UNIWERSYTET WARSZAWSKI
CENTRUM EUROPEJSKICH STUDIÓW REGIONALNYCH I LOKALNYCH

Mechanizmy wykorzystania ewaluacji

Studium ewaluacji średniookresowych INTERREG III

Karol Olejniczak

Publikacja finansowana przez Unię Europejską

Scholar
WYDAWNICTWO NAUKOWE

Mechanizmy wykorzystania ewaluacji

Serdecznie dziękuję profesorowi Grzegorzowi Gorzelakowi za przekonanie mnie do pomysłu „ewaluacji" ewaluacji i pomoc w ukierunkowaniu koncepcji badania, profesorowi Jackowi Szlachcie za cenne uwagi merytoryczne, profesorowi Johnowi Bachtlerowi za udostępnienie danych i komentarz do struktury książki, zaś zespołowi EUROREG za koleżeńską pomoc i przyjazną, twórczą atmosferę.

Dziękuję również moim najbliższym – Rodzicom, Molly, bratu Tomkowi i Luizie za cierpliwość, codzienne wsparcie i mobilizowanie mnie do pracy.

Książkę chciałbym zadedykować moim Rodzicom; dzięki Nim wszystko stało się możliwe.

CENTRUM EUROPEJSKICH STUDIÓW
REGIONALNYCH I LOKALNYCH UW

Mechanizmy wykorzystania ewaluacji

Studium ewaluacji średniookresowych INTERREG III

Karol Olejniczak

Wydawnictwo Naukowe
SCHOLAR

Recenzent: *prof. dr hab. Jacek Szlachta*

Redakcja i korekta: *Sonia Ruszkowska*

Adaptacja projektu okładki: *Katarzyna Juras*

Copyright © by Karol Olejniczak, 2008

Badanie jest efektem projektu naukowego finansowanego przez Ministerstwo Edukacji i Nauki (Nr 1 H02D 019 30)

Publikacja została sfinansowana ze środków Unii Europejskiej w ramach Projektu Środki Przejściowe PL2005/017-488.01.02.06 „Wzmocnienie zdolności ewaluacyjnych w sektorze publicznym". Za jej treść odpowiada Autor. Poglądy w niej wyrażone nie odzwierciedlają w żadnym razie oficjalnego stanowiska Unii Europejskiej

ISBN 978-83-7383-306-7

Wydawnictwo Naukowe Scholar Spółka z o.o.
ul. Krakowskie Przedmieście 62, 00-322 Warszawa
tel./fax 022 828 93 91, 022 826 59 21, 022 828 95 63
dział handlowy: 022 635 74 04 w. 219 lub jw. w. 105, 108
e-mail: info@scholar.com.pl; scholar@neostrada.pl
www.scholar.com.pl

Wydanie pierwsze
Skład i łamanie: WN Scholar (*Stanisław Beczek*)
Druk i oprawa: Paper & Tinta, Warszawa

SPIS TREŚCI

1. Wprowadzenie 9
1.1. Zarys tematyki i przedmiotu badania 9
1.2. Cel książki 13
1.3. Struktura książki 14
2. Teoretyczne podstawy wykorzystania ewaluacji 16
2.1. Wstęp 16
2.2. Ewaluacja – definicje, typy, funkcje 16
2.2.1. Definicje ewaluacji i jej korzenie teoretyczne 16
2.2.2. Funkcje ewaluacji w sektorze publicznym 21
2.2.3. Typologia ewaluacji 25
2.3. Koncepcje wykorzystania ewaluacji 39
2.3.1. Rodzaje wykorzystania ewaluacji 39
2.3.2. Czynniki wpływające na użycie ewaluacji 44
2.3.3. Ewaluacja w kontekście teorii efektywnego uczenia się . . 48
2.3.4. Próby ujęcia systemowego 56
3. Praktyka ewaluacji w sektorze publicznym 64
3.1. Wstęp 64
3.2. Ewaluacja w reformach zarządzania sektorem publicznym . . . 64
3.2.1. Administracja i zarządzanie w sektorze publicznym – uwagi wstępne 64
3.2.2. Początki ewaluacji w Klasycznej Administracji Publicznej 66
3.2.3. Ewaluacja w Neoklasycznej Administracji Publicznej . . 67
3.2.4. Menedżeryzm i użycie ewaluacji 68
3.2.5. Rola ewaluacji w Nowym Zarządzaniu Sektorem Publicznym (NPM) 71
3.2.6. Ewaluacja w reformach sektora publicznego na początku XXI wieku 73
3.3. Praktyka ewaluacji w Unii Europejskiej 79
3.3.1. Ewaluacja w administracjach krajów członkowskich UE . 79
3.3.2. Ewaluacja programów współfinansowanych przez UE . . 83
3.3.3. Praktyka ewaluacji – główne podejścia metodologiczne UE 87
3.3.4. Praktyka ewaluacji programów regionalnych – przebieg badań 94
4. Studium wykorzystania ewaluacji w programach INTERREG III 105
4.1. Wstęp 105
4.2. Charakterystyka Inicjatywy Wspólnotowej INTERREG 105
4.2.1. Współpraca terytorialna jako element Polityki Spójności UE 105
4.2.2. INTERREG jako narzędzie współpracy terytorialnej . . . 110
4.2.3. Organizacja i zarządzanie INTERREG III 113
4.2.4. Specyfika ewaluacji programów współpracy terytorialnej 119
4.3. Model analizy wykorzystania ewaluacji 123

4.3.1. Prezentacja modelu 123
4.3.2. Zastosowanie modelu w studiach przypadków 133
4.4. Studia przypadków 134
4.4.1. Wybór studiów przypadków 135
4.4.2. Przypadek 1: IIIA–Austria – Czechy, Słowacja, Słowenia, Węgry 140
4.4.2. Przypadek 2: IIIA–Irlandia-Irlandia Północna 152
4.4.3. Przypadek 3: IIIB–Europa Północno-Zachodnia 163
5. Wnioski i rekomendacje 174
5.1. Wnioski z badania 174
5.2. Szersze wnioski odnośnie do zarządzania 179
5.3. Sugestie praktyczne 181
6. Aneksy 187
6.1. Metodologia badania 187
Pytania, hipotezy badawcze 187
Uzasadnienie zakresu badania 188
Uzasadnienie podejścia badawczego 188
Plan badawczy 191
Metodologia studiów przypadków 192
Scenariusz wywiadu z przedstawicielami instytucji zlecającej ewaluację 194
Scenariusz wywiadu z ewaluatorami 195
Lista wywiadów 196
6.2. Modele wykorzystania ewaluacji 197
6.3. Listy i mapy programów INTERREG III 206
Bibliografia 211

SPIS TABEL, SCHEMATÓW I MAP

Tabele

Tabela 1. Różnice między ewaluacją, badaniem naukowym, monitoringiem i audytem 18
Tabela 2. Definicja ewaluacji użyta w pracy 21
Tabela 3. Przeciwdziałanie dysfunkcjom zarządzania opartego na wynikach 23
Tabela 4. Zestawienie typologii ewaluacji 37
Tabela 5. Formularz oceny raportu ewaluacyjnego 55
Tabela 6. Odmienne filozofie sektora prywatnego i publicznego 65
Tabela 7. Ewaluacja w paradygmatach zarządzania sektorem publicznym 78
Tabela 8. Liczba ewaluacji zlecanych przez Komisję Europejską 85
Tabela 9. Ewaluacje polityki regionalnej UE w latach 1994–2006 (tylko UE15) 86
Tabela 10. Główne kryteria ewaluacyjne i pytania badawcze 93
Tabela 11. Oferta – zawartość pakietu merytorycznego 97
Tabela 12. Typowe kryteria oceny ofert 98
Tabela 13. Wybrane metody stosowane w ewaluacji 99
Tabela 14. Dobre praktyki raportów ewaluacyjnych 101
Tabela 15. Kanały komunikacji i odbiorcy ewaluacji 103
Tabela 16. Reformy Polityki Spójności Unii Europejskiej 107
Tabela 17. Ewolucja Inicjatywy INTERREG 112
Tabela 18. Typy projektów INTERREG. 115
Tabela 19. Wytyczne KE dotyczące zakresu ewaluacji średniookresowej INTERREG 119
Tabela 20. Przesłanki i rodzaje transferu wiedzy między instytucjami programów UE 138
Tabela 21. Raporty ewaluacyjne wybrane do badania 139
Tabela 22. Możliwości praktycznego zastosowania metamodelu 182

Aneks 6.3

Tabela 1. Programy INTERREG III A 2000–2004 i 2004–2006 206
Tabela 2. Programy INTERREG III B 2000–2004 i 2004–2006 208

Schematy

Schemat 1. Rodzaje użycia ewaluacji 43
Schemat 2. Logika planowania interwencji UE 89
Schemat 3. Logika wdrażania i monitorowania interwencji UE 90
Schemat 4. Cykl interwencji UE 91
Schemat 5. Kryteria ewaluacji a logika interwencji UE 92
Schemat 6. Relacje interesariuszy INTERREG w ewaluacji średniookresowej 118
Schemat 7. Kalendarz planowania i ewaluacji Inicjatywy Wspólnotowej INTERREG 120
Schemat 8. Kalendarz planowania i ewaluacji programów INTERREG dla Austrii 146
Schemat 9. Organizacja badania ewaluacyjnego *mid-term* IIIB NWE . 166
Schemat 10. Dylematy wyboru funkcji ewaluacji 178
Schemat 11. Podejścia ewaluacyjne a orientacja badań 182

Aneks 6.2

Schemat 1. Ramy analityczne użycia ewaluacji 198
Schemat 2. Model procesu ewaluacji 199
Schemat 3. Teoretyczny model wykorzystania ewaluacji 200
Schemat 4. Model logiczny ewaluacji 201
Schemat 5. Metamodel analizy wykorzystania ewaluacji 202
Schemat 6. Skorygowany metamodel analizy wykorzystania ewaluacji . 204

Mapy

Aneks 6.3

Mapa 1. Programy INTERREG IIIA w latach 2000–2004 i 2004–2006 . 209
Mapa 2. Programy INTERREG IIIB w latach 2000–2006 210

1. WPROWADZENIE

1.1. ZARYS TEMATYKI I PRZEDMIOTU BADANIA

Zagadnienie podnoszenia jakości programów publicznych od lat znajduje się w centrum uwagi kolejnych reform zarządzania sektorem publicznym – począwszy od wprowadzanych w latach 60. PPBS (Planowanie–Programowanie–System–Budżetowy, ang. *Planning, Programming, Budgeting System*) i Zarządzanie przez Cele, (ang. *Management by Objectives* – MBO), poprzez ruch Nowego Menedżeryzmu (ang. *New Managerialism*) w latach 80., aż po dominujący obecnie nurt Nowego Zarządzania Sektorem Publicznym (ang. *New Public Management*) i rodzący się nurt *New Public Governance* (Heinrich, 2003; OECD, 2005; Osborne, 2006; Shafritz et al., 2007).

Jednym z narzędzi, po które sięgano we wszystkich tych reformach, jest ewaluacja. Ewaluacja to – w najszerszym ujęciu – użycie metod badań społeczno-ekonomicznych do systematycznego zbierania danych, analizy, interpretacji, oceny i informowania o skuteczności programów publicznych (Mathison, 2005; Rossi et al., 1999). Najogólniej rzecz ujmując, ewaluacja jest narzędziem uczenia się sektora publicznego i jest prowadzona w celu podniesienia jakości badanego programu, jak i przyszłych, analogicznych działań (Chelimsky, 1985: 7; Cronbach, Associates, 1981: 14; Rist, 1990: 7).

