naukom humanistycznym.

W związku z powyższym respektowanie różności dyscyplinarnej i osiągnięcie balansu w zakresie metod oceny może być realizowane poprzez: (1) ocenę w ramach dyscyplin, a nie wydziałów; oraz (2) zbalansowanie zasad oceny publikacji.

WYZWANIA 5: OCENA EKSPERCKA A OCENA PARAMETRYCZNA

Jednym z najczęstszych postulatów przewijających się przez polską debatę jest zwiększenie udziału oceny eksperckiej w instytucjonalnej ocenie nauki, przy jednoczesnej krytyce tej samej oceny eksperckiej. Jak pokazuje przykład REF-u w Wielkiej Brytanii, model ekspercki może się sprawdzić tylko i wyłącznie przy wysokim

wskaźniku zaufania do ekspertów oraz ich dostępności. Dodatkowo należy podkreślić, że system ewaluacji nauki nie może być zbyt kosztowny. Dla porównania jedna edycja polskiej kategoryzacji jednostek naukowych w 2013 roku kosztowała 3 miliony złotych (nie licząc wynagrodzeń członków Komitetu Ewaluacji Jednostek Naukowych), natomiast jedna edycja oceny eksperckiej w Wielkiej Brytanii w 2014 roku kosztowała 33 miliony funtów: w tym 19 milionów na wynagrodzenia dla ekspertów. Jeśli do tego dołożymy 212 milionów funtów, czyli koszty – trudne do oszacowania, więc należy podchodzić z rezerwą do wyliczeń – związane z samą pracą naukowców przy zgłaszaniu dorobku, to widzimy, że przy takich nakładach na naukę w Polsce ocena ekspercka nie jest instrumentem łatwym do implementacji.

Dlatego też efektywniejszą drogą jest modyfikacja obecnego systemu parametrycznego. Na poziomie polityki naukowej jasno zdefiniowany cel ewaluacji nauki (wskazana wcześniej doskonałość naukowa) powinien przekładać się na używane kryteria i parametry. Obecne rozwiązanie ujednolicone dla wszystkich grup nauk oraz jego wypaczone zastosowania koncentrują się na mierzeniu produktywności, a nie na ocenie doskonałości naukowej prowadzonych badań. System oceny nauki powinien być narzędziem wspomagającym osiągnięcie celów wskazanych przez politykę naukową, np. zwiększanie doskonałości w nauce, umiędzynarodowienie czy też awans polskich uczelni w prestiżowych globalnych rankingach uniwersytetów. Nie wszystkie te cele da się łatwo zmierzyć: umiędzynarodowienie można zmierzyć np. liczbą naukowców z zagranicy pracujących w Polsce, jednakże ocena doskonałości w nauce wymaga już przyjęcia różnych wskaźników, które nie zawsze muszą być zaakceptowane przez środowisko naukowe. Tak samo awans w prestiżowych rankingach naukowych nie może być celem samym w sobie: twórcy rankingów poprzez odpowiedni dobór kryteriów (i ich nagłą zmianę) mogą wyrzucić dotychczas funkcjonującą klasyfikację (np. opartą na innych wskaźnikach) do góry nogami. Jednakże wszystkie te cele w obecnym systemie społeczno-gospodarczym są ze sobą powiązane, dlatego system parametryczny musi mieć je na uwadze. Rolą tego systemu powinno być również dostarczanie informacji dla polityki naukowej o osiągnięciach polskich uczelni i instytucji naukowych ze wszystkich dziedzin. Jednocześnie jednak musi on uwzględniać i dopasowywać narzędzia ewaluacji do specyfiki różnych grup nauk: w tym celu może np. korzystać z krajowych baz osiągnięć i publikacji oraz wspomagać rozwój takich narzędzi jak Polska Bibliografia Naukowa czy Polska Bibliografia Prawnicza.

Kluczowe jest to, aby balans pomiędzy oceną ekspercką a oceną parametryczną był utrzymany na tyle, na ile jest to możliwe ze względu na posiadane zasoby (zarówno finansowe, jak i zasoby dostępnych ekspertów).

