Architektura pomiaru wpływu społecznego: od rygoru przyczynowości do dyscypliny wskaźnikowej w ujęciu Matta Stevensona-Dodda



AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje metodologiczne wyzwania pomiaru wpływu społecznego, koncentrując się na rozróżnieniu między prostą chronologią zdarzeń a rzeczywistą przyczynowością. Autor przywołuje koncepcje Matta Stevensona-Dodda, przeciwstawiając atrybucję (bezpośrednie przypisanie zmiany) kontrybucji (wykazanie roli organizacji w złożonym systemie). Kluczowym elementem jest wprowadzenie pojęcia kontrfaktyczności – próby oszacowania, co stałoby się bez interwencji. Dzięki temu możliwe jest oddzielenie zmiany brutto od wpływu netto, co pozwala uniknąć błędów interpretacyjnych wynikających z selekcji uczestników czy zmiennych kontekstowych. Tekst podkreśla, że rzetelna ewaluacja wymaga wyjścia poza schemat „przed i po” na rzecz bardziej zaawansowanych narzędzi statystycznych i analitycznych.

The article analyzes the methodological challenges of measuring social impact, focusing on the distinction between simple chronology and actual causality. The author references Matt Stevenson-Dodd's concepts, contrasting attribution (direct assignment of change) with contribution (demonstrating the organization's role within a complex system). A key element is the introduction of counterfactuals—an attempt to estimate what would have happened without intervention. This makes it possible to separate gross change from net impact, avoiding interpretative errors resulting from participant selection or contextual variables. The text emphasizes that reliable evaluation requires moving beyond 'before and after' schemes toward more advanced statistical and analytical tools.

Artykuł stanowi krytyczną analizę metodologii pomiaru wpływu społecznego, stawiając tezę, że rzetelna ewaluacja wymaga przejścia od naiwnej chronologii zdarzeń do rygorystycznej analizy przyczynowej. Autor argumentuje, że samo stwierdzenie pozytywnej zmiany (zmiana brutto) jest poznawczo niewystarczające; prawdziwy wpływ netto można ustalić jedynie poprzez odseparowanie efektów interwencji od tzw. kontrfaktyczności – czyli tego, co i tak wydarzyłoby się bez udziału organizacji. W tekście przeciwstawiono koncepcję bezpośredniego przypisania sukcesu (attribution) bardziej pokornej analizie wkładu (contribution), która uznaje wielopodmiotową naturę zmian społecznych. Kluczowym elementem wywodu jest ostrzeżenie przed „paraliżem analitycznym" i prawem Goodharta: nadmiar powierzchownych wskaźników KPI często prowadzi do optymalizacji liczb zamiast realnej poprawy dobrostanu beneficjentów. Rozwiązaniem jest budowa hierarchicznego systemu pomiarowego, który łączy epistemiczną skromność z selektywnym doborem strategicznych mierników, pozwalając odróżnić rzeczywistą wartość dodaną organizacji od naturalnych procesów społecznych i przypadkowych korelacji.

SŁOWA KLUCZOWE

pomiar wpływu społecznego | kontrfaktyczność | attribution vs contribution | Theory of Change | deadweight | displacement | drop-off | wpływ netto | selection bias | regresja do średniej | analiza kontrybucji | prawo Goodharta | materialność danych | evidence-informed judgement | SROI

Pomiar wpływu zaczyna się tam, gdzie kończy się chronologia

Przyczynowość, kontrfaktyczność i granice przypisywania wpływu: od „stało się" do „stało się dzięki nam"

Najtrudniejszym momentem w analizie wpływu społecznego nie jest stwierdzenie, że po interwencji zaszła zmiana, lecz rozstrzygnięcie, jaką część tej zmiany wolno racjonalnie powiązać z działaniem organizacji. Właśnie tutaj język zarządzania zderza się z jednym z najstarszych problemów

epistemologii naukowej: różnicą pomiędzy następstwem, współwystępowaniem a przyczynowością. Beneficjent otrzymał wsparcie, a następnie znalazł pracę; uczeń uczestniczył w programie i poprawił swoje wyniki; dzielnica została objęta interwencją społeczną, po której spadła przestępczość.

Żadne z tych zdań samo w sobie nie ustanawia relacji kauzalnej. Pomiędzy interwencją a rezultatem mogły bowiem oddziaływać zmienne kontekstowe: procesy spontaniczne, działania innych podmiotów, selekcja uczestników, zmiany w polityce publicznej, koniunktura gospodarcza, efekt dojrzewania czy regresja do średniej. Pomiar wpływu zaczyna się więc tam, gdzie kończy się prosta chronologia. Autor książki uchwycił tę trudność poprzez rozróżnienie między attribution a contribution. Pierwsze pojęcie odnosi się do przypisywania obserwowanej zmiany bezpośrednio działaniu organizacji; drugie – do wykazania, że organizacja była jednym z istotnych elementów szerszego układu przyczynowego. Autor słusznie wskazuje, że w złożonych systemach społecznych domaganie się wyłącznego autorstwa rezultatu może prowadzić do iluzji kontroli, podczas gdy ujęcie kontrybucyjne lepiej odzwierciedla wielopodmiotową naturę realnej zmiany.

Wymaga to jednak doprecyzowania: przeciwstawienie attribution i contribution nie powinno być traktowane jako wybór między „naukową, lecz nierealistyczną przyczynowością" a „praktycznym wkładem". W rzeczywistości oba pojęcia reprezentują różne poziomy odpowiedzi na pytanie o przyczynowość i mogą być metodologicznie komplementarne. W najbardziej rygorystycznym sensie wpływ można zdefiniować poprzez kontrfaktyczność.

Chcąc wiedzieć, jaki skutek wywołała interwencja, należałoby porównać sytuację tej samej jednostki w tym samym czasie w dwóch stanach świata: jednym, w którym interwencja nastąpiła, i drugim, w którym jej zabrakło. Problem polega na tym, że oba stany nie mogą zostać zaobserwowane jednocześnie. Osoba albo uczestniczyła w programie, albo nie; przedsiębiorstwo albo otrzymało dotację, albo jej nie otrzymało; gmina albo wdrożyła reformę, albo funkcjonowała bez niej.

Wpływ netto wymaga odseparowania zmiany brutto od kontrfaktyczności

Kontrfaktyczny wynik jest zatem poznawczo nieobserwowalny i musi zostać oszacowany za pomocą odpowiednio zaprojektowanego porównania. To właśnie ten problem leży u podstaw eksperymentalnej i quasi-eksperymentalnej ewaluacji wpływu. Współczesny **Magenta Book** definiuje kontrfaktyczność jako próbę odtworzenia tego, co wydarzyłoby się w przypadku braku interwencji, wykorzystując grupę kontrolną, porównawczą lub inne konstrukcje statystyczne;

dopiero różnica między obserwowanym wynikiem grupy objętej programem a wiarygodnym kontrfaktykiem pozwala z większym przekonaniem mówić o efekcie przyczynowym. Podobnie Bank Światowy lokuje problem kontrfaktyczności w centrum **impact evaluation**, budując wokół niego cały aparat randomizacji, **difference-in-differences**, **regression discontinuity**, **instrumental variables** i **matching**.

Można to wyrazić formalnie językiem potencjalnych wyników. Dla jednostki i istnieją hipotetycznie dwa rezultaty: wynik przy interwencji oraz wynik w jej przypadku braku. Indywidualny efekt przyczynowy stanowiłby różnicę między tymi dwiema wartościami, których jednak nie można zaobserwować jednocześnie. Metodologia **impact evaluation** nie „widzi" więc przyczynowości bezpośrednio; konstruuje raczej warunki, w których wynik osób podobnych, lecz nieobjętych interwencją, może służyć jako wiarygodne przybliżenie stanu nieobserwowanego. Ta prosta notacja odsłania, dlaczego sama poprawa wyniku po programie jest poznawczo niewystarczająca. Jeśli uczestnik był bezrobotny przed interwencją i pracuje po niej, obserwujemy zmianę w czasie, lecz nadal nie znamy alternatywnej historii jego życia bez programu.

Kluczowe jest tutaj rozróżnienie między **zmianą brutto** a **wpływem netto**. Zmiana brutto opisuje to, co się wydarzyło. Wpływ netto próbuje ustalić, jaka część tego zdarzenia przekracza to, co i tak by nastąpiło. Rozróżnienie to ma ogromną wagę zarówno naukową, jak i etyczną. Organizacja może rzetelnie raportować wysoką liczbę pozytywnych rezultatów, a jednocześnie wywoływać niewielki efekt netto, jeśli beneficjenci osiągnęliby podobne wyniki bez jej udziału. W takim przypadku nie mamy do czynienia z fałszywymi danymi, lecz z błędną interpretacją danych prawdziwych.

Problematyka kontrfaktyczności pozwala również zrozumieć, dlaczego pomiar „przed i po" bywa zdradliwy. Osoby zgłaszające się do programu często nie stanowią losowej próby populacji; mogą być bardziej zmotywowane, lepiej poinformowane lub przeciwnie – znajdować się w szczególnie trudnym momencie życia. Jeśli program wybiera osoby z najwyższym prawdopodobieństwem sukcesu, późniejsza skuteczność może odzwierciedlać **selection bias**, a nie jakość interwencji. Jeżeli natomiast rekrutuje uczestników w momencie skrajnego pogorszenia ich sytuacji, część poprawy może wynikać z **regresji do średniej**: ekstremalne stany mają statystyczną tendencję do powrotu ku wartościom typowym nawet bez działania programu.