Ewaluacja programów publicznych sięga początków lat 60. (reform związanych z realizowanym w Stanach Zjednoczonych programem *Great Society*) (Cronbach, Associates, 1981; Derlien, 1990). Obecnie jest ugruntowaną praktyką w administracjach wielu rozwiniętych krajów gospodarki rynkowej (Bachtler, 2000; Furubo et al., 2002; Mathison, 2005; OECD, 1999). Znaczący wzrost wydatków w ramach polityk prowadzonych na poziomie Unii Europejskiej połączony z koniecznością podniesienia wiarygodności i przejrzystości działań UE sprawił, że ewaluacja stała się również standardem i wymogiem wszystkich programów Unii Europejskiej (Council of the European Union, 1999: art. 40–43; 2002: art. 27.4; European Commission, 2000c, 2001f). Jednym z pól, na których praktyka ewaluacji rozwinęła się najszerzej, jest polityka regionalna Unii Europejskiej[1].

Przy tak powszechnym wykorzystaniu ewaluacji nasuwa się pytanie o to, ***jaką rzeczywistą rolę odgrywają raporty ewaluacyjne w zarządzaniu programami Unii Europejskiej. Czego i w jaki sposób instytucje i pracownicy sektora publicznego uczą się z ewaluacji?*** Pytanie o rzeczywiste wykorzystanie

[1] Dla dwóch ostatnich okresów programowania (lata 1994–1999 i 2000–2006) tylko w samych krajach „starej" Unii Europejskiej wykonano dotychczas ponad 2000 badań ewaluacyjnych. Po rozszerzeniu UE liczba ta co najmniej się podwoiła.

ewaluacji w zarządzaniu programami publicznymi oraz o czynniki je warunkujące jest centralnym zagadnieniem niniejszej książki.

Problem rzeczywistego wpływu ewaluacji na jakość zarządzania został przedstawiony na gruncie analizy najlepszych jakościowo raportów z ewaluacji *mid-term* (średniookresowej) programów europejskiej współpracy terytorialnej – Inicjatywy Wspólnotowej INTERREG III. Wybór Inicjatywy Wspólnotowej INTERREG jako pola badań wykorzystania ewaluacji programów publicznych jest uzasadniony z trzech względów.

Po pierwsze, INTERREG, którego celem jest wspieranie współpracy terytorialnej (transgranicznej, transnarodowej i międzyregionalnej) pomiędzy regionami UE został zainicjowany przez Komisję Europejską i był nakierowany na ogólnoeuropejskie wyzwania. Co więcej, w odróżnieniu od głównego nurtu funduszy strukturalnych, programy INTERREG były wdrażanie nie przez pojedyncze administracje krajowe, a przez instytucje z różnych krajów w ramach współpracy ponad granicami. To tworzyło prawdziwie ponadnarodowe, międzykulturowe struktury i procesy zarządzania. INTERREG był więc inicjatywą *sensu stricte* europejską, co ewaluacji umożliwiało wyjście poza praktyki narodowe i łączenie różnych podejść i stylów badania.

Po drugie, z racji ograniczonych środków INTERREG w mniejszym stopniu koncentrował się na wyrównywaniu poziomów i infrastrukturze, a w większym na tworzeniu połączeń funkcjonalnych między krajami i instytucjami. Interakcje i współpraca między aktorami z różnych krajów są podstawą efektywnych mechanizmów uczenia się. INTERREG był więc predestynowany do budowania wiedzy.

Po trzecie, inicjatywa INTERREG okazała się wyjątkowo nośna i trwała. Funkcjonuje w Europie od 1990 roku i była kontynuowana przez wszystkie kolejne okresy programowania. Co więcej, w nowym okresie programowania (2007–2013) wzrosła ranga współpracy terytorialnej – stała się ona jednym z trzech głównych celów nowej polityki spójności (Council of the European Union, 2005: art. 3.2c). Taka ciągłość daje możliwość śledzenia długookresowych efektów ewaluacji, szerszej refleksji oraz wyciągania ważkich wniosków na przyszłość.

Czytelnik może być rozczarowany, że książkę w całości oparto na zagranicznych studiach przypadków. Na usprawiedliwienie tej sytuacji mam kilka argumentów. Polska praktyka w 2004 roku (w momencie, gdy rozpoczynałem badanie) uniemożliwiała podjęcie systematycznej analizy. Populacja badań ewaluacyjnych była bardzo niejednolita – zróżnicowana pod względem jakości, tematów, stawianych wymogów i zakresu studiów, a także kontekstu badań (programy pomocowe i przedakcesyjne, a od 2004 roku badania funduszy strukturalnych). Natomiast programy INTERREG, mimo iż kulturowo zróżnicowane, były nakierowane na te same cele, posiadały jednolitą strukturę zarządzania, a ich ewaluacje średniookresowe przeprowadzono w tym samym czasie w całej Europie Zachodniej, zadając te same pytania. Co więcej, tak badania, jak i całe programy badania miały swoją kontynuację. To stworzyło wyjątkową okazję do przeprowadzenia usystematyzowanego, całościowego i porównawczego studium długookresowych efektów ewaluacji.

Prezentowana w książce tematyka, mimo iż oparta na analizie zagranicznych źródeł, zawiera jednak odniesienia do polskiej sytuacji. Co więcej, wnioski i rekomendacje z analizy zostały sformułowane pod kątem krajowych użytkowników ewaluacji.

Główną wartością książki jest przybliżenie polskiemu Czytelnikowi światowych koncepcji i badań roli ewaluacji w zarządzaniu programami publicznymi. Koncepcje te były rozwijane od początku lat 70. Wynikały z rozczarowania słabym użyciem wyników ewaluacji, które początkowo zakładano niejako *a priori*. Poszukiwanie mechanizmów wyjaśniających użycie ewaluacji, jej wpływ (bądź jego brak) na zarządzanie w sektorze publicznym, prowadzone jest więc od prawie półwiecza, jednak głównie w literaturze anglosaskiej[2]. W kolejnych dziesięcioleciach koncentrowano się na różnych aspektach kwestii wykorzystania ewaluacji. Początkowo badania skupiały się na wyjaśnieniu mechanicznego i technicznego użycia wyników (określeniu cech badania – jakości metod i wniosków gwarantujących wdrożenie rekomendacji). Z czasem zaczęto rozszerzać listy czynników potencjalnie wpływających na wykorzystanie ewaluacji. W połowie lat 80. zainteresowano się procesem ewaluacji – relacjami, interakcjami i predyspozycjami uczestników (zlecających, odbiorców, badaczy) jako wyjaśnieniem zróżnicowanego użycia. Użycie interpretowano już nie tylko instrumentalnie (wdrożenie rekomendacji), ale zaczęto dostrzegać jego aspekty interpersonalne – tzw. zmiany koncepcyjne (zmiany w wiedzy i świadomości uczestników ewaluacji). Wykorzystanie ewaluacji analizowano więc w kontekście procesów uczenia się jednostek i organizacji.

W końcu lat 90. sięgnięto po ujęcia systemowe, budowano modele łączące różne podejścia i perspektywy, próbujące uwzględnić zarówno proste czynniki techniczne (jakość raportu, poprawność metod), warunki kontekstowe, jak i relacje między uczestnikami, a nawet szerszym środowiskiem programu czy polityki. Część z rozwijanych modeli stara się nawet ujmować aspekty psychospołeczne procesu uczenia się i zmiany.

W ciągu pięćdziesięciu lat wypracowano dość jednolite rozumienie terminu użycia ewaluacji. Definiuje się go jako rzeczywisty wpływ i rolę, jaką spełnia dane badanie w odniesieniu do programu i sektora publicznego. Zidentyfikowano również szereg typów użycia ewaluacji: w zależności od przedmiotu użycia – wniosków z badania lub procesu badawczego, w zależności od charakteru i świadomości użytkowników, a także od perspektywy czasowej użycia. W latach 80. duży nurt badań koncentrował się na kwestii błędnego i złego użycia ewaluacji, wnosząc do literatury rozpoznanie możliwych ubocznych

[2] Głównymi periodykami naukowymi poświęconymi ewaluacji są: *American Journal of Evaluation*, *Evaluation & Program Planning*, *Evaluation Review*, *Evaluation Practice*, *New Directions for Evaluation*, *Studies in Educational Evaluation*. Wszystkie one wydawane są w Stanach Zjednoczonych. Jedyne europejskie wydawnictwo to kwartalnik *Evaluation* – publikacja naukowa prowadzona przez Europejskie Towarzystwo Ewaluacyjne. Artykuły na temat ewaluacji programów publicznych pojawiają się również w *Science Communication* (zwykle w kontekście analiz mechanizmów uczenia się i transferu wiedzy) oraz w kwartalniku *Regional Studies* (tu zaś w kontekście polityki regionalnej i funduszy strukturalnych UE) wydawanym przez Regional Studies Association. *Gros* publikacji książkowych na temat ewaluacji ukazuje się w wydawnictwie naukowym Sage (specjalizującym się w metodach badań społecznych).

i negatywnych efektów ewaluacji. Obecnie podkreślane są długoterminowe efekty ewaluacji, jej oddziaływanie wykraczające poza ramy jednego programu. Użycie ewaluacji przestaje być traktowane jako cel sam w sobie, umieszcza się je w szerszym kontekście funkcji sektora publicznego, wspierania demokracji i dobrobytu społecznego.

Zarówno koncepcje wykorzystania ewaluacji, jak i związane z nimi badania empiryczne są rozwijane poza Europą, głównie w USA. W badaniach używa się zróżnicowanej metodologii – głównie ilościowej, co nie oznacza, że są one pozbawione ograniczeń, takich jak pomijanie analiz koncepcyjnego i długookresowego wykorzystania ewaluacji, uwzględnianie tylko jednej perspektywy – perspektywy ewaluatorów, bazowanie głównie na opiniach i ocenach samych uczestników procesu, rzadko analizowany kontekst. Analogiczne badania w Unii Europejskiej są całkowicie incydentalne, ograniczone do analiz wykorzystania ewaluacji (głównie użycia instrumentalnego) w jednej tylko instytucji – Komisji Europejskiej. Mają one raczej charakter techniczny i konsultingowy niż naukowy, gdyż są robione na zlecenie Komisji i w niewielkim stopniu odnoszą się do szerszej dyskusji teoretycznej. W Polsce dotychczas niewiele pisano na temat ewaluacji[3]. Zaś temat wykorzystywania ewaluacji przez administrację publiczną jak dotąd poruszony został tylko raz (Górniak, 2005)[4].

Obecnymi wyzwaniami stawianymi we współczesnej literaturze dotyczącej wykorzystania ewaluacji są: wypracowywanie systemowych wyjaśnień wykorzystania ewaluacji, uchwycenie długoterminowego oddziaływania i efektów koncepcyjnych – wpływu na procesy uczenia się, a także wpisanie jej w szerszy kontekst funkcji sektora publicznego, roli we współczesnym rządzeniu (*governance*). Jak wspomniałem, dyskusja ta toczy się jednak głównie w USA. W Europie kwestie wykorzystania ewaluacji praktycznie nie są podejmowane w badaniach naukowych, mimo olbrzymiej skali wykonywanych ewaluacji dla programów UE oraz istniejącej ogólnoeuropejskiej standaryzacji, otwierającej pole do ciekawych analiz. W zasadzie użyteczność ewaluacji programów unijnych i związane z nią mechanizmy uczenia się są traktowane domyślnie, zakładane *a prori* i nie poparte wnioskami z badań na gruncie europejskim.