Dlatego w pierwszym rzędzie rekomendowany jest rozwój systemu parametrycznego, którego nowa wersja powinna osłabiać mechanizmy wytwarzające punktozę i wzmacniać te, które je powstrzymują. Można to uzyskać na przykład: (1) przebudowując zasady budowy „Wykazu czasopism punktowanych” tak, aby indeksowane na nim były tylko najlepsze czasopisma; (2) tworząc katalog uznanych wydawców naukowych; czy też (3) modyfikując zasady podziału punktów za publikacje wieloautorskie – jednakże koniecznie w odniesieniu do specyfiki praktyk publikacyjnych w danej dyscyplinie naukowej. Te instrumenty zostały opisane w Raporcie VIII (Kulczycki 2019).

Jednocześnie należy zwiększyć wagę oceny ekspertów poprzez zastąpienie obecnego Kryterium IV (tzw. eksperckiego) oceną wpływu społecznego dokonywaną przy udziale ekspertów zagranicznych. Jest to niezbędny krok do zbudowania w Polsce takiej kultury ewaluacyjnej, która docelowo pozwoli jeszcze bardziej wzmocnić wagę oceny eksperckiej.

WYZWANIA 6: OCENA WSZYSTKICH A OCENA LIDERÓW

W obecnym rozwiązaniu o sile wydziału lub instytutu badawczego mogą decydować publikacje kilku naukowców. Jeśli nowy model ewaluacji ma mówić o sile i doskonałości naukowej całej dyscypliny w obrębie uczelni, to model oceny musi uwzględniać wszystkich pracowników, a nie jedynie liderów (do czego sprowadza się obecne rozwiązanie). Rekomendowanym rozwiązaniem jest zatem uwzględnianie 3N publikacji, przy czym każdy pracownik powinien zgłosić co najmniej jedną publikację i nie więcej niż 6 publikacji. Rozwiązanie to – proste od strony technicznej – będzie miało kluczowe konsekwencje nie tylko dla systemu ewaluacji, ale przede wszystkim dla sektora nauki i szkolnictwa wyższego, dając naukowcom i kadrom zarządzającym silne instrumenty motywowania (zarówno pozytywne, jak i negatywne).

WYZWANIE 7: OCENA WPLYWU SPOŁECZNEGO

Doskonałość naukowa była dotychczas definiowana (i mierzona) przez pryzmat oddziaływania wyłącznie w obszarze badań naukowych przez wskaźniki prestiżu czy wpływu (np. Impact Factor czy indeks Hirscha). Od lat 90. ubiegłego wieku większą wagę zaczęto jednak przykładac do wpływu badań naukowych na gospodarkę, której miarą jest wysokość przychodów generowanych z tytułu transferu wiedzy oraz wdrożeń. Uwzględnienie społecznego wpływu badań naukowych wpisuje się w koncepcje społeczeństwa oraz gospodarki opartej na wiedzy, które od końca lat 90. XX wieku wytyczają główny kierunek zmian w polityce wobec nauki i szkolnictwa wyższego. Pokazywanie społecznego wpływu badań naukowych wzmocnia pozycję szkolnictwa wyższego czy nauki w ogóle.

Uwzględnienie wpływu społecznego jako osobnego kryterium wymaga jednak zbudowania zupełnie nowego instrumentu lub użycia metody z innego systemu (lecz już przetestowanej, np. w brytyjskim REF-ie). Rozwiązanie brytyjskie, którego zmodyfikowaną wersję można byłoby wdrożyć również w Polsce, opiera się na studiach przypadków, przez co pozwala na rzetelny pomiar i ocenę różnorodnych aspektów wpływu. Stosowanie wskaźników bibliometrycznych nie jest w stanie uchwycić wpływu społecznego nauki na społeczeństwo i gospodarkę. Odnosi się to zwłaszcza do wpływu nauk społecznych i humanistycznych organizujących nasze myślenie o historii i świecie współczesnym. Ich wpływ społeczny, choć zasadniczo niekwestionowany, jest trudny do uchwycenia wyłącznie przy pomocy wskaźników numerycznych.