W naukach społecznych mechanizm ten jest szczególnie istotny przy interwencjach uruchamianych w odpowiedzi na kryzys. Program wsparcia psychologicznego często rozpoczyna się, gdy stan uczestnika jest wyjątkowo zły; program naprawczy w przedsiębiorstwie wdraża się po gwałtownym spadku wyników, a interwencję policyjną nasila po lokalnym wzroście przestępczości. Jeśli sytuacja następnie częściowo się poprawia, organizacja może spontanicznie przypisać tę zmianę interwencji,

choć efekt ten mógłby wystąpić wskutek naturalnej fluktuacji. Dlatego **before-after** jest opisem temporalnym, nie automatycznie projektem przyczynowym.

Dochodzi do tego **history effect**, czyli wpływ wydarzeń zewnętrznych zachodzących równolegle do programu. Kurs aktywizacyjny może zbiegać się w czasie z silnym ożywieniem gospodarczym, program profilaktyki zdrowotnej z ogólnokrajową kampanią medialną, a lokalna interwencja edukacyjna z rokiem reformy podstawy programowej.

Dopasowanie siły twierdzeń do rygoru metodologicznego

Im dłuższy horyzont pomiaru, tym więcej konkurencyjnych procesów może wpływać na wynik. Paradoks pomiaru wpływu polega zatem na tym, że trwały skutek jest poznawczo najbardziej interesujący, a zarazem najtrudniejszy do jednoznacznego przypisania, gdyż z czasem rośnie liczba potencjalnych przyczyn. Najskuteczniejszą ochroną przed problemami selekcji pozostaje randomizacja. Losowy przydział jednostek do grupy interwencyjnej i kontrolnej sprawia, że w wystarczająco dużej próbie różnice systematyczne między grupami stają się mało prawdopodobne; dzięki temu późniejszą rozbieżność wyników można z większym zaufaniem powiązać z samą interwencją. Aktualne brytyjskie wytyczne klasyfikują randomized controlled trials jako projekty o wysokiej zdolności do ustalania attribution, zastrzegając jednak, że każda metoda ma swoje warunki poprawności i zależy od jakości danych. Nie oznacza to jednak, że hierarchię dowodów należy sprowadzić do dychotomii: RCT albo niewiedza. Eksperyment precyzyjnie odpowiada na pytanie o średni efekt interwencji w określonej populacji i warunkach, lecz nie zawsze wyjaśnia mechanizm, kontekst, heterogeniczność efektów czy ich przenoszalność. Nawet silnie zidentyfikowany efekt lokalny może nie zostać odtworzony po zmianie skali, instytucji, populacji lub czasu.

Internal validity i external validity pozostają odrębnymi problemami. Można mieć pewność, że coś zadziałało „tam i wtedy", a wciąż nie wiedzieć, czy zadziała „tu i teraz". Co więcej, części interwencji nie da się realistycznie randomizować. Reforma prawa krajowego nie dopuszcza istnienia „drugiej Polski" jako grupy kontrolnej, a zmiana modelu działania całej instytucji często nie posiada odpowiedniego odpowiednika. Kampania rzecznicza oddziałuje poprzez media, sieci, koalicje, proces legislacyjny i zmianę dyskursu, natomiast programy systemowe zazwyczaj współwystępują z innymi reformami.

W takich przypadkach problemem nie jest rzekoma rezygnacja nauki z przyczynowości, lecz konieczność zastosowania innych strategii inferencyjnych: interrupted time series, synthetic control, comparative case studies, process tracing, realist evaluation, contribution analysis lub kombinacji tych metod. Dojrzała epistemologia wpływu nie ustanawia kultu jednej techniki, lecz dopasowuje siłę twierdzenia do siły projektu badawczego. Jeśli organizacja przeprowadziła rzetelny eksperyment, może formułować silniejsze wnioski o średnim efekcie. Jeśli dysponuje jedynie badaniem przed-po, powinna mówić o obserwowanej zmianie, a nie automatycznie o skutku przyczynowym. Jeżeli natomiast posiada dobrze zrekonstruowaną teorię zmiany, triangulację dowodów, process tracing i analizę alternatywnych wyjaśnień, może budować wiarygodne twierdzenie o contribution, choć nadal nie powinna stylizować go na precyzyjny efekt eksperymentalny. W tym miejscu szczególnego znaczenia nabiera koncepcja Johna Mayne'a.

Analiza kontrybucji jako rygorystyczna alternatywa dla uproszczonej atrybucji

Contribution analysis została rozwinięta właśnie po to, aby oceniać udział programu w wynikach tam, gdzie eksperymentalne odizolowanie pojedynczej przyczyny jest trudne lub nieadekwatne. Nie oznacza to jednak rezygnacji z causal reasoning. Przeciwnie – analiza kontrybucji wymaga zbudowania wiarygodnej Theory of Change, rozpoznania założeń i konkurencyjnych czynników, zgromadzenia dowodów na poszczególne ogniwa, sformułowania **contribution story**, poddania jej krytyce, pozyskania dodatkowych danych i ponownego zweryfikowania narracji. Kluczowe jest tutaj pojęcie **causal package**. W złożonym świecie rezultat nie musi mieć jednej przyczyny; może być wynikiem konfiguracji kilku warunków, z których żaden samodzielnie nie byłby wystarczający.

Młody człowiek może znaleźć stabilną pracę dlatego, że jednocześnie ukończył szkolenie, otrzymał mentoring, poprawiła się jego sytuacja psychiczna, rodzina przejęła część obowiązków opiekuńczych, pracodawca otworzył rekrutację, a lokalny transport umożliwił dojazd. Pytanie o to, która z tych przyczyn „wywołała” sukces, jest źle postawione, jeśli rezultat wymagał ich wspólnej konfiguracji. Organizacja może być zatem przyczyną kontrybucyjną: jej wkład był częścią pakietu warunków, bez których prawdopodobieństwo rezultatu byłoby mniejsze. Mayne podkreśla właśnie możliwość traktowania interwencji jako jednego z niezbędnych lub znaczących elementów szerszego układu przyczynowego. Takie ujęcie pozwala wyjść poza potoczne przeciwstawienie: „albo my spowodowaliśmy rezultat, albo nie mamy na niego wpływu”. W rzeczywistości przyczynowość społeczna ma charakter warstwowy.

Interwencja może być warunkiem inicjującym, wzmacniającym, umożliwiającym albo przyspieszającym. Może zwiększać prawdopodobieństwo rezultatu bez jego determinowania; działać tylko w określonym kontekście lub dla konkretnej podgrupy; być konieczna, lecz niewystarczająca – bądź wystarczająca w jednym środowisku, a zbędna w drugim. Właśnie dlatego kategoria contribution jest poznawczo subtelniejsza niż marketingowe hasło „nasz program zmienił życie". Jednocześnie contribution analysis nie może stać się miękkim alibi dla każdego programu. BetterEvaluation wyraźnie zaznacza, że wiarygodne twierdzenie o wkładzie wymaga racjonalnej teorii zmiany, potwierdzenia realizacji działań, obserwacji oczekiwanego łańcucha rezultatów oraz analizy innych istotnych czynników. Nie jest to zatem opowieść typu „jedna z wielu przyczyn, trudno powiedzieć", lecz **ustrukturyzowana próba zwiększania pewności przyczynowej bez udawania, że świat społeczny można zawsze laboratoryjnie odizolować**. W praktyce organizacyjnej pojęcie attribution bywa niefortunnie traktowane jak własność: „jaki procent wyniku należy do nas?". Taki sposób myślenia może być użyteczny rachunkowo, ale filozoficznie jest niebezpieczny.

Rezultat społeczny nie jest tortem, którego naturalne kawałki należą do poszczególnych przyczyn. Często czynniki oddziałują interaktywnie. Bez A wynik nie nastąpiłby, ale bez B również nie; przypisanie 50% zasługi czynnikowi A i 50% czynnikowi B może nie mieć żadnego głębszego sensu przyczynowego. Przykładowo: jeśli terapia działa tylko wtedy, gdy pacjent ma stabilne warunki mieszkaniowe, a stabilność mieszkaniowa przynosi skutek zdrowotny tylko dzięki terapii, oba elementy są komplementarne. Próba arytmetycznego podzielenia „zasługi" ignoruje tę interakcję. Podobnie w projektach edukacyjnych skuteczność nauczyciela zależy od ucznia, środowiska rodzinnego, programu, warunków szkolnych i wcześniejszych kompetencji.

Nie oznacza to, że kwantyfikacja attribution jest zawsze bezsensowna, lecz że powinna być traktowana jako **model analityczny oparty na założeniach**, a nie obiektywne odkrywanie procentowej własności rezultatu. To rozróżnienie jest istotne również w relacjach między fundatorami a organizacjami. Grantodawca może chcieć wiedzieć, „ile efektu kupił" za przekazane środki, a organizacja może być skłonna przypisać mu proporcjonalną część sukcesu. Tymczasem kapitał mógł być jedynie jednym z komplementarnych warunków obok wiedzy personelu, zaufania społecznego, infrastruktury publicznej i pracy wolontariuszy. Język attribution może więc nieświadomie reprodukować ekonomię własności w obszarze, gdzie społeczna zmiana ma charakter kooperacyjny.

Deadweight i displacement jako korektory wpływu netto

Pierwszym z klasycznych "deflatorów wpływu" jest **deadweight**. Oznacza on tę część obserwowanego rezultatu, która prawdopodobnie nastąpiłaby nawet bez podjętej interwencji. W systemie SROI deadweight pełni funkcję pomniejszania roszczenia organizacji o to, co nie wymagało jej działania. Koncepcja ta jest prosta, lecz niesie ze sobą daleko idące konsekwencje.