Niniejsza książka jest więc jedną z pierwszych analiz wykorzystania ewaluacji w programach Unii Europejskiej. Jest ona zorientowana praktycznie, lecz oparta na szerokiej bazie teoretycznej czerpiącej zarówno z bogatych źródeł amerykańskich, jak i istniejących źródeł europejskich. Prezentowane badanie jest pierwszym, w którym na gruncie polityki regionalnej UE zastoso-

---

[3] Dotychczasowe publikacje dotyczyły głównie pola edukacji (Korporowicz, 1997, 2000; Mizerek, 1997) albo też miały charakter podsumowań czy poradników prowadzenia ewaluacji (Kierzkowski, 2002b; Ministerstwo Rozwoju Regionalnego, 2007; Olejniczak, 2004). Nowsze publikacje prezentowały zaś wyniki samych badań ewaluacyjnych (Gorzelak et al. 2004). Badania empiryczne dotyczące ewaluacji jak dotychczas dotyczyły wyłącznie kształtującego się potencjału ewaluacyjnego polskiej administracji (Kierzkowski, 2002a; Olejniczak, 2002).

[4] Projekt koncentrował się jednak nie na zbadaniu mechanizmów uczenia się, ale na skonstruowaniu systemu wskaźników do oceny jakości ewaluacji, a także podniesienia zdolności syntezy wyników z ewaluacji przez jednostki samorządowe i wykorzystywania ich w kolejnych okresach programowania (a więc metaewaluacji regionalnej polityki publicznej – w tym wypadku województwa małopolskiego).

wano podejście systemowe i procesowe. Uwzględniono kwestie kontekstowe, role europejskich aktorów instytucjonalnych, mechanizmy uczenia się, a także szerszą perspektywę zjawiska wykorzystywania ewaluacji – zarówno czasową, jak i instytucjonalną (poziom nie tylko programu, ale również polityk europejskich). Badanie i jego wnioski odnoszą się do współczesnej dyskusji teoretycznej na temat roli ewaluacji w sektorze publicznym. Jednocześnie jednak mają silny element empiryczny osadzony w praktyce programów Unii Europejskiej. Umożliwiło to przedstawienie we wnioskach praktycznych obserwacji skierowanych tak do zlecających, jak i wykonawców ewaluacji europejskich programów współpracy terytorialnej.

1.2. CEL KSIĄŻKI

Wymóg ewaluacji programów europejskich oraz rosnące znaczenie programów współpracy terytorialnej sprawiają, że kluczowym staje się pytanie o rzeczywistą rolę ewaluacji w podnoszeniu jakości zarządzania tymi programami. To zagadnienie ważne jest również z perspektywy polskiej. Ewaluacja jest jednym z wymogów funduszy strukturalnych, a jej efektywne użycie mogłoby ułatwić polskiej administracji korzystanie ze środków unijnych.

Głównym celem książki jest przedstawienie, w oparciu o empiryczne studia przypadków, rzeczywistego wpływu wykonanych ewaluacji na zarządzanie programami INTERREG. Badanie przeprowadzono na programach INTERREG 2000–2006 z krajów Unii Europejskiej przed rozszerzeniem, gdyż w czasie prowadzenia badań (lata 2004–2006) w Polsce materiał badawczy ewaluacji bieżących praktycznie nie istniał, nie było też wystarczającego odstępu czasowego, by obserwować ewentualne długookresowe efekty ewaluacji.

Cele badawcze zostały ujęte na dwóch płaszczyznach – poznawczej i praktycznej.

Celem poznawczym było zbadanie wykorzystania badań ewaluacyjnych w programach INTERREG oraz wyróżnienie czynników wpływających na ten proces. Na podstawie literatury przedmiotu i praktyki unijnej stworzono model wykorzystania ewaluacji. Został on następnie zweryfikowany w odniesieniu do wybranych badań ewaluacyjnych europejskich programów współpracy terytorialnej.

Celem praktycznym było natomiast dostarczenie wniosków na temat efektywnego wykorzystania ewaluacji w zarządzaniu programami publicznymi. Wnioski te są adresowane do aktorów zaangażowanych w realizację badań ewaluacyjnych – do instytucji publicznych i prywatnych występujących po stronie zlecających i wykonawców.

Głównym rezultatem przedstawionego w książce badania jest zweryfikowany, rozszerzony metamodel wykorzystania ewaluacji w programach Unii Europejskiej. Model ten może być z powodzeniem używany zarówno przez zlecających, jak i wykonujących ewaluację do lepszego planowania i prowadzenia badań pod kątem ich pełnego wykorzystania. Ponadto prezentowane badanie wyjaśniło funkcje, czynniki i procesy, które charakteryzują wykorzystanie ewaluacji *mid-term* (średniookresowych) programów współpracy terytorialnej UE.

Wartość dodana prezentowanego badania obejmuje zarówno aspekty teoretyczne, jak i praktyczne. Wartymi podkreślenia są cztery kwestie.

Po pierwsze, prace teoretyczne na temat wykorzystania ewaluacji programów europejskich są bardzo rzadkie. Stoi to w kontraście z bogatą praktyką europejską i rozwiniętymi koncepcjami amerykańskimi. Książka po części wypełnia tę lukę. Jest głosem w dyskusji nad użytecznością ewaluacji w programach współfinansowanych z budżetu Unii Europejskiej.

Po drugie, choć wymóg ewaluacji w UE jest powszechny, to badania jej rzeczywistych efektów są sporadyczne. Książka stanowi jedną z pierwszych tego typu analiz zorientowanych praktycznie, lecz opartych na szerokiej bazie teoretycznej (czerpiącej ze źródeł amerykańskich i europejskich).

Po trzecie, jest to jedna z pierwszych publikacji na temat wykorzystania ewaluacji uwzględniająca specyfikę INTERREG-u (programów ponadnarodowej współpracy terytorialnej). W założeniu ma praktyczne implikacje dla Komisji Europejskiej i instytucji zaangażowanych w programy współpracy terytorialnej. Jest również wkładem do teorii zarządzania programami w środowisku wielokulturowym.

Wreszcie, po czwarte, książka stanowi jeden z pierwszych polskich głosów w dyskusji nad rolą ewaluacji w lepszym zarządzaniu programami publicznymi, a także wkładem w tworzącą się praktykę ewaluacji w Polsce. Zawarty w niej szeroki przegląd doświadczeń światowych umożliwia ich transfer na grunt Polski, tak z zakresu ewaluacji, jak i zarządzania programami INTERREG. Z uwagi na fakt, że jak dotychczas polskie publikacje dotyczące ewaluacji programów publicznych były incydentalne, praca jest praktycznie w całości oparta na anglojęzycznej literaturze przedmiotu.

1.3. STRUKTURA KSIĄŻKI

Książka składa się z pięciu rozdziałów. Celem rozdziału drugiego jest silne osadzenie dalszej analizy w dotychczasowych koncepcjach ewaluacyjnych. Czytelnik znajdzie w nim podsumowanie bogatej, anglojęzycznej literatury przedmiotu i wprowadzenie do specyfiki systemu ewaluacji programów Unii Europejskiej. W rozdziale omówiono różne definicje ewaluacji, jej teoretyczne korzenie, typy, cele i funkcje. Szczególna uwaga została poświęcona koncepcjom wykorzystania ewaluacji (ang. *evaluation utilization*).

Rozdział trzeci umieszcza ewaluację w szerszym kontekście zarządzania sektorem publicznym, odwołując się i krótko charakteryzując w chronologicznej kolejności reformy i paradygmaty zarządzania publicznego. Ostatnia część rozdziału jest poświęcona przybliżeniu genezy i obecnej praktyki ewaluacji w krajach Unii Europejskiej (w administracjach krajowych i w instytucjach UE). Czytelnicy zainteresowani kwestiami krajowego rynku znajdą tu omówienie polskiej specyfiki na tle europejskich trendów.

Rozdział czwarty wprowadza Czytelników do części empirycznej przeprowadzonych badań wykorzystania ewaluacji. Ich przedmiotem były programy INTERREG. Stąd rozdział otwiera charakterystyka programów współpracy terytorialnej i ich miejsce w unijnej polityce regionalnej. Rdzeniem tego rozdzia-

łu jest część druga – prezentacja syntetycznego metamodelu wykorzystania ewaluacji stworzonego w oparciu o przegląd literatury przedmiotu. Jego praktyczne zastosowanie i analiza wykorzystania ewaluacji w trzech studiach przypadków (w sumie sześć badań ewaluacyjnych) zostało przedstawione w ostatniej części rozdziału.

Ostatni, piąty rozdział zamykający książkę prezentuje wnioski z analizy czynników i efektów wykorzystania ewaluacji programów publicznych, szersze obserwacje odnoszące studia przypadków do paradygmatów zarządzania sektorem publicznym, a także rekomendacje praktyczne kierowane do zlecających i wykonawców ewaluacji. Należy podkreślić, że szczegółowe wnioski i główne rekomendacje dotyczą mechanizmów wykorzystania ewaluacji średniookresowych i bieżących – bo one właśnie były przedmiotem badania. Pokazano je jednak również w odniesieniu do innych typów czasowych ewaluacji, tak by Czytelnik otrzymał całościowy obraz procesu uczenia się z badań ewaluacyjnych.

Cała publikacja opatrzona jest szerokimi odniesieniami do literatury, które mają ułatwić zainteresowanym Czytelnikom poszukiwania dalszych pozycji książkowych dotyczących tematyki ewaluacji. Dodatkowo zaprezentowano w aneksie szczegółową metodologię badania będącego podstawą niniejszej książki (a więc postawione hipotezy badawcze, uzasadnienie podejścia badawczego, narzędzia analizy).

Oddając do rąk Szanownych Państwa niniejszą publikację pozostaje mi mieć nadzieję, że książka okaże się użyteczną inspiracją do dalszych przemyśleń, jak i przydatnym źródłem praktycznym w polu ewaluacji programów publicznych. Czytelników zainteresowanych podjęciem dyskusji zachęcam do kontaktu e-mailowego: k.olejniczak@uw.edu.pl.

2. TEORETYCZNE PODSTAWY WYKORZYSTANIA EWALUACJI[1]

2.1. WSTĘP

Celem niniejszego rozdziału jest stworzenie, w oparciu o bogatą, międzynarodową literaturę przedmiotu, spójnego systemu definicji i typologii, które posłużą w rozdziale czwartym do zbudowania modelu analitycznego.

W pierwszym podrozdziale omawiane są zróżnicowane definicje ewaluacji, jej teoretyczne korzenie, typy, cele i funkcje. Natomiast drugi podrozdział w całości poświęcono przeglądowi koncepcji wykorzystania ewaluacji (*evaluation utilization*) i determinantom użycia badań ewaluacyjnych. Dyskusję osadzono w szerszym nurcie koncepcji uczenia się instytucji i jednostek.

2.2. EWALUACJA – DEFINICJE, TYPY, FUNKCJE

2.2.1. Definicje ewaluacji i jej korzenie teoretyczne

Semantycznie ewaluacja to „oszacowanie wartości" (Oxford University Press, 2002). W takim przypadku termin „ewaluacja" zasadniczo odpowiadałby polskiemu słowu „ocena". Jednak definicje ewaluacji programów pochodzące z literatury przedmiotu, a także praktyka, podkreślają, że nie jest ona tożsama z oceną[2].

W literaturze znaleźć można szereg definicji ewaluacji, podkreślających różne aspekty samego jej procesu, uwypuklających jej zróżnicowane cele czy też odmienne kryteria samego procesu oceny.