Ekspertka ocena wpływu społecznego wymaga paneli z udziałem międzynarodowych ekspertów, gdyż stwarza to większą szansę intersubiektywności i rzetelności formułowanych ocen. Dla nauki w Polsce (podobnie jak dla większości systemów ewaluacyjnych) punktem odniesienia jest przede wszystkim brytyjski REF, na którego potrzeby wpływ społeczny zdefiniowano jako „efekt, zmiana lub korzyść dla gospodarki, społeczeństwa, kultury, polityki publicznej czy sektora zdrowia, środowiska, jakości życia poza akademią” (Wróblewska 2017). Ważne jednak, aby kryterium wpływu odnosiło się do badań naukowych, a nie politycznej, publicystycznej czy rozrywkowej aktywności naukowców, a sposób ich sprawozdawania przez uczelnie pokazywał wyraźny związek pomiędzy wynikami badań (udokumentowanymi w formie publikacji) oraz intersubiektywnie weryfikowaną zmianą poza akademią.

Na potrzeby polskiej ewaluacji instytucjonalnej można wykorzystać zmodyfikowany model brytyjski, w którym wpływ społeczny był oceniany przez ekspertów na podstawie nadesłanych przez jednostki opisów przypadków wpływu. Jednostka w ramach dyscypliny mogłaby przysłać odpowiednią liczbą takich przypadków w zależności od liczby pracowników. Zaistnienie takiego wpływu musiałoby być podparte dowodami (np. publikacjami, raportami potwierdzającymi wykorzystanie wyników badań).

Co niezwykle istotne dla badań humanistycznych i społecznych, ważne jest to, aby umożliwić zgłaszanie przypadków wpływu badań zrealizowanych przed ewaluowanym okresem, lecz mających wpływ w okresie ewaluowanym. Innymi słowy, w kategoryzacji w 2021 roku opisy przypadków wpływu muszą być podparte dowodami opisującymi ten wpływ opublikowanymi w latach 2017–2020. Same badania, które wygenerowały ów wpływ mogły być zrealizowane wcześniej. Takie rozwiązanie pozwoli docenić długofalowe konsekwencje badań, co ma istotne znaczenie dla badań podstawowych.

WYZWANIE 8: OCENA PUBLIKACJI NAUKOWYCH

Publikacje naukowe można oceniać bibliometrycznie, biorąc za punkt wyjścia dwa podejścia: analizę cytowań pojedynczych publikacji oraz analizę kanałów publikacji, w którym dana publikacja się ukazała. Pierwsze podejście pozwoli nam powiedzieć np. że artykuł opublikowany 2 lata temu został zacytowany 10 razy. Drugie podejście wskaże nam, że dany artykuł został opublikowany w uznanym czasopiśmie, gdyż wskaźnik bibliometryczny tego czasopisma jest wyższy niż średnia wartość wskaźnika w tej dziedzinie. Oba rozwiązania mają istotne wady i zalety, które jednakże trzeba rozpatrywać przede wszystkim w kontekście możliwości ich wykorzystania w polskim systemie, w którym musimy ocenić aktywność publikacyjną ponad 100 tysięcy naukowców.

Model oparty na wyliczaniu cytowań dla każdej zgłoszonej publikacji jest dobry do ewaluacji indywidualnych naukowców czy zespołów prowadzonej w dłuższym okresie albo wręcz obejmujących ich całą karierę naukową. Rozwiązanie to sprawdza się: gdy (1) ograniczana jest liczba zdarzeń podlegających ocenie oraz podmiotów przedstawiających zdarzenia do oceny; (2) gdy celem jest wybór najlepszych/liderów, podczas gdy pozostała część aplikujących nie ma znaczenia; (3) gdy zastosowane kryterium pozwala różnicować badane jednostki (uczelnie) wzdłuż całego kontinuum. Natomiast do interwałowej oceny instytucji akademickich po prostu się nie nadaje. Kluczowa jest niewykonalność takiego rozwiązania: badania pokazują, że po 5 latach od publikacji w humanistyce cytowanych jest zaledwie ok. 10% prac, a w naukach twardych 60–70%. Oznacza to, że gdyby w następnym cyklu ewaluacji w 2021 roku wziąć do oceny najlepsze prace humanistów indeksowane w bazach w latach 2017–2020, okazałoby się, że praktycznie żadne nie są cytowane. Dlatego drugie podejście, oparte na ocenie kanałów publikacji, jest nie tylko dobre, ale i wykonalne w kontekście polskim.