Jeśli 80% uczestników programu znalazło pracę, nie można od razu raportować 80% "wpływu". Jeżeli w porównywalnej grupie osób, które nie brały udziału w programie, zatrudnienie znalazłoby 60%, realny efekt dodatkowy jest znacznie mniejszy. Jeśli ponadto uczestnicy byli wybierani spośród osób o wysokich kwalifikacjach i dużej motywacji, deadweight może być jeszcze wyższy. Program może zatem wyglądać imponująco w liczbach absolutnych, a mimo to generować niewielką wartość marginalną. Zarządczo prowadzi to do pojęcia **additionality**. Organizacja powinna pytać nie tylko o to, czy tworzy dobro, lecz czy tworzy dobro dodatkowe względem scenariusza bazowego.

Taki sposób myślenia jest szczególnie istotny w politykach dotacyjnych. Dotacja dla przedsiębiorstwa może sfinansować inwestycję, którą firma i tak zamierzała zrealizować; subsydium zatrudnieniowe może wspierać osobę, która zostałaby zatrudniona również bez zewnętrznej pomocy, a grant kulturalny może finansować wydarzenie posiadające wystarczający popyt komercyjny. W każdym z tych przypadków output zostaje osiągnięty, ale additionality może być niewielka.

Deadweight jest więc czymś więcej niż korektą statystyczną – stanowi próbę odpowiedzi na pytanie o **koszt alternatywny interwencji**. Jeżeli kapitał publiczny lub społeczny finansuje rezultat, który wydarzyłby się samoistnie, środki te nie zostały użyte tam, gdzie mogłyby wygenerować większą dodatkową zmianę. Pomiar deadweight wiąże się zatem bezpośrednio z efektywnością alokacyjną.

Drugim korektorem jest **displacement**, czyli przesunięcie rezultatu. Interwencja może poprawić sytuację jednej grupy, nie zwiększając całkowitego dobrobytu systemu, lecz przenosząc korzyść kosztem innej grupy. SROI traktuje displacement jako element rezultatu, który należy odjąć przy ustalaniu wartości netto. Najbardziej oczywisty przykład występuje na rynku pracy: program może zwiększyć zatrudnienie uczestników, ale jeśli liczba dostępnych stanowisk jest stała, sukces ten oznacza jedynie, że beneficjenci zostali zatrudnieni zamiast osób spoza programu. Z perspektywy uczestnika interwencja jest skuteczna; z perspektywy społecznej całkowite zatrudnienie może pozostać bez zmian.

Podobny problem pojawia się w rewitalizacji miejskiej, gdy poprawa atrakcyjności dzielnicy wypiera uboższych mieszkańców, w subsydiowaniu biznesu, gdy wsparcie jednej firmy osłabia jej niewspieranych konkurentów, czy przy intensyfikacji działań policji w jednym obszarze, co skutkuje geograficznym przemieszczeniem się przestępczości. Displacement pokazuje, dlaczego analiza wpływu musi precyzyjnie określić **jednostkę i poziom analizy**. Program może być skuteczny dla jednostki, neutralny dla miasta i nieefektywny dla całego kraju – wpływ mikro nie sumuje się automatycznie do wpływu makro. Zjawisko to jest szczególnie istotne w Polsce w kontekście programów regionalnych i funduszy wspierających konkurencyjność: wzrost zatrudnienia w jednej lokalizacji może w rzeczywistości oznaczać jedynie przesunięcie aktywności z innego powiatu, przedsiębiorstwa lub sektora.

Korekta efektów brutto i pułapka monetyzacji wpływu

Trzecim istotnym korektorem jest **drop-off**, czyli stopniowe wygasanie efektu w czasie. Rezultat osiągnięty bezpośrednio po zakończeniu programu nie musi utrzymywać się przez rok, trzy lata czy dekadę. Social Value International definiuje drop-off jako pogarszanie się rezultatu wraz z upływem czasu, co wymusza ograniczenie wartości przypisywanej przyszłym okresom. Pojęcie to wprowadza do teorii wpływu kluczowe rozróżnienie pomiędzy **efektem punktowym, trwałością i samopodtrzymywaniem**. Kurs może chwilowo zwiększyć wiedzę, ale bez praktycznego zastosowania kompetencja zanika.

Interwencja psychologiczna może przynieść poprawę, która jednak wymaga okresowego wsparcia. Zatrudnienie może nastąpić, by zakończyć się już po miesiącu. Zmiana zachowania często utrzymuje się tylko tak długo, jak istnieje bodziec zewnętrzny. W każdym z tych przypadków „osiągnięcie outcome” w chwili t_1 nie jest równoważne trwałemu wpływowi. Ma to istotne następstwa dla projektowania mierników: organizacja badająca rezultat wyłącznie przy zamknięciu programu ma naturalną tendencję do przeceniania jego trwałości. Im szybciej następuje pomiar, tym łatwiej wykazać pozytywną zmianę, lecz tym mniej wiemy o jej stabilności. Powstaje zatem konflikt między wygodą ewaluacyjną a istotą pojęcia impact. Jeżeli wpływ ma oznaczać zmianę trwałą, obserwacja długookresowa nie jest luksusem metodologicznym, lecz częścią definicji przedmiotu.

Teoria wpływu byłaby jednak wciąż zbyt wąska, gdyby uwzględniała jedynie pożądane rezultaty. Interwencje w systemach społecznych generują również **skutki niezamierzone**, zarówno dodatnie, jak i ujemne. Program może zrealizować główny cel, a jednocześnie zwiększyć nierówności, stygmatyzację, koszty administracyjne lub zależność od instytucji. Z drugiej strony

może nie osiągnąć celu podstawowego, lecz wywołać wartościowe efekty uboczne, których nie przewidziano. OECD wprost definiuje impact jako obejmujący również skutki pośrednie, niezamierzone i wyższego rzędu, a współczesne standardy impact evaluation podkreślają konieczność ich badania.

To zmienia status Theory of Change. Nie wystarczy pytać, czy zrealizował się przewidywany łańcuch zdarzeń; trzeba pozostawić miejsce dla **emergent outcomes**, czyli rezultatów pojawiających się poza pierwotnym modelem. W przeciwnym razie system ewaluacyjny może być strukturalnie ślepy na skutki, których nie umieszczono w arkuszu KPI. Paradoks polega na tym, że im silniej organizacja skupia uwagę na kilku wybranych wskaźnikach, tym większe ryzyko przeoczenia zmian nieprzewidzianych. Prostota musi więc zostać uzupełniona systemem wykrywania anomalii.

W tym kontekście kluczowe są zasady sformułowane przez Stevenson-Dodda. Autor książki przywołuje trzy z nich: zakaz wyolbrzymiania własnych osiągnięć, podawanie liczb bezwzględnych obok procentów oraz wymóg audytowalności rezultatów. Są to rygorystyczne zasady rzetelności sprawozdawczej, jednak **nie należy utożsamiać audytowalności outcome z dowodem attribution**. Pasek wynagrodzenia może bezspornie potwierdzić, że beneficjent podjął pracę, lecz nie dowodzi, że to program spowodował zatrudnienie. Zaświadczenie szkolne potwierdza ukończenie kursu, ale nie jest dowodem na to, że kurs wywołał długotrwałą zmianę. Dokumentacja źródłowa rozwiązuje problem **measurement validity** na jednym poziomie – upewnia nas, że zdarzenie zaszło – ale nie rozstrzyga automatycznie kwestii **causal validity**. To rozróżnienie powinno zajmować centralne miejsce w dojrzałym systemie impact management.

Można zatem zaproponować trzy kolejne warstwy dowodowe. Pierwsza odpowiada na pytanie: **czy rezultat zaszedł?** Druga: **czy rezultat jest związany z interwencją zgodnie z wiarygodnym mechanizmem i dowodami?** Trzecia: **jaka część lub prawdopodobieństwo zmiany może zostać przypisana interwencji w porównaniu z wiarygodnym kontrfaktykiem?** Audytowalność w stylu Street League znakomicie wzmacnia pierwszą warstwę. Contribution analysis może znacząco wspierać drugą, natomiast projekty eksperymentalne i quasi-eksperymentalne są najsilniejsze w trzeciej. Dopiero połączenie tych poziomów pozwala uniknąć dwóch symetrycznych błędów: **marketingowego overclaimingu** i **metodologicznego perfekcjonizmu**.

Pierwszy przypisuje organizacji wszystko, co wydarzyło się po interwencji. Drugi uznaje, że bez RCT nie wolno powiedzieć nic o wpływie. Dojrzała metodologia operuje skalą uzasadnionej pewności i nazywa dokładnie to, co pozwalają stwierdzić dane. W tym świetle interesujące stają się narzędzia Social Return on Investment. Autor książki jest wobec SROI sceptyczny: dostrzega zdolność tej metody do porządkowania nakładów i rezultatów, ale wskazuje na ryzyko arbitralnych **financial**

proxies, monetyzacji dóbr niemających oczywistej ceny oraz fiksację na końcowym wskaźniku zwrotu. Ta krytyka jest zasadna, lecz SROI posiada także ważny walor poznawczy: zmusza analityka do jawnego pomniejszania surowego rezultatu o deadweight, displacement, attribution i drop-off. Social Value International przedstawia SROI jako ramę do mierzenia, zarządzania i rozliczania wartości społecznej, a praktyka ta traktuje wspomniane korekty jako zabezpieczenie przed overclaimingiem. W pewnym sensie SROI próbuje **zamienić epistemiczną pokorę na procedurę rachunkową**. Zamiast ogólnego stwierdzenia „nie wszystko jest naszą zasługą", każe oszacować, jaka część zmiany wydarzyłaby się i tak, jaka została przesunięta, jaka wynikała z udziału innych aktorów oraz jak szybko efekt zanika. Jest to intelektualnie wartościowe, gdyż wymusza ujawnienie założeń. Problem zaczyna się jednak wtedy, gdy niepewne przesłanki zostają skondensowane do precyzyjnie wyglądającej liczby, np.: „każda złotówka wygenerowała 4,37 zł wartości społecznej".