---

[1] Przegląd światowych koncepcji ewaluacji był możliwy tylko i wyłącznie dzięki stypendium badawczemu Indiana University – Bloomington i Uniwersytetu Warszawskiego. Chciałbym podziękować pracownikom IU i UW za stworzenie możliwości skorzystania z przepastnych zasobów biblioteki IU. Szczególne podziękowania kieruję do Christy Borders, Andrew Hinnanta oraz „załogi Bloomi" w osobach prof. Justyny Beinek, Tomka Bilczewskiego, Lidiji Cvik, prof. Hanny Gosk i *last but not least* Adriana Zdrady, za wspaniałą atmosferę podczas mego pobytu na obczyźnie oraz za to, że nigdy nie rozmawialiśmy o ewaluacji!

[2] W polskich dokumentach na określenie ewaluacji stosowany jest najczęściej, choć nie zawsze konsekwentnie, termin „ocena". Użycie anglojęzycznego terminu jest bardziej wskazane z kilku powodów. „Ewaluacja" jest procesem. „Ocena" stanowi tylko jeden z elementów tego procesu. Według niektórych ujęć osąd nie zawsze jest kluczowy w procesie ewaluacji, a może być nawet pominięty (akcent jest wtedy postawiony na interakcje i animacje negocjacji pomiędzy uczestnikami programu). Użycie terminu „ocena" koncentruje uwagę wyłącznie na jednym elemencie, przypisując mu cechy całego procesu, a także wyłącza cały szereg alternatywnych podejść. Wreszcie, po trzecie, „ewaluacja" to nie tylko pojedyncze działanie, ale dziedzina wiedzy. Jak zauważa Kierzkowski (2002b: 10), ze względów semantycznych trudno byłoby więc określać terminem „ocena" zespół koncepcji teoretycznych.

Najszerzej rzecz ujmując, ewaluacja to systematyczne badanie społeczno-ekonomiczne oceniające i informujące o jakości i wartości programów publicznych (Forunier, 2005; Patton, 2004; Rossi et al., 2004; Weiss, 1998a)[3].

W literaturze przedmiotu podkreślane są szczególnie trzy elementy. Po pierwsze, uwypuklany jest analityczny i systematyczny (systemowy) charakter procesu (ang. *systematic inquiry*). Oznacza to stosowanie podejścia naukowego popartego danymi i przestrzeganie rygoru badawczego wynikającego z kanonu badań nauk społecznych (niezależnie od jakościowej lub ilościowej orientacji badania czy łączenia metod badawczych) (Langbein, Felbinger, 2006: 33; Patton, 2004; Rossi et al., 2004; Scriven, 1991: 131)[4]. Po drugie, zaznacza się, iż ewaluacja bada jakość danego programu (ang. *merit – quality*) i jego wartość w kategoriach ekonomicznych (ang. *worth – value*)[5] (Davidson, 2005: 1–2; Paddock, 1998). Wreszcie trzecim elementem podkreślanym w definicjach ewaluacji jest to, iż badanie ewaluacyjne obejmuje ocenę zarówno procesów – działania programu, jak i jego efektów (Bingham, Felbinger, 2002: 3; Weiss, 1998a: 4).

Tym, co odróżnia ewaluację od innych narzędzi badania i dyscyplinowania sektora publicznego (takich jak audyt, monitoring, badania *stricte* naukowe) są trzy kwestie. Po pierwsze, ewaluacja łączy w sobie element empiryczny (bada program i zjawiska pojawiające się jako konsekwencje programów) z charakterem normatywnym (ocenia strukturę działań, ich logikę, ich wartość) (Forunier, 2005: 139). Po drugie, rdzeniem ewaluacji jest jej utylitaryzm – ma wspomagać konkretnych odbiorców w konkretnej sytuacji (Stufflebeam, 2001a: 11), zaspokajać ich potrzeby informacyjne, wywoływać pozytywne zmiany i poprawiać jakość tak bieżącej, jak i przyszłych, analogicznych interwencji (Weiss, 1998a: 5). Po trzecie, „ewaluacja" jest procesem uczenia się, w którym kluczowe są interakcje i negocjacje (Greene, 2003: 592). Szczegółowe różnice podsumowano w poniższej tabeli.

---

[3] W niniejszej pracy dyskusja na temat szczegółowych celów, kryteriów, metod ewaluacji została wyłączona z głównej definicji i jest analizowana w dalszych częściach tego rozdziału. Dzięki temu operacyjna, wypracowana w tym miejscu definicja wyjściowa ewaluacji jest krótsza, bardziej elastyczna i możliwa do stosowania w różnych kontekstach – nie tylko badanych tu ewaluacji bieżących.

[4] Centralnymi pojęciami są tu: trafność wewnętrzna – na ile relacja między badanymi czynnikami jest przyczynowo-skutkowa, a na ile może być przypisana czynnikom zewnętrznym; trafność zewnętrzna – do jakiego stopnia obserwacje danej relacji mogą być generalizowane i wykorzystane w innych programach, populacje, lokacje; trafność teoretyczna – na ile badanie mierzy to, co zakłada (czy mierzone koncepty zostały właściwie zdefiniowane i zoperacjonalizowane), rzetelność – na ile użyte narzędzia gwarantują powtarzalność wyników i obserwacji (Shadish et al. 2001; Mathison 2004). Terminy te są szerzej dyskutowane w dalszej części książki.

[5] W niektórych przypadkach terminy te są stosowane rozłącznie, w innych natomiast wymiennie.

Tabela 1. Różnice między ewaluacją, badaniem naukowym, monitoringiem i audytem

Rodzaj cechy	Ewaluacja	Badanie naukowe	Monitoring	Audyt
Cel badania	Oszacowanie rzeczywistych efektów programu Ocena zasadności programu w danej formie	Poszerzenie wiedzy teoretycznej Identyfikacja i weryfikacja teorii wyjaśniających działanie progamu	Zweryfikowanie, czy program postępuje zgodnie z planem	Sprawdzenie, czy działania wykonywane są zgodnie z procedurami
Zakres badania	Szeroki i/lub wąski Szersze spojrzenie na program oraz sposób jego prowadzenia. Zidentyfikowanie i ocena: – celów danego przedsięwzięcia (w tym założonych relacji przyczynowo-skutkowych) – sposobów realizacji (ewentualnych słabości i pól potencjalnych usprawnień) – długookresowych zamierzonych i niezamierzonych efektów, wybiegających poza czas i miejsce wdrażania programu	Szeroki Szerokie spojrzenie na program jako element większego systemu	Wąski Rejestrowanie postępów z prowadzonych operacji Bieżące weryfikowanie tempa i kierunku, w którym zmierza program	Wąski Poprawność procedur Poprawność rozliczeń Zgodność z zapisami prawnymi i zasadami kontraktu
Metody badawcze	Zwykle multidyscyplinarne (ekonomia, socjologia itp.) Nie zawsze oparte na ścisłej metodologii naukowej	Zakorzenione zwykle w jednej dyscyplinie Ściśle przestrzegające metodologii badań naukowych	Oparte przede wszystkim na procedurach administracyjnych – zapisanych w dokumentach regulujących funkcjonowanie programu	Prawnie określone zasady, procedury i metody prowadzenia kontroli

Rodzaj cechy	Ewaluacja	Badanie naukowe	Monitoring	Audyt
Sposób realizacji	Zebranie danych – negocjacje – analiza, osąd (*judgement*) i wnioski – rekomendacje Pozyskiwanie dodatkowych danych na temat procesów, efektów programu i jego kontekstu Negocjacje z różnymi interesariuszami – identyfikowanie i uwzględnienie ich pozycji, interesów, oczekiwań Krytyczna analiza istniejących i zebranych danych – oszacowanie efektów programu Ocena efektywności struktur, procesu i skutków programu Rekomendacje zmian	Zebranie danych – analiza – opis i wnioski Postawienie hipotez Pozyskiwanie danych – głównie na temat szerokiego kontekstu działań, danych z dotychczasowych, zbliżonych tematycznie badań Szeroka analiza (często porównawcza) Opis i wnioski weryfikujące hipotezy badawcze – w zależności od podejścia budujące (indukcja) lub potwierdzające teorię (dedukcja)	Zbieranie danych – weryfikacja Systematyczne zbieranie ustalonego na początku programu zestawu danych (dotyczących postępów programu) Zestawianie istniejących danych dotyczących postępów inwestycji z uprzednio przyjętym planem	Kontrola – ocena Sprawdzanie faktur, rozliczeń finansowych, proceduralnej strony wykonanych operacji
Moment badania	Punktowy – na początku, w trakcie i po zakończeniu programu (ewaluacje *ex ante, średniookresowe, ex post*) Ciągły – w trakcie trwania programu (ewaluacja *on-going*)	Punktowy – dowolny, zwykle jednak po zakończeniu programu	Ciągły – w czasie trwania programu	Punktowy – okresowo w czasie trwania działań, główne badanie w momencie zamykania programu
Podmioty realizujące badanie	Eksperci zewnętrzni (konsultanci lub naukowcy) Niezależny, wewnętrzny departament ds. ewaluacji (w ramach danej instytucji zarządzającej programem)	Naukowcy	Instytucja zarządzająca programem (przy współpracy z głównymi interesariuszami)	Zewnętrzne – wyspecjalizowana firma posiadająca certyfikat audytora Wewnętrzne (w ramach danej instytucji zarządzającej programem) niezależna komórka audytorska
Efekty – konsekwencje badania	Działanie Przedstawienie rekomendacji dotyczących usprawnienia procesów zarządzania, wzmocnienia skuteczności i oddziaływania programu, ewentualnej zmiany orientacji programu (obecnego lub przyszłego), jego zaniechania lub kontynuacji	Opis Przedstawienie opisu, wniosków, które podnoszą stan wiedzy na temat danych procesów w danej dziedzinie Przedstawienie teorii, szerszych praw, zasad, modeli, które wyjaśniają praktykę	Działanie Podjęcie decyzji o kontynuacji programu w jego bieżącym kształcie lub ewentualnych działań administracyjnych niwelujących odchylenia od planu (np. przesunięcia środków pomiędzy poszczególnymi celami)	Działanie W przypadku wykrycia uchybień konsekwencje prawne, postępowanie administracyjne, zablokowanie ostatniej transzy lub zwrot całości wydanych środków

Źródło: (Olejniczak, 2007: 363–365).

Oficjalne rozporządzenia Unii Europejskiej z zakresu polityki budżetowej i polityki regionalnej nie zawierają definicji ewaluacji *sensu stricte*, opisują ją natomiast określając jej cele i dopuszczalne formy (porównaj: Council of the European Economic Community, 1988: art. 6; 1993: art. 26; Council of the European Union, 1999: art. 40–43; 2002; 2006 art. 47–49). Podobnie w pierwszych przewodnikach Komisji Europejskiej unikano jednej definicji, wymieniając natomiast cechy, które ewaluacja powinna spełniać – analityczny i systematyczny charakter badania (ang. a*nalytical & systematic approach*), solidność badania (ang. *reliability*) i użyteczność – jego charakter praktyczny (ang. *utility – issue oriented approach*) (European Commission, 1997)[6].