Podstawową zasadą nowego modelu oceny publikacji powinna być tożsamość metod oceny wszystkich kanałów publikacji poprzez uznanie – funkcjonującej od ponad pół wieku w naukometrii oraz wdrożonej w systemach ewaluacji nauki m.in. w Danii, Finlandii, Norwegii i Hiszpanii – zasady dziedziczenia prestiżu (artykuł jest wart tyle, ile czasopismo, w którym jest opublikowany, a książka jest warta tyle, ile wydawnictwo ją wydające). Takie rozwiązanie jest możliwe do przyjęcia, ponieważ dotyczy instytucjonalnej oceny nauki na poziomie całych uczelni: dlatego nie można wykluczyć, że w dobrym czasopiśmie czy wydawnictwie mogą również pojawić się kiepskiej jakości publikacje i one też będą punktowane, podobnie jak w słabych czasopismach pojawią się prace na wysokim poziomie. W tym przypadku trzeba jednoznacznie odróżnić regułę od zdarzeń wyjątkowych, które biorąc pod uwagę skalę oceny instytucjonalnej, nie będą miały wpływu na jej końcowy efekt.

Sposób monitorowania i ewaluowania wprowadzanych zmian

Wyzwanie	Wskaźnik	Oczekiwane efekty	Sposób ewaluacji wdrożonej zmiany
<p>Cel instytucjonalnej oceny nauki</p>	<p>Definicja doskonałości naukowej oraz jej miary</p> <p>System zachęt (silne impulsy) do kierowania się doskonałością naukową</p>	<p>Jasno wskazana definicja doskonałości naukowej oraz ilościowe i jakościowe sposoby jej mierzenia zarówno w dokumentach programowych i kierunkowych tworzonych przez Ministerstwo Nauki i Szkolnictwa Wyższego, jak i w aktach prawnych regulujących instytucjonalną ewaluację nauki oraz awanse naukowe</p> <p>Uznawanie w instytucjonalnej ewaluacji tylko takich efektów działalności pracy naukowej, które charakteryzują się doskonałością naukową w najwyższym stopniu</p>	<p>Analiza aktów prawnych oraz sprawdzenie, w jaki sposób prezentowana definicja doskonałości naukowej oraz jej miary odpowiadają miarom używanym w polityce naukowej Komisji Europejskiej czy też kryteriom głównych rankingów uniwersytetów</p> <p>Analiza, w jaki sposób zasady instytucjonalnej ewaluacji nauki zabezpieczają przed punktozami oraz jakie zostały wprowadzone mechanizmy zachęt (np. czy została istotnie zwiększona liczba punktów za publikacje w czasopiśmie w stosunku do czasopiśm o zasięgu lokalnym)</p>
<p>Transparentność procesu oceny i wyników ewaluacji</p>	<p>Informowanie środowiska naukowego o sposobie ustalania zasad instytucjonalnej oceny nauki</p>	<p>Wzrost klarowności i skuteczności komunikatów tworzonych przez Ministerstwo Nauki i Szkolnictwa Wyższego, który może przełożyć się do celowo na zwiększenie wiedzy naukowców, dziekanów i rektorów na temat celu ewaluacji oraz jej zasad</p>	<p>Analiza czytelności i klarowności komunikatów tworzonych przez Ministerstwo Nauki i Szkolnictwa Wyższego oraz analiza liczby i zakresu spotkań informacyjnych oraz konsultacyjnych.</p>