Wpływ jako model decyzyjny i gramatyka dowodów

Im bardziej precyzyjny wydaje się wynik końcowy, tym łatwiej zapomnieć, że jego fundamentem mogą być szerokie przedziały niepewności i sporne proxy. Dlatego SROI należy interpretować nie jako odkrywanie "prawdziwej ceny" społecznego dobra, lecz jako model decyzyjny zależny od jawnych założeń. Dojrzałe zastosowanie tej metody wymaga analizy wrażliwości: sprawdzenia, jak wynik zmienia się po modyfikacji parametrów takich jak deadweight, attribution, drop-off, wartości proxy czy czas trwania efektu. Współczesna praktyka Social Value promuje właśnie takie podejście, testując wpływ zmian tych zmiennych na końcowy wskaźnik. Jeśli niewielka korekta założenia radykalnie zmienia SROI, wynik należy uznać za mało stabilny.

Jeszcze głębszy problem pojawia się, gdy ten sam program oddziałuje odmiennie na poziomie jednostki, instytucji i systemu. Można zwiększyć dobrostan konkretnych uczestników, nie dotykając strukturalnej przyczyny problemu. Można poprawić rezultaty jednej szkoły, przyciągając do niej najlepszych nauczycieli z sąsiednich placówek, co w efekcie pogorszy równowagę całego systemu. Podobnie można skutecznie aktywizować pojedyncze osoby bez realnego zwiększenia liczby miejsc pracy na rynku.

Z drugiej strony istnieją działania o charakterze systemowym – poprzez zmianę prawa lub normy – które przy niewielkiej liczbie bezpośrednich beneficjentów przynoszą szerokie skutki. Dlatego impact nie powinien być traktowany jako pojedynczy punkt na końcu liniowej drabiny. Trafniej jest mówić o poziomach oddziaływania: indywidualnym, relacyjnym, organizacyjnym, instytucjonalnym i systemowym. Mechanizmy przyczynowe na każdym z nich mogą być odmienne;

rezultat mikro może być warunkiem zmiany makro, ale nie musi. Zmiana makro może natomiast nastąpić bez mierzalnego efektu u wszystkich bezpośrednich uczestników – kampania rzecznicza może oddziaływać poprzez legislację i debatę publiczną, a nie przez klasyczną ścieżkę usługową.

Taka wielopoziomowość jest kluczowa dla organizacji obywatelskich, think tanków i podmiotów rzeczniczych. Liczenie uczestników spotkań czy liczby publikacji bywa niemal całkowicie oderwane od realnego wpływu. Z drugiej strony przypisanie sobie zmiany ustawowej tylko dlatego, że organizacja prowadziła kampanię, byłoby typowym overreachem. Konieczna jest analiza ścieżek contribution: sprawdzenie, czy organizacja dostarczyła argumentów wykorzystanych w procesie, czy jej przedstawiciele uczestniczyli w konsultacjach, czy konkretna propozycja została przejęta przez decydentów, jakie inne podmioty działały równolegle oraz kiedy dokładnie zmieniło się stanowisko instytucji.

W tym kontekście process tracing może okazać się bardziej adekwatny niż próba budowania sztucznej grupy kontrolnej. Perspektywę tę pogłębia epistemologia bayesowska: organizacja nie musi traktować pytania o wpływ w sposób binarny (udowodnione / nieudowodnione). Może zamiast tego przyjąć początkowe przekonanie o prawdopodobieństwie działania określonego mechanizmu, a następnie aktualizować je wraz z napływem danych. Teoria Zmiany tworzy tu prior – uzasadniony stan wiedzy wyjściowej. Wyniki programu, dane jakościowe i obserwowane mechanizmy stanowią kolejne porcje evidence, które zmieniają stopień zaufania do hipotezy. Process tracing może być formalizowany właśnie poprzez Bayesian updating, co podkreśla również aktualny Magenta Book.

W takim ujęciu kluczowa staje się diagnostyczność dowodu. Fakt, że uczestnik deklaruje zadowolenie, jest prawdopodobny zarówno wtedy, gdy program działa, jak i gdy nie działa (ludzie bywają uprzejmi wobec organizatorów). Taki dowód niewiele zmienia nasze przekonanie. Natomiast obserwacja rezultatu, który byłby mało prawdopodobny bez konkretnego mechanizmu, ma znacznie większą wartość inferencyjną. To pozwala zrozumieć, dlaczego więcej danych nie zawsze oznacza więcej wiedzy. Setki słabych wskaźników mogą wnosić mniej niż kilka dowodów o wysokiej mocy rozróżniającej między konkurencyjnymi hipotezami. Stevenson-Doddowska intuicja „mierzyć mniej, lecz ważniej” otrzymuje tu mocną podstawę metodologiczną: selekcja mierników nie powinna wynikać z łatwości zbierania danych, lecz z ich zdolności do aktualizowania przekonań o najważniejszych elementach Theory of Change.

Z tej analizy wyłania się potrzeba precyzyjniejszego języka organizacyjnego. Zamiast jednego, wszechobecnego czasownika „wpłynęliśmy”, warto rozróżniać: zaobserwowaliśmy, osiągnęliśmy, przyczyniliśmy się, zwiększyliśmy prawdopodobieństwo, współspowodowaliśmy, możemy z dużą pewnością przypisać, a w najsilniejszych przypadkach – oszacowaliśmy efekt przyczynowy. Każde z

tych sformułowań odpowiada innemu poziomowi dowodu. Tak rozumiana „gramatyka wpływu" ma znaczenie nie tylko metodologiczne, lecz również aksjologiczne.

Epistemiczna skromność jako fundament rzetelnego przypisywania sprawstwa

Overclaiming jest formą epistemicznej niesprawiedliwości: organizacja przywłaszcza sobie zasługę innych aktorów, samych beneficjentów oraz procesów społecznych, nad którymi nie sprawowała kontroli. Może to reprodukować paternalistyczną narrację, w której człowiek staje się biernym odbiorcą „zmiany dostarczonej" przez instytucję. Contribution language przypomina natomiast, że beneficjent sam jest podmiotem procesu; jego decyzje, wysiłek, relacje i sprawczość współtworzą rezultat.

Stevenson-Doddowskie „never over-claim" można zatem odczytać szerzej niż tylko jako zasadę marketingową. Jest to norma epistemicznej skromności i sprawiedliwego przypisywania sprawstwa. Organizacja typu „Difference" nie powinna dowodzić własnej wielkości, lecz precyzyjnie określać zakres, w którym jej obecność zmieniła prawdopodobny przebieg zdarzeń. Im bardziej złożony system, tym bardziej dojrzałość mierzenia przejawia się nie w głośności roszczenia, ale w dokładności granic roszczenia.

Wpływ społeczny nie jest zatem liczbą dodawaną na końcu raportu, lecz argumentem przyczynowym. Składa się on z Theory of Change, dowodów na realizację działań, obserwacji rezultatów, kontrfaktycznego rozumowania, analizy mechanizmów i konkurencyjnych wyjaśnień, a także uwzględnienia udziału innych aktorów, deadweight, displacement, trwałości oraz skutków niezamierzonych. Dopiero po przejściu przez tę architekturę organizacja może racjonalnie odpowiedzieć na pytanie nie tylko „czy coś się zmieniło?", lecz również „co zmieniło się dzięki naszej obecności, w jakim stopniu, dla kogo, na jak długo i z jaką pewnością możemy to powiedzieć?".

W tym kontekście kluczowy staje się spór o samą architekturę pomiaru: dlaczego organizacje mierzą zbyt wiele, jak powstaje paraliż wskaźnikowy, czym różni się użyteczny outcome od wygodnego KPI oraz gdzie przebiegają granice SROI, Social Value, wspólnych ram wskaźnikowych i opowieści o wpływie. Koncepcja Stevenson-Dodda zderza się tu z prawem Goodharta, teorią zarządzania przez cele, psychologią uwagi i kosztami transakcyjnymi pomiaru. Pojawia się problem wskaźnika, który zaczynając rządzić organizacją, może stopniowo przestać opisywać rzeczywistość.

Autor książki wskazuje, że problemem systemu pomiarowego może być nie niedostatek danych, lecz ich nadmiar. Cyfryzacja obniżyła krańcowy koszt gromadzenia informacji: systemy CRM umożliwiły rejestrowanie tysięcy pól, ankiety internetowe można powielać niemal bez ograniczeń, a dashboardy pozwalają prezentować dziesiątki wykresów na jednym ekranie. Technologiczna możliwość pomiaru łatwo przekształca się jednak w poznawczy imperatyw: skoro coś można policzyć, pojawia się pokusa, aby to liczyć.

Selektywność i hierarchia w odróżnianiu telemetrii od wpływu

Stevenson-Dodd przeciwstawia tej tendencji zasadę selektywności – system zarządczy powinien koncentrować się na kilku rezultatach, które rzeczywiście informują o tym, czy organizacja realizuje swój zasadniczy cel (purpose). Autor ujmuje tę dyscyplinę jako ograniczenie głównego systemu pomiarowego do około trzech-pięciu kluczowych wyników (outcomes) oraz przeniesienie ciężaru sporu z pytania „jakie jeszcze dane możemy zebrać?" na pytanie „które dane zmienią decyzję?". Nie należy jednak rozumieć tej zasady dosłownie, jakby liczba pięć posiadała szczególną wartość epistemologiczną. Przedział trzech-pięciu mierników jest heurystyką zarządczą, a nie prawem psychometrii czy uniwersalnym limitem złożoności. Jej sens polega na wymuszeniu hierarchii informacji. Organizacja może potrzebować setek zmiennych operacyjnych do zarządzania bezpieczeństwem, finansami, personelem i zgodnością prawną, a jednocześnie posługiwać się kilkoma wskaźnikami najwyższego poziomu jako sygnałami realizacji celu strategicznego.