W nowszych dokumentach Unii Europejskiej ewaluacja jest definiowana nieco bardziej technicznie – jako „niezależne, obiektywne badanie wartości i jakości interwencji publicznej w odniesieniu do konkretnych kryteriów i standardów" (European Commission, 1999b: vol. 6), bądź też jako ocena działań publicznych w zależności od ich rezultatów, oddziaływania i potrzeb, które mają zaspokoić (European Commission, 2004a). Tym samym akcent jest więc położony na kryteria – punkt odniesienia przyjęty do oceny wartości i jakości programów poddanych ewaluacji (ta perspektywa jest dyskutowana szerzej w rozdziale 3). Najnowsze rozporządzenie UE (Council of the European Union 2006: art. 47.1) definiuje ewaluację dużo bardziej szczegółowo – przez pryzmat celów, jakie ma spełnić. Są to: poprawa jakości, skuteczności i – co ważne w przypadku rozległych, multisektorowych interwencji UE – spójności strategii i programów. Mocno wyeksponowany został także kontekst trwałego rozwoju i jego aspekty środowiskowe.

Wielość definicji wynika ze zróżnicowanych, potrójnych korzeni teoretycznych będących podstawą ewaluacji (Bachtler, Wren, 2006: 146; Tavistock Institute et al., 2003: 19). Pierwszą inspiracją jest pozytywizm – założenie o możliwości odkrycia wiedzy obiektywnej poprzez systematyczną, ustrukturalizowaną i powtarzalną obserwację prowadzoną według ścisłych, naukowych standardów. Drugą tradycją (zasadniczo stojącą w opozycji do pozytywizmu) jest konstrukcjonizm (nazywany również konstruktywizmem, ang. *constructivism*), według którego nie ma jednej obiektywnej rzeczywistości społecznej, lecz kilka współistniejących, zależnych od perspektywy obserwatora odczuć i interpretacji. Tworzą je aktorzy oraz instytucje biorące udział w danym procesie, operujące w określonym kontekście. W tej interpretacji rzeczywistość społeczna jest wytworem jednostek, efektem działalności aktorów zaangażowanych w dany proces, a nie czymś z góry danym, „zewnętrznym". Kluczem do zrozumienia sytuacji jest więc uchwycenie i zinterpretowanie kontekstu i procesów tak, jak są one postrzegane z różnych perspektyw przez różnych aktorów (Greene, 2003: 597). Wreszcie trzecią tradycją jest realizm (ang. *realism*), mający założenia zbliżone do konstruktywizmu, ale silniej podkreślający badanie mechanizmów – interakcji między uczestnikami programu a osiąganymi efektami. Te dwa ostatnie podejścia stoją w silnej sprzeczności z pozytywistyczny-

[6] Zmieniały się też terminy. W pierwszym Rozporządzeniu dotyczącym polityki regionalnej ewaluacja była określana terminem *assessment*, w latach 1994–1999 funkcjonował *appraisal and evaluation* (to pierwsze jako określenie ewaluacji *ex ante*), a dopiero w dwóch ostatnich okresach programowania zaczęto stosować jednolity termin *evaluation*.

mi postulatami poszukiwania obiektywnej, zewnętrznej wiedzy (Pawson, Tilley, 1997). Ponadto ewaluacja czerpie tak teorie, jak i narzędzia z co najmniej czterech różnych pól: metodologii badań społecznych, teorii ekonomicznych i wyboru publicznego, teorii zarządzania i organizacji, a także nauk politycznych i administracji (Tavistock Institute et al., 2003: 14–15).

Połączone, a jednocześnie wewnętrznie sprzeczne korzenie teoretyczne mają oczywiście silne implikacje dla zróżnicowania metod i podejść do ewaluacji, typów ewaluacji i przypisywanych jej funkcji. Przesunięcie akcentu na jedną z tradycji w oczywisty sposób warunkuje podejście badawcze i metodologię. Kwestie te będą szczegółowo omawiane w rozdziale na temat podejść badawczych do ewaluacji.

Podsumowując powyższy przegląd literatury, w niniejszej pracy stosowana będzie następująca definicja ewaluacji (patrz poniższa tabela).

Tabela 2. Definicja ewaluacji użyta w pracy

Definicja ewaluacji stosowana w niniejszej pracy
Ewaluacja to użycie zróżnicowanych metod badań społeczno-ekonomicznych do systematycznego zbierania danych, analizy, interpretacji, oceny i informowania o jakości i wartości programów publicznych.

Powyższa definicja jest szeroka i nie zawiera informacji na temat celów ewaluacji czy samych kryteriów analizy i oceny. Takie uproszczenie daje jednak jasną, spójną definicję wyjściową, nie wykluczając żadnej z trzech tradycji teoretycznych, na których zbudowana jest ewaluacja. Na tej podstawie w dalszych częściach publikacji budowane będą i uszczegółowiane funkcjonalne i operacyjne aspekty ewaluacji – wspomniane już kwestie celów i typów badania, kryteriów oceny.

2.2.2. Funkcje ewaluacji w sektorze publicznym

Procesy modernizacji społecznej, wyższa świadomość obywatelska, krytyczna niezależność współczesnej opinii publicznej, a wreszcie ograniczone zasoby sektora publicznego sprawiają, że współczesne społeczeństwa kładą nacisk na naukowe dowody efektywności, wydajności i jakości działań prowadzonych we wszystkich dziedzinach życia społecznego (Kardorff, 2004: 138). Co więcej, we współczesnych społeczeństwach funkcjonuje przeświadczenie, że nauka może i powinna pomagać w rozwiązywaniu problemów społecznych – Patton (2004: 337) nazywa to za E. Schumanem „the experimenting society". W sektorze prywatnym skuteczność działań jest w prosty sposób weryfikowana przez mechanizm rynkowy. Działania sektora publicznego, choć często prowadzone w warunkach rynkowych, nie mogą być oceniane wyłącznie z perspektywy osiągniętego zysku (Stewart, Ranson, 1994). Kryteria osądu wartości programów publicznych są zróżnicowane i zależne od wielu czynników, w tym dominujących przekonań czy aktualnych potrzeb społeczności. Ewaluacja, bazując na systematycznym procesie badawczym, dostarcza kryteriów, metod i środków do oceny racjonalności działań publicznych. Spełnia w sek-

torze publicznym rolę analogiczną do mechanizmu rynkowego w sektorze prywatnym (Shadish et al., 1991: 19).

Najogólniej rzecz ujmując, celem ewaluacji jest więc przyczynianie się do dobrobytu społecznego poprzez wspomaganie instytucji demokratycznych w doborze, nadzorowaniu, wartościowaniu i doskonaleniu programów i polityk publicznych (Mark et al., 2000: 3). Ewaluacja jest prowadzona w celu podniesienia jakości i skuteczności (efektywności) badanego programu, jak i przyszłych, analogicznych działań (Chelimsky, 1985: 7; Council of the European Union, 1999: art. 40; Cronbach, Associates, 1981: 14; Rist, 1990: 5).

Ta szeroko pojęta rola może być realizowana przez szereg funkcji zwykle określanych mianem celów ewaluacji. W literaturze przedmiotu przedstawiane są różne kategoryzacje.

Chelimsky (1997) wyróżnia trzy perspektywy funkcjonalne ewaluacji programów publicznych. Pierwsza to perspektywa odpowiedzialności publicznej (ang. *accountability perspective*) koncentrująca się na mierzeniu rezultatów działań, ocenie wartości i kosztów, a co za tym idzie wydajności programów. Druga to perspektywa gromadzenia wiedzy (ang. *knowledge perspective*) – zrozumienie natury problemów i procesów społecznych oraz tworzenie nowych i weryfikowanie starych modeli i metod radzenia sobie z nimi. Wreszcie perspektywa trzecia – perspektywa rozwoju (ang. *development perspective*) to wzmacnianie instytucji, budowanie zdolności instytucjonalnych. Analogiczną optykę przyjmuje Górniak (Górniak, 2005), pisząc o funkcji rozliczenia i egzekwowania odpowiedzialności, funkcji poznawczej (gromadzenia wiedzy) i funkcji rozwojowej.

Mark (Mark et al., 2000: 13) podaje cztery ogólne funkcje ewaluacji, w zasadzie pokrywające się z kategoryzacją Chelimsky. Oprócz funkcji odpowiedzialności (przedstawianej jednak jako określanie jakości i wartości programów), funkcji budowania wiedzy i wzmacniania instytucji, wyodrębnia funkcję kontroli i weryfikacji zgodności z regułami (na ile działania są zgodne z formalnymi regulacjami). Gary T. Henry (2001) w ciekawy sposób wiąże te funkcje z szerszym kontekstem – z wymogami płynącymi ze współczesnych systemów demokratycznych i z wymogiem przejrzystości działań publicznych, informowania społeczeństwa i sprawności instytucji publicznych (w swej analizie odwołuje się do Konstytucji Stanów Zjednoczonych).

Niektórzy autorzy (Greene, 2006; Ryan, Destefano, 2000) wskazują również na szerszą funkcję ewaluacji – wzmacnianie demokracji – która wiąże się z i nformowaniem społeczeństwa, rozwijaniem dialogu społecznego i włączaniem w proces decyzyjny.

W europejskich źródłach odnoszących się do programów społeczno-gospodarczych podawanych jest pięć funkcji ewaluacji (Tavistock Institute et al., 2003). Są to:

- poprawianie planowania (ang. *planning & determining efficiency*) – zagwarantowanie uzasadnienia dla polityki/programu oraz wydajnej i racjonalnej alokacji zasobów;
- wzmocnienie odpowiedzialności (ang. *improving accountability*) – informowanie społeczeństwa, na ile cele programu zostały osiągnięte i jak efektywnie użyto przypisanych zasobów publicznych;

- poprawa wdrażania (ang. *improving management & implementation process*) – podnoszenie jakości struktur zarządzania programem oraz samego procesu implementacji działań;
- poszerzanie wiedzy i wspieranie procesów uczenia się (ang. *knowledge production*) – wyjaśnianie czynników i procesów warunkujących sukces programu lub jego porażkę, wskazywanie w jaki sposób można podnieść efektywność przyszłych interwencji, promowanie dobrych wzorców;
- wzmacnianie instytucji (ang. *institutional strengthening*) – podnoszenie umiejętności i zasobów uczestników programu, w tym wzmacnianie powiązań i relacji między nimi.

Bachtler (2001: 42–48), analizując funkcje ewaluacji w polu europejskiej polityki regionalnej, zwraca uwagę na dodatkowy aspekt poszerzania wiedzy. Podkreśla, iż ewaluacja powinna weryfikować teorie i założenia leżące u podstaw danych polityk (ang. *testing assumptions*).

Tabela 3. Przeciwdziałanie dysfunkcjom zarządzania opartego na wynikach

Nazwa i opis dysfunkcji	Przeciwdziałanie ewaluacji
Nacisk na mierzenie celów i pomiar sukcesu zamiast na rzeczywiste efekty (ang. *Measure fixation*).	Analiza rzeczywistych efektów: planowanych i ubocznych (pozytywnych i negatywnych).
Zabijanie innowacji poprzez sztywny system wskaźników i pomiarów (ang. *Ossification*).	Szersze spojrzenie, użycie różnych metod. Krytyczna analiza samego systemu wskaźników (szerzej monitoringu) – ich zasadności, praktycznego zastosowania.
Koncentrowanie uwagi na kwantyfikowalnych efektach ilościowych i ignorowanie trudno mierzalnych, jakościowych efektów programu (ang. *Tunnel Vision*).	Badanie efektów jakościowych, miękkich, relacji pomiędzy interesariuszami (wywiady, ankiety, obserwacja uczestnicząca).
Koncentrowanie się na wąskich technicznych celach lub celach krótkookresowych kosztem celów strategicznych (ang. *Suboptimization & myopia*).	Weryfikowanie, na ile projekt (pozytywnie zakończony od strony formalnej) przyczynił się do szerszej strategii, wpłynął na cele polityki.
Błędne interpretowanie danych kontekstowych i w rezultacie budowanie błędnych założeń programu i ocen (ang. *Misinterpretation*).	Poprawne użycie metod badawczych, spojrzenie z kilku perspektyw, używanie kilku rodzajów narzędzi do oceny jednej kwestii (triangulacja)[7].
Manipulowanie danymi tak, by pokazać na papierze efekty programu różne od rzeczywistości (ang. *Misrepresentation*).	Weryfikowanie efektów na podstawie różnych źródeł. Utrzymanie dystansu do zlecającego – niezależność i obiektywizm sądów, przestrzeganie standardów zawodowych.