Wyzwanie	Wskaźnik	Oczekiwane efekty	Sposób ewaluacji wdrożonej zmiany
Transparentność procesu oceny i wyników ewaluacji	Publikacja aktów prawnych definiujących kryteria i parametry oceny	Niezwłoczne poinformowanie środowiska naukowego o celu, zakresie, przedmiocie i sposobie instytucjonalnej oceny nauki	Analiza tego, czy opublikowanie aktów prawnych nastąpiło szybciej w stosunku do poprzedniego okresu ewaluacji oraz – co najważniejsze – czy akty prawne nie uległy wielu zmianom do momentu przybrania ostatecznego kształtu
	Usunięcie prerogulowania systemu	Zmniejszenie liczby kryteriów oraz parametrów, a w konsekwencji ilości raportowanej przez uczelnie i instytuty badawczej informacji na temat efektów działalności naukowej	Porównanie liczby kryteriów i parametrów ostatecznie uwzględnianych w 2021 roku w odniesieniu do liczby z 2017 roku
Gromadzenie i analiza danych naukowych	Opublikowanie wyników ewaluacji oraz informacji o zgłoszonych osiągnięciach	Jasne i klarowne poinformowanie środowiska naukowego o wynikach ewaluacji, jej szczegółowym przebiegu oraz o zgłoszonych osiągnięciach naukowych będących podstawą ostatecznej oceny	Analiza czy forma udostępnienia wyników w 2021 roku została w istotny sposób udoskonalona względem 2017 roku oraz czy informacje o zgłoszonych osiągnięciach będące podstawą oceny zostały ogłoszone
	Zbieranie danych na poziomie pojedynczego naukowca	Wdrożenie nowego systemu teleinformatycznego oraz obowiązkowego i trwałego identyfikatora naukowca, tj. ORCID	Analiza aktów prawnych w zakresie wymaganych i raportowanych informacji przez uczelnie i instytuty badawcze
Respektowanie różnorodności dyscyplinarnej	Ewaluowanie instytucji w ramach dyscyplin	Poszanowanie różnych kultur epistemicznych i kultur cytowań w obrębie dyscyplin naukowych	Analiza aktów prawnych, czy jednostka ewaluacji została przeniesiona z podstawą jednostki organizacyjnej na dyscyplinę naukową lub artystyczną w danej uczelni lub instytucie badawczym

Wyzwanie	Wskaźnik	Oczekiwane efekty	Sposób ewaluacji wdrożonej zmiany
Respektowanie różnorodności dyscyplinarnej	<p>Punktowanie czasopism z zakresu nauk humanistycznych</p> <p>Relacja punktów za monografie do punktów za artykuły w czasopismach</p>	<p>Zaprzestanie punktowego deprecjonowania artykułów w naukach</p> <p>Punktowe docenienie monografii</p>	<p>Analiza, czy czasopisma z nauk humanistycznych mogą otrzymać maksymalną możliwą punktację</p> <p>Analiza, czy za monografie wydane u najbardziej prestiżowych wydawców można uzyskać większą liczbę punktów niż za artykuły</p>
Ocena ekspercka a ocena parametryczna	Rozwój modelu oceny parametrycznej	<p>Oslabienie mechanizmów wytwarzających punktozę przy jednoczesnym wdrożeniu silnych bodźców zachęcających do kierowania się doskonałością naukową w publikowaniu wyników prac naukowych</p>	<p>Analiza, czy została zmniejszona liczba parametrów, którymi można było zdobywać małą liczbę, lecz możliwych do zsumowania punktów. Jednocześnie należy przebadać, czy zmianom uległy zasady oceny kanałów publikacji naukowych oraz zasady podziału punktów za publikacje wieloautorskie</p>
	Zmiana formy oceny eksperckiej przy jednoczesnym zwiększeniu jej wagi	Zwiększenie wagi oceny eksperckiej pozwoli rozpocząć budowanie takiej kultury ewaluacyjnej, która docelowo umożliwi jeszcze bardziej wzmocnić wagę oceny eksperckiej	Analiza zmiany wagi oceny eksperckiej w ostatecznej ocenie oraz formy i zakresu oceny (m.in. uwzględnienie ekspertów międzynarodowych)
Ocena wszytkich a ocena liderów	Wymóg raportowania ograniczonej liczby publikacji od każdego pracownika	Stworzenie silnych impulsów do kierowania się doskonałością naukową w badaniach oraz stworzenie instrumentów motywowania (zarówno pozytywnych, jak i negatywnych)	Analiza aktów prawnych w zakresie liczby wymaganych publikacji od każdego pracownika oraz liczby możliwych do zgłoszenia publikacji przez uczelnie w danej dyscyplinie naukowej lub artystycznych