Błąd pojawia się wówczas, gdy miesza się telemetrię organizacji z pomiarem jej wpływu. Liczba logowań do systemu, czas obsługi sprawy, liczba telefonów, frekwencja czy wykorzystanie budżetu mogą być niezwykle ważne operacyjnie, lecz nie sprawia to, że automatycznie stają się miernikami społecznej zmiany. System wskaźników powinien zatem posiadać strukturę hierarchiczną. Na najniższym poziomie znajdują się dane procesowe umożliwiające kontrolę codziennego działania. Wyżej można umieścić wskaźniki jakości i dostępności usług. Jeszcze wyżej – outcomes opisujące zmianę w sytuacji odbiorców. Na szczycie pojawiają się nieliczne miary lub zestawy dowodów odpowiadające na pytanie, czy organizacja zbliża rzeczywistość do stanu określonego w jej wizji.

Różne poziomy nie konkurują ze sobą; odpowiadają na odmienne pytania. Patologia pomiarowa zaczyna się wtedy, gdy system nie rozróżnia tych poziomów i przedstawia zarządowi kilkadziesiąt KPI o równym statusie, pozostawiając kierownictwu samodzielne odkrycie, które z nich naprawdę mają znaczenie. Każdy wskaźnik jest z definicji redukcją – nie jest zjawiskiem, lecz jego reprezentacją. Bezrobocie rejestrowane nie jest tym samym co brak pracy; frekwencja szkolna nie

jest edukacją; dochód nie jest dobrostanem; liczba konsultacji nie jest zdrowiem; deklarowane poczucie bezpieczeństwa nie jest całym doświadczeniem bezpieczeństwa.

Miara wybiera pewien wymiar rzeczywistości, operacjonalizuje go i pomija inne. Dzięki redukcji zyskujemy możliwość porównania, agregacji, monitorowania czasu i wykrywania odchyleń, lecz zawsze za cenę utraty części znaczenia. Dlatego najbardziej podstawowym pytaniem metodologicznym nie jest to, czy wskaźnik jest „obiektywny", lecz co dokładnie reprezentuje, jak wiernie oddaje interesujący konstrukt i w jakich sytuacjach przestaje być dobrym przybliżeniem. W psychometrii problem ten pojawia się jako trafność konstruktu; w ekonomii jako jakość proxy; w zarządzaniu jako problem KPI; w statystyce publicznej jako zależność między definicją operacyjną a zjawiskiem społecznym. Postulat prostoty Stevenson-Dodda osiąga naukową dojrzałość dopiero wtedy, gdy zostanie uzupełniony właśnie tym pytaniem o trafność. Łatwy do zebrania miernik nie staje się dobrym miernikiem dlatego, że jest tani, jednoznaczny i efektownie wygląda na dashboardzie. W tym miejscu szczególnie użyteczna jest tradycja kojarzona z prawem Goodharta.

Prawo Goodharta i pułapka optymalizacji wskaźników

Pierwotnie problem ten dotyczył polityki monetarnej: gdy władza zaczyna wykorzystywać obserwowaną regularność statystyczną jako instrument kontroli, zachowanie uczestników systemu ulega zmianie, a sama regularność może przestać być stabilna. Późniejsze uogólnienie tej myśli bywa formułowane znacznie prościej: **kiedy miernik staje się celem, może przestać być dobrym miernikiem**. Źródła historyczne wiążą intuicję Goodharta właśnie ze zmianami zachowań podmiotów pod wpływem systemu kontroli. Dla zarządzania wpływem oznacza to, że system pomiarowy nie jest neutralnym obserwatorem organizacji. W chwili, gdy od wyniku zależą fundusze, reputacja, przedłużenie kontraktu czy kariera menedżera, wskaźnik zaczyna **kształtować zachowanie, które miał wyłącznie opisywać**. Jeśli urząd oceniany jest według czasu rozpatrywania spraw, powstaje bodziec do skracania tego czasu, co niekoniecznie idzie w parze z poprawą jakości rozstrzygnięć. Gdy szpital premiowany jest za określony czas oczekiwania, organizacja może nauczyć się zarządzać definicją kolejki. Jeśli organizacja aktywizacyjna otrzymuje wynagrodzenie za liczbę osób zatrudnionych, może zacząć selekcjonować tych, którym najłatwiej pomóc. Z kolei gdy uczelnia jest oceniana przez pryzmat jednego rodzaju publikacji, system badawczy zaczyna produkować głównie ten konkretny typ treści. Empirycznym przykładem tych zjawisk są badania Gwyn Bevana i Christophera Hooda nad angielskim systemem ochrony zdrowia.

Autorzy ci wykazali, że silne zarządzanie przez cele może generować zarówno rzeczywistą poprawę, jak i zachowania typu **gaming**, ponieważ uczestnicy uczą się optymalizować formalne kryteria

oceny. Ich analiza jest szczególnie cenna, gdyż nie ogranicza się do banalnego twierdzenia, że „wskaźniki są złe", lecz ukazuje endogeniczność pomiaru: **ludzie reagują na sposób, w jaki są mierzeni**. W konsekwencji sprawny system **impact measurement** musi być projektowany nie tylko jako narzędzie epistemiczne, ale również jako **system bodźców**. Należy pytać, jakie zachowania staną się racjonalne po wprowadzeniu danego wskaźnika. To istotna zmiana perspektywy. Nie wystarczy zadać pytania, czy miara koreluje z pożądanym rezultatem; trzeba przewidzieć, co zrobią aktorzy, gdy dowiedzą się, że to właśnie ona decyduje o ich ocenie.

W programie reintegracji zawodowej można przykładowo wprowadzić miernik sześciomiesięcznego utrzymania zatrudnienia. Z punktu widzenia epistemicznego jest on znacznie lepszy od prostej liczby ukończonych szkoleń. Jednak jako bodziec może zachęcać do tzw. **creaming** – selekcji osób o najwyższym prawdopodobieństwie sukcesu – oraz **parking** osób szczególnie trudnych, których obsługa jest kosztowna, a szansa na dobry wynik wskaźnikowy niska. Wskaźnik, który miał zwiększyć orientację na **outcome**, może w ten sposób uderzyć w sprawiedliwość programu. Pojawia się tu konflikt między teorią efektywności a teorią sprawiedliwości: jeśli celem jest maksymalizacja liczby trwałych zatrudnień na jednostkę wydatku, selekcja uczestników jest racjonalna; jeśli jednak program ma zapewniać pierwszeństwo osobom w najgorszej sytuacji, taka strategia jest normatywnie wadliwa. System pomiarowy nie tylko obserwuje wartości – **materializuje hierarchię wartości w strukturze nagród**. Dlatego jednym z najważniejszych rozwinięć modelu Stevensona-Dodda powinno być pojęcie **adversarial metric design**: projektując KPI, należy wyobrazić sobie nie tylko zachowanie idealnego użytkownika, lecz także najbardziej racjonalny sposób uzyskania dobrego wyniku bez osiągnięcia rzeczywistego celu.

Jeżeli wskaźnik można łatwo „wygrać" poprzez zmianę sposobu rejestracji, selekcję populacji, przesunięcie kosztów, sztuczne rozdrobnienie zdarzeń albo skrócenie horyzontu obserwacji, jego wartość zarządcza jest ograniczona. Dobry miernik powinien być zatem nie tylko trafny statystycznie, lecz również odporny na manipulację strategiczną. Stevenson i Dodd wprowadzają w tym miejscu argument często pomijany w dyskusjach metodologicznych: **pomiar nie jest darmowy**.

Zbieranie danych pochłania czas pracowników, beneficjentów, analityków i kierownictwa oraz angażuje infrastrukturę IT. Autorzy proponują, aby organizacje wyceniały koszt dodatkowych obowiązków pomiarowych narzucanych przez fundatorów, a zasadniczy system wspólny ograniczały do niewielkiej liczby kluczowych **outcomes**. Argument ten można ująć w kategoriach ekonomicznego kosztu alternatywnego.

Nadmiar pomiaru osłabia realny wpływ społeczny

Jeśli pracownik pierwszego kontaktu poświęca godzinę na wypełnianie formularza, czas ten zostaje odebrany pracy z człowiekiem. Jeśli beneficjent wypełnia dziesiątą ankietę, ponosi koszt czasu i obciążenia poznawczego, a niekiedy musi ponownie przeżywać trudne doświadczenia. Z kolei gdy każda instytucja finansująca wymaga odrębnego zestawu kategorii, dane mogą stać się nieporównywalne, mimo ogromnej ilości pracy włożonej w ich zbieranie.

Nadmiar pomiaru może więc paradoksalnie zmniejszać sam wpływ, który miał mierzyć. Prowadzi to do pojęcia **value of information** – informacja ma wartość tylko wtedy, gdy realnie może zmienić decyzję. W ekonomii decyzji teoretycznie można wycenić, ile warto zapłacić za dodatkową informację przed dokonaniem wyboru. W praktyce organizacyjnej nie musi to przyjmować formy formalnego rachunku probabilistycznego; wystarczy dyscyplina pytania: „co zrobimy inaczej w zależności od wartości tego wskaźnika?". Jeśli odpowiedź brzmi „nic", uzasadnienie dla jego regularnego zbierania słabnie, chyba że wymaga tego prawo, audyt lub inny zewnętrzny obowiązek.

Taka selekcja wskaźników może być radykalniejsza niż zwykłe ograniczenie dashboardu. Organizacja powinna rozróżnić informacje niezbędne do decyzji strategicznych, dane operacyjne, wymogi regulacyjne, dane badawcze oraz te zbierane z przyzwyczajenia. Ta ostatnia kategoria bywa zaskakująco szeroka, gdyż bazy danych wykazują zależność od ścieżki: pole raz dodane do formularza często zostaje w nim przez lata tylko dlatego, że „zawsze je zbieraliśmy". Psychologia poznawcza wyjaśnia z kolei, dlaczego mnożenie KPI nie przekłada się liniowo na lepsze zarządzanie.