Źródło: Opracowanie własne na podstawie (Leeuw, 2002; Smith, 1995).

Warto też zwrócić uwagę, że ogólna funkcja ewaluacji podawana jako „poprawa zarządzania" mieści w sobie minimalizowanie negatywnych efektów ubocznych zarządzania opartego na wynikach (ang. *performance management*). Jak zostanie to szerzej omówione w części poświęconej praktyce ewa-

[7] Triangulacja (ang. *triangulation*) oznacza „więcej niż jeden sposób zbierania danych w ramach jednolitego planu badawczego, w celu przetestowania tej samej hipotezy" (Frankfort–Nachmias, Nachmias, 2001: 611). W ewaluacjach oznacza zarówno zróżnicowanie metod, jak i źródeł danych.

luacji (rozdział 3), ewaluacja jest silnie złączona z nowym podejściem do zarządzania w sektorze publicznym – zarządzaniem opartym na wynikach (*nota bene* czerpie ono wzorce z zarządzania biznesowego). Podejście to, mimo iż dynamizuje i dyscyplinuje sektor publiczny, może prowadzić do pewnych zniekształceń. Umiejętnie stosowana ewaluacja może minimalizować ryzyko ich wystąpienia. W Tabeli 3 przedstawiono dysfunkcje mogące wystąpić przy zarządzaniu opartym na wynikach oraz sposoby, w jakie mogą być one neutralizowane przez ewaluację.

W odniesieniu do funduszy strukturalnych UE (w tym także Inicjatywy Wspólnotowej INTERREG), Batterbury (2006: 181–182) konkluduje, że obecne ewaluacje spełniają trzy funkcje: wzmacniają odpowiedzialność i legitymizują działania, poprawiają planowanie, podnoszą jakość i skuteczność wdrażania. Zaznacza jednak, że badania ewaluacyjne mają więcej funkcji i wymienia kolejne cztery, niewystępujące lub marginalizowane w funduszach strukturalnych. Są to: budowanie zdolności ewaluacyjnych (ang. *building capacity*), uczenie się (ang. *learning*), włączanie interesariuszy i tworzenie w nich poczucia współwłasności programu (ang. *developing a sense of ownership among programme participants*), uprawnianie interesariuszy – włączanie we współdecydowanie o programie tych grup, które były marginalizowane (ang. *empowerment*).

Podsumowując literaturę dotyczącą funkcji ewaluacji, na potrzeby niniejszej analizy można przedstawić następującą listę pięciu potencjalnych funkcji badań ewaluacyjnych w odniesieniu do programów publicznych współfinansowanych przez Unię Europejską.

Funkcja poprawy planowania (ang. *improving planning*) oznacza:

- zagwarantowanie racjonalnego uzasadnienia dla polityki/programu poprzez zweryfikowanie jego spójności zewnętrznej, przystawalności celów do potrzeb i problemów;
- zapewnienie wydajnej, racjonalnej i skutecznej alokacji zasobów poprzez zweryfikowanie logiki programu, racjonalności wyznaczonych celów, spójności zewnętrznej planowanej strategii.

Funkcja poprawy wdrażania i bieżącej kontroli jakości (ang. *improving performance & quality*) obejmuje:

- poprawienie istniejącej struktury, procedur i procesu zarządzania programem (ang. *improving programme management*);
- minimalizowanie negatywnych efektów ubocznych zarządzania opartego na wynikach;
- podnoszenie efektywności dostarczania programu poprzez weryfikowanie, na ile program realizuje zakładane cele, jaka jest jakość produktów (ang. *improving programme delivery*).

Funkcja wzmacniania partnerstwa i poczucia współwłasności (ang. *improving partnership & ownership*) obejmuje:

- wzmacnianie powiązań, relacji, współpracy, interakcji między uczestnikami programu;
- aktywizowanie interesariuszy – włączanie ich w działania programu, dyskusję nad jego oceną i kierunkami rozwoju;

- uprawnianie interesariuszy – włączanie we współdecydowanie o programie tych grup, które były marginalizowane (ang. *empowerment*).

Funkcja wzmacniania odpowiedzialności (ang. *improving accountability*) to:

- informowanie szerszej opinii publicznej o skuteczności programu – na ile cele programu zostały osiągnięte;
- informowanie szerszej opinii publicznej o wydajności podjętych działań – jak efektywnie użyto przypisanych zasobów publicznych.

Funkcja poznawcza – wspieranie procesów uczenia się oznacza:

- wyjaśnianie czynników i procesów warunkujących działania programu, jego sukces lub porażkę (w tym także czynników ubocznych);
- weryfikowanie teorii leżących u podstaw danych polityk, interwencji;
- wyciąganie wniosków dla innych programów i promowanie dobrych wzorców;
- budowanie zdolności ewaluacyjnych danych instytucji zaangażowanych w badanie lub szerzej – systemu w danym reigonie, kraju (ang. *building evaluation capacity*).

Pojedyncze badanie ewaluacyjne może spełniać kilka funkcji. Wydaje się jednak, że w praktyce niektóre z funkcji nie mogą być ze sobą łączone w ramach jednego badania. Potencjalna sprzeczność leży pomiędzy funkcjami zorientowanymi na bieżące zarządzanie i procesy a funkcjami nakierowanymi na ocenę efektów (czyli – jak określa to najnowsze prawodawstwo UE – charakterem operacyjnym lub strategicznym) (Council of the European Union, 2006: art. 47.2). Wzajemne wykluczanie się w ramach jednego badania wynikać może z kilku kwestii. Po pierwsze z ograniczeń czasu i zasobów – trudno w jednym badaniu objąć dokładną analizą tak odmienne tematy jak bieżące operacje programu i relacje między jego uczestnikami, a jego długoterminowe oddziaływanie. Po drugie, badanie zjawisk zarządzania i analizowanie oceny efektów wymagają odmiennych podejść i warsztatów badawczych. Po trzecie, badani i badający, koncentrując się na jednym z dwóch wyróżnionych biegunów, przyjmują odmienne role (pomagających i wspomaganych przy ewaluacji procesów, a oceniających i ocenianych przy ewaluacji efektów). Wydaje się, że jedynie badania zorientowane na funkcję planowania mogą godzić w sobie analizę procesów i efektów, gdyż do obydwu odnoszą się pobieżnie i *a priori* (dopiero projektują system wdrażania i przewidują efekty interwencji). Te potencjalne dylematy wyboru będą weryfikowane w trakcie analizy empirycznej studiów przypadków. Odniosę się do nich również we wnioskach i rekomendacjach.

2.2.3. Typologia ewaluacji

W literaturze przedmiotu funkcjonuje szereg typologii ewaluacji (Greene, 2003; Herman et al., 1987; Owen, Rogers, 1999; Tavistock Institute et al., 2003). To zróżnicowanie wynika po pierwsze z odmiennych inspiracji teoretycznych i podstaw metodycznych, po drugie zaś z przyjmowania różnych kryteriów dla typologii ewaluacji (ze względu na moment wykonania ewaluacji,

cel, użyte metody, etc.). Upraszczając, możemy wyróżnić pięć podstawowych wymiarów kategoryzowania ewaluacji. Są to:

- przedmiot ewaluacji,
- ogólny cel ewaluacji,
- moment podjęcia badania,
- administracyjna organizacja badania,
- zastosowane podejście badawcze – model badawczy (ang. *evaluation approach*).

Poniżej krótko omówione jest każde z nich. Należy jednak podkreślić, że te kategorie nie są rozłączne. W praktyce konkretnego badania ewaluacyjnego nakładają się na siebie.

Podział ze względu na przedmiot ewaluacji

Stosując kryterium przedmiotu badania, tzw. ewaluandy (ang. *evaluand*) można wyróżnić pięć typów odnoszących się do interwencji publicznych.

Ewaluacja projektu to badanie koncentrujące się na analizie pojedynczej interwencji, z reguły dotyczy nowego, innowacyjnego projektu. W praktyce UE takie ewaluacje są elementem większej całości (badania programu) i później wykorzystywane są jako studia przypadku lub też robione są na potrzeby działań pilotażowych i mają pomóc w podjęciu decyzji, czy kontynuować interwencję w większej skali.

Ewaluacja programu to najczęściej spotykany typ ewaluacji – jej przedmiotem może być zarówno pojedynczy program sektora publicznego, jak i cały pakiet unijnych interwencji – na przykład działanie, priorytet.

Ewaluacja polityki, częściej nazywana analizą polityk (ang. *policy analysis*) składa się z szeregu badań i analiz (zwykle na przestrzeni wielu lat). W administracjach państw członkowskich często praktykowana, na poziomie UE jest kwestią delikatną politycznie (np. próby analizy wspólnej polityki rolnej). Niemniej niektóre polityki UE doczekały się szczegółowych, regularnych i krytycznych ewaluacji (np. polityka spójności – raporty kohezyjne).

Ewaluacja tematyczna/horyzontalna (ang. *thematic/horizontal evaluation*), której przedmiotem jest konkretna kwestia dotycząca zarządzania lub aspekt programu (np. zastosowanie zasady partnerstwa w zarządzaniu funduszami strukturalnymi, realizacja celów horyzontalnych, wsparcie innowacyjności MŚP). Analiza danego aspektu jest prowadzona przekrojowo pomiędzy kilkunastoma programami.

Metaewaluacja (ang. *meta-evaluation*) to badanie, którego przedmiotem są inne, wcześniejsze ewaluacje. Jest to analiza porównawcza, podsumowująca obserwacje, wnioski i rekomendacje kilkunastu ewaluacji (z reguły dotyczących analogicznych programów) oraz oceniająca zastosowane techniki i poprawność metodologii (a więc jakość poszczególnych ewaluacji).

Niektórzy polscy autorzy (Zawicki, 2007) wyróżniają prócz ewaluacji programów publicznych również ewaluację skutków regulacji, czyli ocenę jakości prawa (prowadzoną z reguły *ex ante*). Taka kategoryzacja znacznie poszerza pojęcie badań ewaluacyjnych. Nie ma ona jednak szerszego zastosowania

w literaturze przedmiotu i praktyce analizowanego tutaj pola polityki regionalnej. W niniejszej publikacji koncentruję się więc na dominującym nurcie rozumienia przedmiotu ewaluacji – interwencjach publicznych, przede wszystkim programach.

Podział ze względu na ogólny cel ewaluacji

Zasadniczo w literaturze wyróżniane są dwa ogólne rodzaje ewaluacji (Scriven, 1991: 168–9, 340).

Pierwszy to **ewaluacja formatywna** (ang. *formative evaluation*). Jej celem jest przede wszystkim ocena formy, planu oraz procesu wdrażania danej interwencji przed jej rozpoczęciem lub w czasie jej trwania. Koncentruje się na doradzeniu osobom zarządzającym programem, jakie cele ustalić, lub jak poprawić płynność wdrażania programu i podnieść jego ostateczną jakość. Często jest prowadzona przez samą instytucję wdrażającą interwencję (ang. *in-house evaluation*).