Wyzwanie	Wskaźnik	Oczekiwane efekty	Sposób ewaluacji wdrożonej zmiany
Ocena wpływu społecznego	Uwzględnienie Wpływu społecznego jako nowego kryterium	Dehermetyzacja nauki w Polsce i wyjście poza mury akademii. Jednocześnie dostarczenie argumentów na rzecz praktyczności i przydatności badań podstawowych, a w szczególności nauk humanistycznych i społecznych. Istotnym oczekiwanym efektem jest również legitymizacja badań naukowych w staraniach nad podniesieniem wysokości publicznych nakładów na badania naukowe w strukturze wydatków publicznych (w relacji do PKB)	W krótkim okresie możliwa jest jedynie ewaluacja tego, czy i w jaki sposób kryterium to zostało uwzględnione. Jak pokazuje przykład brytyjskiego REF-u, długofalowa ewaluacja jest możliwa dopiero po kilku latach po ogłoszeniu wyników oceny wpływu społecznego
Ocena publikacji naukowych	Zasada dziedziczenia prestiżu stosowana dla każdego kanału publikacji naukowych	Dowartościowanie monografii, rozdziałów oraz prac pod redakcją wydanych u najlepszych wydawców naukowych	Analiza sposobu przyznawania punktów za monografie, rozdziały i prace pod redakcją

7. LITERATURA

- Abramo, G. i D'Angelo, C.A. (2015). The relationship between the number of authors of a publication, its citations and the impact factor of the publishing journal: Evidence from Italy. *Journal of Informetrics*. 9(4): 746–761.
- Ancaiani, A., Anfossi, A.F., Barbara, A., Benedetto, S., Blasi, B., Carletti, V. ... i Sileoni, S. (2015). Evaluating scientific research in Italy: The 2004–10 research evaluation exercise. *Research Evaluation*. 24(3): 242–255.
- Antonowicz, D. i Brzeziński, J. M. (2013). Doświadczenia parametryzacji jednostek naukowych z obszaru nauk humanistycznych i społecznych 2013 – z myślą o parametryzacji 2017. *Nauka*. 4: 51–85.
- Butler, L. (2003). Explaining Australia's increased share of ISI publications – the effects of a funding formula based on publication counts. *Research Policy*. 32(1): 143–155.
- Ferrara, A. i Bonaccorsi, A. (2016). How robust is journal rating in Humanities and Social Sciences? Evidence from a large-scale, multi-method exercise. *Research Evaluation*. 25(3): 1–13.
- Good, B., Vermeulen, N., Tiefenthaler, B., & Arnold, E. (2015). Counting quality? The Czech performance-based research funding system. *Research Evaluation*. 24(2): 91–105.
- Haddow, G. i Genoni, P. (2010). Citation analysis and peer ranking of Australian social science journals. *Scientometrics*. 85(2): 471–487.
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*. 41(2): 251–261.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. i Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*. 520(7548): 429–431.
- Jablecka, J. i Lepori, B. (2009). Between historical heritage and policy learning: The reform of public research funding systems in Poland, 1989–2007. *Science and Public Policy*. 36(9): 697–708.
- Kulczycki, E. (2017a). Assessing publications through a bibliometric indicator: The case of comprehensive evaluation of scientific units in Poland. *Research Evaluation*. 26(1): 41–52.
- Kulczycki, E. (2017b). Punktoza jako strategia w grze parametrycznej w Polsce. *Nauka i Szkolnictwo Wyższe*. 1(49): 63–78.
- Kulczycki, E. (2019). *Raport VIII: Ocena publikacji jako instrument polityki naukowej*. Centrum Studiów nad Polityką Publiczną UAM: Poznań.
- Kulczycki, E., Drabek, A. i Rozkosz, E.A. (2015). Publikacje a zgłoszenia ewaluacyjne, czyli zniekształcony obraz nauki w Polsce. *Nauka*. 3: 35–58.
- Kulczycki, E., Rozkosz, E.A. i Drabek, A. (2015). Publikacje polskich badaczy w czasopiśmie z list ERIH w kontekście ewaluacji jednostek naukowych. *Kultura i Edukacja*. 1(107): 149–172.