Uwaga i pamięć robocza są ograniczone. Decydent otrzymujący pięćdziesiąt wskaźników nie analizuje ich wszystkich z jednakową wagą; tworzy uproszczenia, skupiając się na najbardziej wyrazistych liczbach, ostatnich odchyleniach lub miernikach, które zna najlepiej. Nadmiar informacji może zatem zwiększać arbitralność selekcji zamiast ją ograniczać. Dashboard nie jest więc lustrem rzeczywistości, lecz interfejsem uwagi. Kolejność wskaźników, rozmiar wykresów, progi alarmowe, częstotliwość aktualizacji i sposób agregacji kształtują to, co kierownictwo uzna za pilne. Pozornie techniczne decyzje projektowe niosą konsekwencje strategiczne. Jeśli cash flow aktualizuje się codziennie i sygnalizuje czerwonym alarmem, a wellbeing beneficjentów oceniany jest raz na rok, system poznawczy zarządu naturalnie preferuje dane finansowe – nawet jeśli formalnie deklaruje, że impact jest centralnym elementem strategii. Proponowane przez Stevensona-Doddowskie trzy do pięciu wskaźników można zatem interpretować jako próbę zbudowania wąskiego pasma strategicznej uwagi.

Nie oznacza to likwidacji danych niższego poziomu, lecz konieczność, by członkowie zarządu wiedzieli natychmiast, które sygnały odpowiadają na pytanie o zasadniczą zmianę. Autor książki wiąże ten problem bezpośrednio z paraliżem analitycznym organizacji pozarządowych: rozbudowane raporty powstają często na niższych szczeblach, podczas gdy decyzje strategiczne pozostają zdominowane przez dane finansowe i operacyjne. Warto tutaj rozróżnić dwie funkcje wskaźników: prezentacyjną i diagnostyczną. Pierwsza odpowiada na pytanie, jak organizacja wygląda w oczach otoczenia; druga – co powinna zmienić. Ten sam miernik może pełnić obie funkcje, choć nie zawsze tak jest.

Liczba osób „objętych oddziaływaniem" dobrze prezentuje się w raporcie, komunikując skalę, lecz niewiele mówi kierownictwu o mechanizmach skuteczności. Z kolei wysoki wskaźnik rezygnacji z programu po trzecim tygodniu jest komunikacyjnie nieatrakcyjny, ale zarządczo niezwykle użyteczny, gdyż wskazuje miejsce awarii Theory of Change. Organizacja nastawiona na reputację naturalnie wybiera informacje potwierdzające jej wielkość; organizacja nastawiona na uczenie – te, które ujawniają, gdzie jej model jest błędny.

Tutaj powraca rozróżnienie „Do-Difference". Organizacja typu „Do" może dysponować imponującym zestawem mierników, lecz większość z nich opisuje jedynie samą organizację. Organizacja „Difference" potrzebuje wskaźników opisujących zmianę u innych. Jeszcze dojrzalsza organizacja wymaga dodatkowo wskaźników diagnostycznych, dzięki którym wie nie tylko, czy outcome się poprawia, ale także które elementy mechanizmu wymagają korekty.

SROI jako narzędzie analizy a pułapka monetyzacji

Autor książki poświęca osobne miejsce metodzie Social Return on Investment. SROI dąży do wyrażenia wartości społecznej w języku zbliżonym do finansowego zwrotu, co pozwala porównywać efekty wytwarzane przy określonych nakładach. Autor słusznie zauważa podwójny charakter tego podejścia. Z jednej strony zmusza ono do uporządkowania nakładów (inputs), rezultatów (outcomes), efektu martwego (deadweight), atrybucji i innych korekt; z drugiej zaś może wytwarzać pozorną precyzję poprzez dobór finansowych proxy, skłaniając użytkowników do nadmiernej koncentracji na finalnym współczynniku. Obecne standardy Social Value International pokazują jednak, że SROI nie powinno być redukowane do prostej formuły „1 zł = X zł wartości społecznej". SVI opisuje dziś szerszy system ośmiu zasad, obejmujących włączenie interesariuszy, zrozumienie zmiany, wartościowanie tego, co istotne, materialność, zakaz overclaimingu, przejrzystość, weryfikację oraz responsywność.

Organizacja powinna więc nie tylko policzyć wartość, lecz przede wszystkim ustalić, dla kogo nastąpiła zmiana, co ma dla tych osób znaczenie, jakie skutki były pozytywne i negatywne oraz jaka część rezultatu może być uczciwie przypisana interwencji. Co więcej, współczesne standardy SVI wprost dopuszczają kontekstowość rezultatów i wartości, nie tworząc jednej uniwersalnej taksonomii wpływu. SVI wyraźnie stwierdza, że ramy zasad mają zwiększać porównywalność i wiarygodność procesu, ale nie prowadzą automatycznie do systemu pomiaru typu „one-size-fits-all". Jest to ważna korekta wobec zbyt mocnej krytyki SROI zawartej w książce. Problemem nie jest sama metodologia, lecz sposób jej użycia: SROI może być bogatym procesem analizy wartości albo zostać zredukowane do atrakcyjnej liczby wykorzystywanej marketingowo. Najpoważniejszy problem epistemologiczny pojawia się przy monetyzacji. Nie wszystkie wartości społeczne mają cenę rynkową, dlatego SROI posługuje się proxy: kosztami unikniętymi, willingness-to-pay, wellbeing valuation lub innymi przybliżeniami. Zabieg ten umożliwia wprowadzenie wspólnej jednostki rachunku, lecz jednocześnie wprowadza arbitralność – różne, choć rozsądne proxy mogą prowadzić do odmiennych wyników.

Właśnie dlatego wartość końcowa powinna być przedstawiana wraz z założeniami i analizą wrażliwości, a nie jako odkryty odpowiednik „społecznego PKB" programu. SROI przypomina model ekonomiczny: jego siła polega na wymuszaniu jawności założeń i konsekwencji, a nie na magicznej przemianie wartości moralnych w obiektywne pieniądze. Model jest najbardziej użyteczny wtedy, gdy zmusza zarząd do konfrontacji z pytaniem, dlaczego jeden rezultat uznaje za ważniejszy od drugiego i jak wynik zmienia się po modyfikacji parametrów. Staje się niebezpieczny natomiast wtedy, gdy współczynnik finalny zaczyna funkcjonować jako ranking moralny programów.

Interoperacyjność zamiast standaryzacji i narracja zakotwiczona w danych

Autor odróżnia SROI od bardziej ogólnych ram Social Value i sugeruje "drogę środka": wymogi formalne można realizować w sposób proporcjonalny, natomiast wewnętrzny system zarządczy powinien łączyć kilka zweryfikowanych wskaźników z jakościowym wyjaśnieniem transformacji. Ta intuicja jest trafna, ale wymaga aktualizacji. W brytyjskim systemie zamówień publicznych Social Value Model nadal ewoluuje. W sierpniu 2026 r. Cabinet Office opublikował PPN 026, który ma być stosowany do zamówień rozpoczynanych od 1 stycznia 2027 r.; model koncentruje obecnie społeczną wartość między innymi na dobrych miejscach pracy, umiejętnościach i szansach dla

lokalnych społeczności, a deklarowane zobowiązania mają być monitorowane poprzez mechanizmy kontraktowe i KPI.

Ten przykład dobrze obrazuje różnicę między social value jako kategorią normatywną a Social Value Model jako instrumentem administracyjnym. W pierwszym przypadku pytamy o pełny bilans znaczących zmian społecznych i środowiskowych. W drugim – potrzebujemy operacyjnych kategorii, które można wpisać do przetargu, ocenić, zakontraktować i monitorować. Administracja musi redukować złożoność, ponieważ kontrakt nie może opierać się wyłącznie na ogólnych deklaracjach dobra społecznego. Każda taka redukcja niesie jednak ryzyko, że wykonawcy zaczną optymalizować jedynie zakontraktowane elementy, nawet jeśli nie wyczerpują one pełnego pojęcia wartości społecznej. To napięcie jest nieusuwalne. Nie da się zarządzać bez uproszczenia, ale każde uproszczenie może deformować zarządzany system. Dojrzałość nie polega więc na znalezieniu idealnej taksonomii, lecz na projektowaniu systemów zdolnych rozpoznawać własne ślepe plamy.

Kolejnym problemem jest standaryzacja. Autor jest sceptyczny wobec wielkich uniwersalnych ram, argumentując, że setki wskaźników mogą być zbyt abstrakcyjne dla bieżącego zarządzania. Jako przykład przywołuje globalny system SDG i wskazuje liczbę 232 wskaźników. Tu potrzebna jest aktualizacja faktograficzna: po kompleksowym przeglądzie w 2025 r. oficjalny globalny framework ONZ obejmuje obecnie 234 unikalne wskaźniki, przy 251 pozycjach na liście, ponieważ część z nich powtarza się pod kilkoma celami.

Sama korekta liczby jest mniej istotna niż problem, który odsłania. Globalny wskaźnik SDG i miernik zarządczy lokalnej organizacji pełnią odmienne funkcje. Wskaźniki SDG mają umożliwiać monitorowanie celów na poziomie państw i świata; organizacja pierwszego kontaktu potrzebuje danych pozwalających korygować codzienną interwencję.