Drugi to **ewaluacja podsumowująca** (ang. *summative evaluation*), której celem jest podsumowanie zakończonej interwencji – jej efektów i procesu oraz pomoc w decyzji „co dalej?" (kontynuować czy zatrzymać program, zmienić jego formę). Taka całościowa ocena jest podejmowana po zakończeniu programu. Ma charakter opisowy i wyjaśniający i jest adresowana do szerszej publiki, do społeczeństwa, a nie wyłącznie do instytucji realizujących dany program[8].

Zasadniczo temu dualnemu podziałowi odpowiada rozróżnienie zastosowane w nowych dokumentach polityki regionalnej Unii Europejskiej (Council of the European Union, 2006: art. 47.2). Wyróżniane są **ewaluacje o charakterze operacyjnym** – wspomagające bieżące zarządzanie programem (a więc odpowiednik ewaluacji formatywnej) oraz **ewaluacje o charakterze strategicznym** – badające ewolucję programu i jego wartość w świetle szerszych celów polityki, a także w odniesieniu do innych programów (tym samym odpowiadające ewaluacji podsumowującej).

Tak zarysowany podział współgra z zaprezentowanymi w poprzednim podrozdziale dylematami wyboru funkcji ewaluacji. Ewaluacje formatywne, operacyjne zorientowane są na bieżące zarządzanie i procesy, podczas gdy ewaluacje podsumowujące (o charakterze strategicznym) – na ocenę efektów.

Podział ze względu na moment podjęcia badania

Zasadniczo wyróżnia się trzy rodzaje ewaluacji: ewaluacje wykonywane przed rozpoczęciem badanej interwencji, w jej trakcie oraz badania po zakończeniu programu będącego przedmiotem analizy i oceny (Bingham, Felbinger, 2002; Forunier, 2005; Scriven, 1991).

---

[8] W niektórych źródłach obydwa terminy są stosowane zamiennie z „ewaluacją procesów" (ang. *process evaluation*) i „ewaluacją efektów" (ang. *outcome evaluation*) (Bingham R.D., Felbinger C.L. 2002: 4). Scriven (1991) dowodzi, że nie jest to poprawne, gdyż przy ewaluacji podsumowującej ogranicza podejście tylko do efektów, wyłączając analizę procesów wdrażania (a tym ewaluacja podsumowująca również się zajmuje, tyle że z perspektywy *ex post*). Jednak w praktyce funduszy strukturalnych akcenty rzeczywiście są położone na procesy w ewaluacji formatywnej i na efekty w ewaluacji podsumowującej.

W praktyce Unii Europejskiej wypracowany został nieco bardziej szczegółowy podział, ugruntowany zapisami prawnymi. I tak, w polu polityki regionalnej UE w latach 2000–2006 wyróżniono pięć rodzajów ewaluacji (Council of the European Union, 1999; European Commission, 1999b)[9].

Ewaluacja wstępna/szacunkowa (ang. *ex ante evaluation*) prowadzona przed rozpoczęciem interwencji. Jej celem jest optymalizacja alokacji zasobów poprzez zweryfikowanie struktury, logiki i uzasadnienia programu.

Ewaluacja średniookresowa (ang. *mid-term evaluation*). Jest przeprowadzana w połowie trwania danego programu. Kładzie nacisk na ocenę procesu zarządzania, aktualność założeń strategii, postępy w osiąganiu efektów.

Uaktualnienie ewaluacji średniookresowej (ang. *mid-term update evaluation*). To badanie jest dokładnym powtórzeniem ewaluacji średniookresowej, ale w późniejszym terminie – jest przeprowadzana w 3/4 trwania danego programu (rok, dwa po badaniu średniookresowym). Przyczyną tego powtórzenia jest z reguły niski stopień zaawansowania programu w połowie okresu jego wdrażania (szerzej o przyczynach tego zjawiska i realiach UE w rozdziale 3).

Ewaluacja *ad hoc/*bieżąca lub ciągła (ang. *on-going/ad hoc/interim evaluation*). Wykonywana jest w dowolnym momencie implementacji programu (poza momentami przypadającymi na ewaluację średniookresową i jej uaktualnienie). Może być badaniem punktowym (mówimy wtedy o ewaluacji bieżącej) albo składać się z całego szeregu wykonywanych po sobie badań. Ta druga opcja, choć rzadko spotykana, zbliża ewaluację do monitoringu i nazywa się ewaluacją ciągłą (co podkreśla jej permanentny charakter). Ma zapewnić wsparcie dla zarządzania programem – jest zlecana po to, by wyjaśnić i zaproponować poprawę bieżących problemów występujących w programie.

Ewaluacja *ex post/*końcowa/pełna/zamykająca (ang. *ex post evaluation*). Jest wykonywana po zakończeniu interwencji (najczęściej po dwóch latach od zakończenia wszystkich działań). Skupia się na ocenie rezultatów i długookresowych efektów (oddziaływania) programu, w tym trwałości uzyskanych pozytywnych zmian i zidentyfikowaniu ewentualnych efektów ubocznych (pozytywnych i negatywnych).

W nowym okresie programowania (lata 2007–2013) zdecydowano się na pewne uproszczenia, idące w kierunku wyznaczonym przez praktykę amerykańską. I tak, wymogiem prawnym w odniesieniu do funduszy strukturalnych pozostały tylko ewaluacje *ex ante* i *ex post*. Dodatkowo, UE dopuszcza ewaluacje bieżące, z tym że decyzję o ich podjęciu i zakresie badania pozostawia instytucjom wdrażającym programy w krajach członkowskich (Council of the European Union, 2006: art. 48–49). Zniknęły więc wymogi prawne prowadzenia ewaluacji średniookresowych i ich uaktualnień. Jednak ich zakres może być pokryty badaniami bieżącymi.

[9] W polskich dokumentach występuje kilka różnych tłumaczeń terminów angielskich. W tym miejscu pracy przytaczane są wszystkie. W dalszej części pracy używane są terminy łacińskie i te stosowane w Unii Europejskiej.

Podział ze względu na administracyjną stronę badania

Ewaluacja wewnętrzna, czyli prowadzona przez osoby z danej instytucji odpowiedzialnej za wdrażanie analizowanej interwencji (najczęściej przez odpowiedni departament lub komórkę ewaluacyjną funkcjonującą w ministerstwie kierującym danym programem).

Ewaluacja zewnętrzna, czyli badanie zlecane poza instytucję prowadzącą dany program, wykonywane przez zespoły niezależnych ekspertów wyłonione w ramach postępowania przetargowego.

Podział ze względu na zastosowane podejście badawcze

Dotychczas zaprezentowane typologie są w miarę jednolite, prezentowane przez różnych autorów tylko z nieznacznymi modyfikacjami. Natomiast w wypadku podejść badawczych istnieje wiele odmiennych prób ujęcia tematu, brak też w literaturze jednego dominującego ujęcia.

Modele ewaluacji są warunkowane przede wszystkim momentem podjęcia badania i jego celem (pomoc w bieżącym zarządzaniu vs. decyzje strategiczne). Stufflebaum (2001a) wyróżnia aż dwadzieścia dwa podejścia badawcze (biorąc pod uwagę jednocześnie kryteria szczegółowego celu badania adresatów i używane metody), Paddock (1998) identyfikuje dwanaście modeli, różnicując je przede wszystkim pod względem roli, jaką odgrywa w badaniu ewaluator i tego, w czym badanie ma pomóc (decyzja polityczna, bieżące zarządzanie, negocjacje między interesariuszami). Koncentrując się tylko na trzech elementach: celu ewaluacji, roli ewaluatora i metodach, szereg współistniejących podejść można pogrupować w sześć głównych typów: podejście eksperymentalne i *quasi*-ekspertymentalne, podejście wyłączające cele programu, podejście partycypatywne/uczestniczące, podejście uprawniające interesariuszy, podejście skoncentrowane na użyciu badania, podejście realistyczne.

Podejście eksperymentalne i *quasi*-ekspertymentalne (ang. *experimental & quasi-experimental design*) opiera się na paradygmacie pozytywistycznym i skupia się na badaniu relacji przyczynowo-skutkowej między zaplanowanymi działaniami danego programu a osiągniętymi zmianami w danej populacji na danym obszarze (Bingham, Felbinger, 2002; Campbell, Stanley, 1967; Shadish et al., 2001). Ewaluator jest naukowcem – „strażnikiem prawdy", kierującym się przy ocenie programu obiektywnymi kryteriami naukowymi. W badaniu wykorzystywane są metody ilościowe, eksperymentalne i analiza statystyczna. Jego rdzeniem jest badanie sytuacji sprzed i po interwencji w formie eksperymentu (jeśli zastosowany jest losowy wybór grupy) lub *quasi*-eksperymentu (gdy grupa kontrolna nie jest losowa)[10]. Zasadniczą prze-

[10] Przy określaniu relacji przyczynowo-skutkowej zastosowany może być jeden z kilku podtypów tego modelu – w zależności od dostępnych danych, charakteru grupy docelowej, możliwości kontrolowania zmiennych (Shadish et al., 2001). Po pierwsze, wyróżniane jest preeksperymentalne podejście (ang. *reflexive design*) w dwóch opcjach: porównanie danej grupy poddanej programowi przed i po programie (ang. *one group pre-test-posttest design*) lub tej samej grupy w kilku momentach przed programem z jednorazowym badaniem po programie (ang. *simple time-series design*). Po drugie, wyróżniane jest właściwe *quasi*-eksperymentalne podejście

wagą tego modelu nad pozostałymi jest możliwość jasnego zidentyfikowania – w oparciu o obiektywnie naukową analizę – relacji przyczynowo-skutkowej (lub jej braku) między działaniem programu a jego efektem. Natomiast słabą stroną są trudności w zastosowaniu przy dużych, zróżnicowanych programach realizowanych na dużym obszarze (problem stworzenia poprawnej grupy kontrolnej i kontrolowania wszystkich zmiennych), a także brak wyjaśnienia samego procesu, który spowodował ewentualne efekty.

Podejście wyłączające cele programu (ang. *goal-free evaluation model*) opiera się na założeniu, że prawdziwa wartość programu może być oceniona na podstawie rzeczywistych zmian w otoczeniu, a nie na podstawie tego, co program starał się zmienić (Scriven, 1976). Oznacza to, że cała organizacja badania pomija zapisane cele programu, zaś koncentruje się na otwartej obserwacji i identyfikacji rzeczywistych zmian wywołanych interwencją (pozytywnych i negatywnych)[11]. Obserwacje te są następnie odnoszone do potrzeb i problemów grupy docelowej (Alkin, 2004: 33 za Scriven 1974). Ewaluator wartościuje i ocenia, kierując się swoim doświadczeniem, krytyczną interpretacją tego, co ważne, a przede wszystkim kryteriami wypracowanymi na bazie analizy alternatywnych efektów programu i rzeczywistych potrzeb odbiorców (dobry program to taki, który spełnił potrzeby odbiorców). Autor tego podejścia – Scriven – podkreśla użycie metod ilościowych. Metody jakościowe są jednak również dopuszczalne. Zaletą tego podejścia jest to, iż analizuje efekty programu całościowo – włączając w to efekty uboczne – i wiąże je z oczekiwaniami grupy docelowej. Pomijanie zapisanych celów programu wywołuje jednak opór ze strony zarządzających programem.