- Kulczycki, E., Korzeń, M. i Korytkowski, P. (2017). Toward an excellence-based research funding system: Evidence from Poland. *Journal of Informetrics*. 11(1): 282–298.
- Langfeldt, L., Bloch, C.W. i Sivertsen, G. (2015). Options and limitations in measuring the impact of research grants – Evidence from Denmark and Norway. *Research Evaluation*. 24(3): 256–270.
- Rijcke, S. de, Wouters, P. F., Rushforth, A. D., Franssen, T. P. i Hammarfelt, B. (2016). Evaluation practices and effects of indicator use: A literature review. *Research Evaluation*. 25(2): 161–169.
- Saarela, M., Kärkkäinen, T., Lahtonen, T. i Rossi, T. (2016). Expert-based versus citation-based ranking of scholarly and scientific publication channels. *Journal of Informetrics*. 10(3): 693–718.
- Schneider, J.W., Aagaard, K. i Bloch, C.W. (2014). What happens when funding is linked to (differentiated) publication counts? New insights from an evaluation of the Norwegian Publication Indicator. W: E. Noyons (red.). *Proceedings of the science and technology indicators conference 2014 Leiden 'Context Counts: Pathways to Master Big and Little Data'* (543–550). Leiden: Universiteit Leiden.
- Schneider, J.W., Aagaard, K. i Bloch, C.W. (2016). What happens when national research funding is linked to differentiated publication counts?: A comparison of the Australian and Norwegian publication-based funding models. *Research Evaluation*. 25(3): 244–256.
- Sivertsen, G. (2017). Finansowanie oparte na publikacjach – Model norweski. *Nauka i Szkolnictwo Wyższe*. 49(1): 47–59.
- Sugimoto, C.R., Work, S., Larivière, V. i Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*. 68(9): 2037–2062.
- van den Besselaar, P., Heyman, U. i Sandström, U. (2017). Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics*. 11(3): 905–918.
- Vanclay, J.K. (2011). An evaluation of the Australian Research Council's journal ranking. *Journal of Informetrics*. 5(2): 265–274.
- Wróblewska, M.N. (2017). Ewaluacja „wpływu społecznego” nauki: przykład REF 2014 a kontekst polski. *Nauka i Szkolnictwo Wyższe*. 49(1): 79–104.

O AUTORZE



Emanuel Kulczycki, dr hab. prof. UAM – zajmuje się ewaluacją oraz socjologią nauki. Jest profesorem nadzwyczajnym w Uniwersytecie im. Adama Mickiewicza w Poznaniu, gdzie kieruje grupą badawczą Scholarly Communication Research Group. Doktorat (2011) i habilitację (2016) zrobił z filozofii. Obecnie jest przewodniczącym European Network for Research Evaluation in the Social Sciences and the Humanities. Ostatnio publikował m.in. w “Nature”, “Research Evaluation”, “Journal of Informetrics”, “Scientometrics” oraz “Aslib Journal of Information Management”. Więcej na: emanuelkulczycki.com.

CENTRUM STUDIÓW NAD POLITYKĄ PUBLICZNĄ UAM

Centrum Studiów nad Polityką Publiczną UAM działa od 2002 r., a jego założycielem i dyrektorem jest prof. Marek Kwiek, kierownik Katedry UNESCO Badań Instytucjonalnych i Polityki Szkolnictwa Wyższego. Centrum skupia młodych naukowców zajmujących się międzynarodowymi badaniami naukowymi szkolnictwa wyższego (*higher education research*) i systematycznie promuje je w Polsce jako istotną subdyscyplinę nauk społecznych. Centrum wydaje półrocznik *Nauka i Szkolnictwo Wyższe* i organizuje Ogólnopolskie Konferencje Badaczy Szkolnictwa Wyższego. W pracach Centrum regularny udział bierze 10 osób, m.in. kilkoro finansowanych zewnętrznie postdoków (filozofowie, socjologowie, ekonomiści, prawnicy i statystycy). Sekretarzem naukowym Centrum jest dr Krystian Szadkowski.