Oczekiwanie, że jedna taksonomia spełni obie funkcje, jest błędem kategorii. Abstrakcyjność globalnego wskaźnika nie jest wadą, jeśli służy porównaniu makro; staje się problemem dopiero wtedy, gdy próbuje się go mechanicznie użyć jako KPI małej organizacji. Dlatego zamiast dążyć do pełnej standaryzacji, bardziej rozsądna jest interoperacyjność. Organizacje mogą posiadać własne mierniki outcomes odpowiadające Theory of Change, mapując je jednocześnie do wyższych kategorii wspólnych dla sektora, państwa czy SDG. Analogicznie w informatyce różne systemy nie muszą przechowywać identycznych danych, jeśli posiadają uzgodnione interfejsy i semantyczne mapowanie.

Taki model pozwala zachować lokalną trafność bez utraty możliwości agregacji. Różnica między standaryzacją a interoperacyjnością ma również znaczenie polityczne. Pełna standaryzacja przenosi władzę definiowania dobra do podmiotu tworzącego ramę. Interoperacyjność pozostawia część tej

władzy społecznościom i organizacjom, wymagając jedynie, aby potrafiły wyjaśnić relację między własnymi kategoriami a wspólnym językiem. Dla sektora społecznego może to być rozwiązanie bardziej pluralistyczne.

Autor trafnie odrzuca również fałszywy wybór pomiędzy "twardymi liczbami" a "miękkimi historiami". Proponuje model Data + Story, w którym wskaźniki dostarczają struktury dowodowej, natomiast narracja jakościowa wyjaśnia kontekst i mechanizm zmiany. To podejście zasługuje na rozwinięcie, ponieważ storytelling bywa zarazem jednym z najcenniejszych i najbardziej manipulowalnych narzędzi komunikacji wpływu. Historia pojedynczego beneficjenta może ujawnić zjawiska niewidoczne w średniej: mechanizm, punkt przełomowy, znaczenie relacji, barierę proceduralną czy niezamierzony skutek. W badaniach jakościowych przypadek nie musi reprezentować statystycznie populacji, aby posiadać wysoką wartość wyjaśniającą. Może być przykładem krytycznym, przypadkiem odstającym, dowodem mechanizmu albo źródłem nowej hipotezy.

Jednocześnie organizacja dysponująca tysiącem przypadków może zawsze znaleźć historię spektakularnego sukcesu. Jeżeli wybór historii następuje po poznaniu wyniku i bez informacji o rozkładzie wszystkich rezultatów, storytelling może stać się cherry-pickingiem. Jeden wzruszający przypadek nie mówi, jak często podobna transformacja zachodzi ani ilu osobom program nie pomógł. Historia wymaga zatem zakotwiczenia w denominatorze: odbiorca powinien wiedzieć, czy opisany przypadek jest typowy, wyjątkowy, czy dobrany dlatego, że ujawnia konkretny mechanizm. Naukowo dojrzała narracja wpływu powinna łączyć trzy rodzaje wiedzy:

Dane ilościowe odpowiadają głównie na pytania "ile?", "jak często?" i "jak duża jest różnica?". Dane jakościowe pomagają zrozumieć "jak?" i "dlaczego?". Theory of Change pozwala zaś umieścić oba rodzaje dowodu w strukturze przyczynowej.

Mierzyć mniej, ale lepiej i w równowadze

Żaden z tych elementów nie jest w pełni samowystarczalny. Krytyka nadmiaru wskaźników nie może prowadzić do drugiej skrajności – tworzenia za każdym razem całkowicie autorskich mierników. W wielu dziedzinach nauki wypracowano bowiem walidowane instrumenty, których właściwości psychometryczne zostały wielokrotnie sprawdzone. Dobrym przykładem są skale WEMWBS oraz ich krótsza wersja SWEMWBS, służące do pomiaru dobrostanu psychicznego. Warwick Medical School wskazuje, że narzędzia te zostały szeroko zwalidowane w populacjach dorosłych i badane w innych grupach, a ich wyniki są wrażliwe na zmiany wywołane interwencjami w obszarze dobrostanu. Znaczenie walidowanego instrumentu jest dwojakie: po pierwsze,

ogranicza ryzyko stworzenia pytania, które brzmi intuicyjnie, lecz nie mierzy stabilnie interesującego nas konstruktu; po drugie, zwiększa możliwość porównywania wyników.

Nie oznacza to jednak automatycznego transferu między kulturami, populacjami i kontekstami. Walidacja jest bowiem właściwością użycia instrumentu w określonych warunkach, nie magiczną pieczęcią przenoszącą się bez ograniczeń na każdą populację. Zasada „mierzyć mniej" powinna zatem iść w parze z zasadą „mierzyć lepiej". Redukcja liczby źle dobranych wskaźników nie zwiększa wiedzy. Wybór jednego miernika może być bardziej odpowiedzialny niż wybór dziesięciu, ale tylko wtedy, gdy jest on trafny, czuły na zmianę, interpretowalny i użyteczny decyzyjnie.

Najważniejsza korekta wobec radykalnego minimalizmu Stevenson-Dodda dotyczy ryzyka single-metric governance. Im bardziej organizacja redukuje system do jednego lub kilku wskaźników, tym silniejsza może być presja Goodharta. Ograniczenie liczby mierników zmniejsza przeciążenie informacyjne, ale jednocześnie zwiększa wagę każdego z pozostałych. Jeśli jeden KPI zaczyna decydować o wszystkim, bodziec do jego optymalizacji staje się ogromny. Rozwiązaniem nie musi być powrót do setek miar, lecz balanced measurement. Kluczowy outcome można zabezpieczać miernikami równoważącymi. Przykładowo: jeśli celem jest zwiększenie zatrudnienia, obok liczby osób zatrudnionych należy monitorować trwałość i jakość pracy oraz udział uczestników w najtrudniejszej sytuacji. Jeśli szpital skraca czas oczekiwania, potrzebuje równolegle miar bezpieczeństwa i jakości.

Analogicznie, urząd cyfryzujący usługę powinien monitorować nie tylko udział kanału online, ale również wykluczenie osób niezdolnych do korzystania z niego. Z punktu widzenia teorii systemów oznacza to, że wskaźnik główny powinien być otoczony constraint metrics – miarami pilnującymi, aby poprawa w jednym wymiarze nie następowała kosztem degradacji innego. Jest to dokładnie ten sam problem, który ekonomia definiuje jako optymalizację wielokryterialną. Organizacja społeczna rzadko bowiem posiada tylko jeden cel.

Materialność i rozdzielenie funkcji raportowania od zarządzania

Główny cel systemu jest ograniczony warunkami etycznymi, prawnymi, jakościowymi i sprawiedliwościowymi. W tym miejscu kluczowe staje się pojęcie materialności. Social Value International uznaje zasadę „Only include what is material": raport powinien obejmować te informacje, których pominięcie mogłoby zniekształcić prawdziwy i uczciwy obraz wartości wytwarzanej lub niszczonej przez dane działanie. Materialność jest zasadą dojrzalszą niż proste

„mierz tylko pięć rzeczy”, ponieważ łączy redukcję z odpowiedzialnością. Nie trzeba raportować wszystkiego, ale nie wolno pominąć czegoś istotnego tylko dlatego, że wynik wygląda źle lub pomiar jest trudny. Pojęcie to pozwala również rozstrzygnąć problem negatywnych skutków. Organizacja może posiadać trzy znakomite KPI, lecz jeśli jej program generuje poważny efekt uboczny nieuwzględniony w systemie, minimalizm pomiarowy staje się epistemicznym kamuflażem. Materialność wymaga zatem uzupełnienia informacji w momencie, gdy rzeczywistość okaże się bogatsza niż pierwotna architektura mierników. Dlatego dobry system jest minimalny, ale otwarty. Posiada niewielki rdzeń, lecz może się rozszerzać, gdy teoria zmiany, dane jakościowe albo anomalie wskazują na nowy materialny wymiar. Prostota nie jest tu zamknięciem katalogu, lecz dyscypliną pierwszeństwa.

Należy rozdzielić dwa cele pomiaru, które organizacje często próbują realizować jednocześnie. Raportowanie służy rozliczalności wobec fundatorów, regulatorów, obywateli, inwestorów i interesariuszy. Zarządzanie wpływem służy natomiast zmianie działania. Pierwsza funkcja wymaga stabilności definicji i porównywalności okresów; druga zdolności adaptacji i eksperymentowania. Raportowanie preferuje dane audytowalne i standaryzowane, podczas gdy zarządzanie wpływem może korzystać z wczesnych sygnałów, danych jakościowych i hipotez. Autor książki proponuje w tym kontekście „drogę środka”: formalne wymogi zewnętrzne należy realizować proporcjonalnie, natomiast wewnętrzny system opierać na kilku twardych miernikach oraz jakościowym opisie transformacji. Koncepcję tę można rozwinąć w model dwóch prędkości danych. Dane rozliczeniowe powinny być stabilne, audytowalne i porównywalne, natomiast dane rozwojowe mogą być szybsze, bardziej eksperymentalne i adaptacyjne.

Próba zbudowania jednego systemu idealnego dla obu tych funkcji może sprawić, że żaden z nich nie będzie efektywny. Rozróżnienie to jest szczególnie ważne w administracji publicznej. Instytucja nie może co kwartał zmieniać oficjalnej definicji rezultatu tylko dlatego, że uczy się nowych rzeczy, gdyż utraciłaby możliwość porównania i rozliczalności. Może jednak równolegle prowadzić elastyczniejsze analizy diagnostyczne. System dojrzały łączy więc stały kręgosłup rozliczalności z ruchomą warstwą uczenia się. Najgłębszym problemem kultury wskaźnikowej jest bowiem możliwość delegowania odpowiedzialności moralnej na liczbę.

Menedżer może stwierdzić: „KPI został wykonany”, jakby samo spełnienie miernika rozstrzygało o słuszności decyzji. Tymczasem każdy KPI jest produktem wcześniejszego wyboru: ktoś zdecydował, co liczyć, gdzie ustawić próg, jak długo obserwować rezultat i które skutki pominąć. Liczba nie eliminuje osądu. Przesuwa osąd wstecz, do momentu projektowania miary. Jeżeli ten etap zostanie ukryty, wskaźnik zyskuje pozór neutralności.