Centrum prowadzi badania naukowe w ramach rozbudowanej współpracy międzynarodowej, do której systematycznie włącza młodą kadre. Najważniejsze międzynarodowe projekty badawcze Centrum w ostatniej dekadzie dotyczyły relacji państwo – uniwersytet, relacji uniwersytetów z gospodarką i społeczeństwem, transformacji kadry akademickiej w ramach zmieniającego się finansowania uczelni, zarządzania szkolnictwem wyższym, przedsiębiorczości akademickiej, globalizacji i równości szans edukacyjnych oraz dostępu do szkolnictwa wyższego. Centrum prowadzi od wielu lat wspólne projekty badawcze z najlepszymi ośrodkami badań szkolnictwa wyższego w Europie (CHEPS w Holandii, INCHER w Kassel, Institute of Education w Oslo i UCL Institute of Education w Londynie) i w USA (CIHE w Bostonie i SUNY w Albany). Najważniejsze duże międzynarodowe projekty dotyczące badań szkolnictwa wyższego realizowane w Centrum to EDUWEL: *Education and Welfare* (2009-2013), WORKABLE: *Making Capabilities Work* (2009-2012), EUROAC: *The Academic Profession in Europe* (2009-2012), EUEREK: *European Universities for Entrepreneurship* (2004-2007) oraz GOODUEP: *Good Practices in University-Enterprises Partnerships* (2007-2009).

W latach 2012-2018 w Centrum realizowany był pięcioletni projekt MAESTRO: *Program Międzynarodowych Badań Porównawczych Szkolnictwa Wyższego*, finansowany przez NCN, a w latach 2014-2018 – projekt NCN HARMONIA (*Europejskie uniwersytety flagowe*). W latach 2015-2017 Centrum realizowało również

projekt w programie MISTRZ finansowany przez Fundację na rzecz Nauki Polskiej (FNP), a od 2016 r. realizuje projekt w programie DIALOG *Doskonałość naukowa: konkurencyjność, mierzalność, umiędzynarodowienie (EXCELLENCE)*. W 2016-2017 10-osobowy zespół prowadził w Centrum projekt przygotowujący założenia do nowej ustawy, zwanej Ustawą 2.0, promujący badawczą wizję funkcjonowania najlepszych polskich uczelni i silne publiczne finansowanie badań naukowych.

Centrum jest najważniejszym ośrodkiem naukowym zajmującym się badaniami szkolnictwa wyższego w Polsce i w Europie Środkowej.

www.cpp.amu.edu.pl

SERIA RAPORTÓW CENTRUM STUDIÓW NAD POLITYKĄ PUBLICZNĄ UAM

- 1. Marek Kwiek**, *Umiędzynarodowienie badań naukowych i widzialność polskiej nauki w świecie*
- 2. Marek Kwiek**, *Modele kariery naukowej i atrakcyjność profesji akademickiej*
- 3. Marek Kwiek**, *Indywidualna produktywność naukowa i konsekwencje rosnącej stratyfikacji społecznej w nauce*
- 4. Joanna Wolszczak-Derlacz**, *Efektywność i produktywność szkół wyższych*
- 5. Dominik Antonowicz**, *Stopniowe różnicowanie systemu szkolnictwa wyższego i jego konsekwencje*
- 6. Dominik Antonowicz**, *Zarządzanie szkolnictwem wyższym i jego przejrzysty ustrój a autonomia instytucji akademickich*
- 7. Emanuel Kulczycki**, *Procedury ewaluacji jednostek podstawowych i instytucji*
- 8. Emanuel Kulczycki**, *Procedury ewaluacji czasopism, współczynniki wpływu i listy czasopism*
- 9. Krystian Szadkowski**, *Problematyka wylaniania wiodących jednostek naukowych, centrów doskonałości i instytucji flagowych*
- 10. Krystian Szadkowski**, *Globalne rankingi uniwersytetów a długoterminowa strategia wzmacniania pozycji polskich uczelni*