Dlatego dobra governance powinna traktować definicje KPI jako decyzje strategiczne, podlegające podobnej kontroli jak budżet czy strategia. Zarząd powinien rozumieć, jakie założenia zawiera wskaźnik, jakie zachowania może wywołać, które grupy może dyskryminować i w jakich okolicznościach powinien zostać zmieniony.

Prostota pomiaru jako wynik złożonego rozumowania

W tym miejscu spotykają się myśli Stevenson-Dodda, Goodharta, współczesne podejście do Social Value oraz teoria decyzji. Wszystkie te tradycje, choć wychodzą z różnych punktów, prowadzą do podobnego wniosku: **pomiar nie jest neutralnym dodatkiem do zarządzania**. Jest on integralną częścią samego mechanizmu zarządzania.

Dobra organizacja nie powinna zatem aspirować do „data-driven management" rozumianego jako zastąpienie ludzkiego osądu danymi. Trafniejszym ideałem jest **evidence-informed judgement**: decyzja powinna być korygowana i ograniczana przez dowody, lecz odpowiedzialność za nią musi pozostać po stronie człowieka. Dane mogą wskazać, że dany outcome rośnie, ale nie odpowiedzą samodzielnie na pytanie, czy koszt tej poprawy jest sprawiedliwy, czy rezultat jest wystarczająco istotny, czy efekt uboczny jest moralnie akceptowalny lub czy społeczeństwo akceptuje taką wymianę wartości. Postulat Stevenson-Dodda, by „mierzyć mniej", okazuje się zatem bardziej wymagający, niż mogłoby się wydawać. Minimalizm pomiarowy nie może oznaczać antyintelektualizmu ani rezygnacji z rzetelnej metodologii.

Aby odpowiedzialnie wybrać trzy wskaźniki, trzeba często najpierw zrozumieć trzydzieści innych możliwości. Niezbędna jest znajomość Theory of Change, mechanizmów przyczynowych, ryzyk Goodharta, właściwości psychometrycznych, kosztów zbierania danych, oczekiwań interesariuszy oraz potencjalnych skutków niezamierzonych i możliwości manipulacji. **Prostota na końcu jest rezultatem złożonego rozumowania na początku.**

Dobry system mierzenia wpływu przypomina zatem bardziej kokpit samolotu niż archiwum. Nie bombarduje on pilota wszystkimi danymi z każdego sensora z taką samą intensywnością; prezentuje przede wszystkim sygnały niezbędne do podjęcia bezpiecznej decyzji, zachowując jednocześnie możliwość zejścia do szczegółu w razie anomalii. Zarząd organizacji społecznej potrzebuje analogicznej struktury: kilku wskaźników wysokiego poziomu, dostępu do warstwy diagnostycznej, zdolności analizy przyczyn oraz systemu wykrywania nieprzewidzianych skutków. Nie należy więc przeciwstawiać dużych zbiorów danych małej liczbie wskaźników. Organizacja może dysponować ogromną bazą danych, stosując jednocześnie bardzo prosty interfejs decyzyjny. Może wykorzystywać AI do analizy tysięcy obserwacji, by zarządowi przedstawić jedynie trzy

dynamiczne sygnały. Kluczowym pytaniem nie jest ilość przechowywanej informacji, lecz **architektura przejścia od danych do decyzji**.

Dopiero w tym kontekście właściwe znaczenie zyskuje krytyka „paraliżu analitycznego" sformułowana przez Stevenson-Dodda. Nie jest ona argumentem przeciwko nauce, statystyce czy złożonym modelom, lecz przeciwko systemowi, w którym koszt poznawania przewyższa wartość samego poznania, wskaźniki wypierają cel, a produkcja raportów staje się odrębną gałęzią działalności. Właściwa odpowiedź nie polega na mierzeniu mniej dla samej prostoty, lecz na wymogu, by każda dodatkowa zmienna uzasadniała swoje miejsce poprzez znaczenie dla decyzji, trafność wobec Theory of Change lub obowiązek rozliczalności. To prowadzi naturalnie do kolejnego problemu: **jak mierzyć to, co wymyka się prostemu liczeniu**.

Będzie to zderzenie danych ilościowych i jakościowych, walidowanych skal z obserwacją doświadczenia, psychometrii z fenomenologią, średnich populacyjnych z heterogenicznością jednostek oraz klasycznych ankiet z możliwościami NLP i sztucznej inteligencji. Centralnym pytaniem nie będzie już to, czy „miękkie" zjawiska można mierzyć, lecz **co właściwie dzieje się z człowiekiem, kiedy jego dobrostan, godność, zaufanie, sprawczość albo poczucie bezpieczeństwa zostają przekształcone w zmienną**.

> Ostatecznie, dążenie do absolutnej precyzji w liczbach może stać się formą ucieczki przed odpowiedzialnością za realnego człowieka. Prawdziwa dojrzałość pomiaru nie objawia się w głośności roszczeń o sukces, lecz w dokładnym wyznaczeniu granic własnego sprawstwa. Czy zatem w świecie zdominowanym przez dashboardy i KPI potrafimy jeszcze odróżnić organizację, która jedynie sprawnie zarządza danymi, od tej, która faktycznie zmienia życie?

Inspiracje

[1] 

"The Difference How to Understand Measure and Maximise Social Impact" Matt Stevenson-Dodd

Practical Inspiration Publishing | ISBN: 9781788609562

Matt Stevenson-Dodd to brytyjski lider sektora społecznego, konsultant i autor specjalizujący się w pomiarze wpływu społecznego, zarządzaniu organizacjami charytatywnymi i strategiach non-profit. Jest założycielem i dyrektorem generalnym Trust Impact, firmy konsultingowej założonej w 2019 roku, która pomaga organizacjom non-profit, przedsiębiorstwom społecznym i instytucjom publicznym w transparentnym pomiarze i komunikowaniu wpływu. Przed założeniem Trust Impact, Stevenson-Dodd spędził ponad dwie dekady jako dyrektor organizacji charytatywnych, w tym w latach 2010-2019 pełniąc funkcję dyrektora generalnego Street League, wiodącej brytyjskiej organizacji charytatywnej promującej sport dla pracowników. Tam był pionierem w zakresie transparentnego raportowania wpływu i udostępniania danych w czasie rzeczywistym, zyskując uznanie w całym sektorze za szczere ujawnianie niedociągnięć organizacyjnych. Ukończył studia MBA z wyróżnieniem na Uniwersytecie w Nottingham i studiował w Harvard Business School. W 2020 roku został mianowany Czcigodnym Czcigodnym Zakonem Świętego Jana.

Matt Stevenson-Dodd *is a British social sector leader, consultant, and author specializing in social impact measurement, charity governance, and non-profit strategy. He is the founder and chief executive officer of Trust Impact, a consultancy established in 2019 to help non-profits, social enterprises, and public bodies measure and communicate impact transparently. Prior to establishing Trust Impact, Stevenson-Dodd spent over two decades as a charity executive, notably serving from 2010 to 2019 as CEO of Street League, the UK's leading sport-for-employment charity. There, he pioneered transparent impact reporting and real-time open data dashboards, gaining sector-wide recognition for candidly disclosing organizational shortcomings. He holds an MBA with Distinction from Nottingham University and studied at Harvard Business School. In 2020, he was appointed a Member of the Venerable Order of Saint John.*

Trust Impact

Literatura uzupełniająca

Książki

Książki autorów przywoływanych i literatura pokrewna:

[1] "The Journal of John Mayne During a Tour on the Continent Upon Its Reopening After the Fall of Napoleon, 1814. Edited by His Grandson John Mayne Colles. With Numerous Illustrations"

John MAYNE (of the Irish Bar.)

1909



[2] "How did Britain come to this? A century of systemic failures of governance"

Gwyn Bevan

LSE Press | ISBN: 9781911712114

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] "Does 'naming and shaming' work for schools and hospitals? Lessons from natural experiments following devolution in England and Wales"

Gwyn Bevan, Deborah Wilson

2013 | Public Money & Management | DOI: 10.1080/09540962.2013.799801

Źródło

[2] "Hitting and Missing Targets by Ambulance Services for Emergency Calls: Effects of Different Systems of Performance Measurement within the UK"

Gwyn Bevan, Richard Hamblin

2008 | Journal of the Royal Statistical Society Series A (Statistics in Society) | DOI: 10.1111/j.1467-985x.2008.00557.x

Źródło

[3] "A scoping review protocol to map the evidence on interventions to prevent overweight and obesity in children"

Peer-Benedikt Vincent Bussiek, Chiara De Poli, G. Bevan

2018 | BMJ Open | DOI: 10.1136/bmjopen-2017-019311

Źródło

[4] "The Tools of Government"

Christopher C. Hood

1983 | DOI: 10.1007/978-1-349-17169-9

Źródło

[5] **"The Art of the State: Culture, Rhetoric, and Public Management"**
Christopher C. Hood
1998
Źródło

[6] **"Government as a Tool-Kit"**
Christopher C. Hood
1983 | DOI: 10.1007/978-1-349-17169-9_7
Źródło

[7] **"Tools Approaches"**
Helen Zerlina Margetts, Christopher C. Hood
2026 | International series on public policy | DOI: 10.1007/978-3-032-06026-6_8
Źródło

Treść tworzy Fundacja Dobre Państwo.
Redaguje i publikuje APA ONE, autorski system redakcyjny Fundacji oparty na AI.
APA ONE Publishing System
Fundacja Dobre Państwo © 2026
Article ID: 909e9bb0-a702-4566-bffd-410414a01396