Podejście partycypatywne/uczestniczące (ang. *participatory evaluation*) opierając się na filozofii konstrukcjonizmu i teoriach uczenia się organizacji, stawia dwa cele ewaluacji. Po pierwsze, zrozumienie programu poprzez uwzględnienie różnych perspektyw jego postrzegania i włączenie interesariuszy w proces ewaluacji (a szczególnie w moment oceny). Po drugie, poprawę programu poprzez wypracowanie konsensusu wśród interesariuszy i kumulo-

---

w dwóch opcjach: badania grupy docelowej poddanej interwencji z grupą porównawczą, niepoddaną interwencji (ang. *pretest-posttest comparison group*). Wreszcie „prawdziwie naukowe" podejście tzw. klasyczny plan eksperymentalny, gdzie pomiary początkowe i końcowe są wykonywane i porównywane między grupą poddaną interwencji (eksperymentalną) oraz specjalnie stworzoną na potrzeby badania grupą kontrolną, dla której wszystkie zmienne są kontrolowane przez badaczy (ang. *pretest-posttest control group*), a dobór próby odbywa się na zasadzie losowej. Ten ostatni model jest możliwy do prowadzenia wyłącznie w warunkach laboratoryjnych. Nie ma więc zastosowania do programów publicznych rozwoju regionalnego. Dodatkowo wyróżniany jest również plan czynnikowy (ang. *factorial design*) – podtyp badania eksperymentalnego, dający możliwość jednoczesnego badania wpływu kilku zmiennych zależnych na zmienną niezależną i wykrywanie relacji między nimi. W takim wypadku próbuje się kontrolować i mierzyć siłę oddziaływania poszczególnych czynników, stosując dodatkowe testy statystyczne (Bingham, Felbinger, 2002).

[11] W tym celu wykorzystywana jest tzw. metoda *Modus Operandi* – ciąg przyczynowy. Ewaluator tworzy pełną listę potencjalnych efektów programu (identyfikując alternatywne rozwiązania i sprzeczne efekty). Zawęża potencjalne wybory poprzez określenie, które z potencjalnych przyczyn występowały zanim pojawiły się efekty i który ciąg przyczynowy w całości pasuje do zaobserwowanych zmian i zdarzeń.

wanie wiedzy instytucjonalnej (Cousins, Whitmore, 1998; King, 2005). W założeniu siła ewaluacji (pozytywna zmiana, którą wywołuje) jest proporcjonalna do tego, w jakim stopniu są reprezentowane opinie różnych interesariuszy, w jakim stopniu interesariusze są włączani w badanie (określenie jego celów, ocenę efektów, identyfikację problemów), na ile rozumieją proces i utożsamiają się z nim (ang. *sense of ownership*).

W tym podejściu ewaluatorzy pełnią przede wszystkim rolę animatorów dyskusji, doradców wprowadzających nowe perspektywy i wiedzę, asystentów technicznych (gromadzących i podsumowujących dane i informacje do dyskusji, sugerujących narzędzia), koordynatorów procesu, a także negocjatorów pomiędzy włączonymi uczestnikami ewaluacji.

Definiowanie roli ewaluatora jest wśród propagatorów tego podejścia zróżnicowane. Niektórzy autorzy zaznaczają, że ewaluator włącza uczestników w proces formułowania kryteriów, jednak ostatecznej oceny dokonuje sam, niezależnie. Poszukuje konsensusu, ale nie stara się go budować, gdy ten nie istnieje (Stake, 2000: 104). Inni stoją na stanowisku, że ewaluator identyfikuje potrzeby informacyjne uczestników, zbiera informacje merytoryczne, a następnie jedynie ułatwia dialog i negocjacje pomiędzy interesariuszami dokonującymi ostatecznej oceny. Działa więc aktywnie nad budowaniem konsensusu i wiedzy (Alkin, 2004: 39 za MacDonalds 1979; Torres, Preskill, 2001).

Głównymi narzędziami badania są metody jakościowe i interaktywne: wywiady, panele dyskusyjne, warsztaty zarówno na etapie tworzenia pytań badawczych, kryteriów oceny programu, zbierania opinii, jak i samej oceny. Niektórzy autorzy silnie podkreślają również rolę studiów przypadku (Guba, Lincoln, 1989: 185). Siłą tego podejścia jest aktywizowanie szerokiej rzeszy odbiorców badania oraz proces uczenia się. Trudnym elementem jest natomiast zachowanie przez ewaluatorów obiektywizmu w badaniu, utrzymanie jego integralności (Stufflebeam, 1994), a także zapewnienie, by ewaluacja nie pomijała kwestii kontrowersyjnych czy niewygodnych dla głównych interesariuszy, ugruntowując jedynie *status quo* (Chelimsky, 1997: 17). Główną osią krytyki jest również to, że ewaluacja – miast być obiektywnym procesem oceny – staje się animowaną, wielogłosową dyskusją nad programem (w praktyce zdominowaną przez głównych decydentów).

Podejście uprawniające interesariuszy (ang. *empowerment evaluation*) – jego celem jest włączenie (tak w proces oceny, jak i zarządzania programem) tych interesariuszy, którzy dotychczas byli marginalizowani lub zupełnie wyłączeni, a po drugie nauczenie interesariuszy samodzielnego określania własnej wartości i wartości programu (Fetterman, 1996). Ewaluator spełnia tu kilka ról. Po pierwsze, reprezentuje głos tych interesariuszy, którzy są nieobecni. Po drugie, dostarcza decydentom informacji, które umożliwią sprawiedliwszą alokację zasobów. Po trzecie, jest animatorem dyskusji i wspomaga ocenę, pilnując by kryteria sprawiedliwości, uczciwości i etyki były przestrzegane. Wreszcie, po czwarte, uczy interesariuszy, tak by mogli w przyszłości sami prowadzić swoje ewaluacje (Fetterman, 1996; House, 2001). Podobnie jak w podejściu partycypatywnym, tak i tu dominują metody jakościowe.

Podejście skoncentrowane na użyciu badania (ang. *utilization-focused evaluation*) jest w zasadzie odmianą podejścia partycypatywnego[12]. Zostało jednak wyróżnione w tej pracy, gdyż szczególny akcent stawia na wykorzystanie ewaluacji. Cała filozofia badawcza opiera się na założeniu, że badanie ma pomóc konkretnej grupie adresatów w konkretny sposób. Ewaluatorzy mają za zadanie zmaksymalizować wykorzystanie badania przez konkretnych użytkowników, dostosowując wszystkie jego elementy do zakładanego użycia. Ewaluator nie ocenia samodzielnie, ale współpracuje i pomaga przyszłym użytkownikom ewaluacji w wypracowaniu kryteriów oceny i samej oceny (Alkin, 2004: 50). Wymaga to: zidentyfikowania grupy potencjalnych użytkowników, włączenia ich w proces ustalania celów i formy badania (w tym dyskusji nad metodami), aktywnej współpracy z nimi w procesie zbierania danych, ich interpretacji, oceny i wypracowywania rekomendacji, wspólnego stworzenia strategii rozpowszechniania wyników (Patton, 1997, 2005). Ewaluator działa analogicznie jak w podejściu partycypatywnym, ale szczególny akcent jest kładziony na elastyczność i dostosowywanie się do potrzeb użytkowników (Owen, Lambert, 1998; Stake, 2005)[13]. Część autorów definiuje użytkowników wąsko – jako zarządzających programem, inni widzą grupę interesariuszy dużo szerzej (jako opinię publiczną) lub też uzależniają skład interesariuszy od kontekstu i specyfiki programu.

Dobór metod i technik uzależniany jest od decyzji uczestników badania. Wybierają oni te narzędzia, które najlepiej spełniają ich oczekiwania – mogą to być metody ilościowe, eksperymentalne, jakościowe. Niemniej samo podejście wymaga użycia metod jakościowych i interaktywnych, po to by nawiązać i utrzymać kontakt z interesariuszami.

Podejście realistyczne (ang. *realistic/realist evaluation*)[14] skupia się na kontekście i procesach programu – w jakich warunkach, w jaki sposób i w stosunku do jakich grup dana interwencja działa lub nie (Henry, 2005; Pawson, Tilley, 2005). Rdzeniem jest skonstruowanie, a następnie przetestowanie teoretycznego modelu działania programu (Chen, 1990, 2005). Procedura badawcza jest czteroetapowa. Punktem wyjścia jest zidentyfikowanie mechanizmu programu – teorii lub założeń, którymi program się kieruje (założoną przez operatorów programu relację przyczynowo-skutkową) i stworzenie wstępnej hipotezy jego działania. Krokiem drugim jest opis kontekstu interwencji, w tym

---

[12] Termin został wprowadzony przez Michela Pattona (I edycja 1978; 1997). Zbliżone merytorycznie koncepcje to: *responsive evaluation* promowana przez Roberta Stake'a (2005) czy *user-oriented evaluatio*n Marvina Alkina (2004: 49).

[13] Stosowana zasada to ang. *active-reactive-adaptive* – bądź aktywny (w identyfikowaniu potencjalnych użytkowników, włączaniu ich, precyzowaniu pytań) – odpowiadaj (na zmieniającą się sytuację badania) – dostosowuj się (pytania badawcze i metody wraz ze zmieniającą się sytuacją i pogłębianym rozumieniem programu) (Patton, 2005: 431).

[14] Mathison (2005: 359–366) odróżnia *realist evaluation* od *realistic evaluation*. Wynika to raczej z podkreślenia wkładu dwóch różnych grup autorów (G.T. Henry, M.M. Mark z US vs. R. Pawson, N. Tilley z UK), nie zaś z samego meritum podejścia. Bardzo zbliżonym podejściem jest również *theory-based evaluation*. Promujący je autorzy, tacy jak Peter Rossi (1999) czy Huey-Tsyh Chen (1990) stawiają co prawda dużo silniejszy akcent na metody ilościowe (w tym łączenie realizmu z metodami *quasi*-eksperymentalnymi), ale główne założenia są analogiczne.

systemu relacji społecznych, interpersonalnych i instytucjonalnych, w których działa program. Element trzeci to zidentyfikowanie efektów – zamierzonych i niezamierzonych konsekwencji programu. Wreszcie ostatni krok to stworzenie modelu, który wiąże wszystkie trzy wcześniejsze elementy, przeanalizowanie relacji między nimi i zbadanie zróżnicowania efektów pod kątem weryfikacji wstępnej hipotezy działania programu[15]. Ewaluator w tym podejściu pełni rolę analityka odtwarzającego logikę relacji. Podejście realistyczne nie określa preferowanych metod, choć niektórzy z autorów preferują metody ilościowe i *quasi*-eksperymenty (Chen, Rossi, 1983). Mogą być one zarówno jakościowe, jak i ilościowe. Są dobierane w zależności od skali programu, charakteru badanych relacji, etc. Z reguły ewaluatorzy korzystają z modeli i map logicznych, które są następnie testowane i weryfikowane w oparciu o badania *quasi*-eksperymentalne, wywiady, panele dyskusyjne czy przegląd literatury (Henry et al., 1998; Pawson, 2006; Pawson, Tilley, 1997).

W rzeczywistości programów publicznych rzadko zdarza się, by któreś z powyższych podejść było reprezentowane w formie czystej. Należy je traktować raczej jako sposób myślenia o ewaluacji. W praktyce, szczególnie przy ewaluacji dużych programów, niektóre z tych sposobów myślenia są łączone (np. podejście realistyczne z podejściem partycypatywnym, *quasi*-eksperymentalne z realistycznym). Dlatego też w odniesieniu do programów UE konieczna jest inna, nieco uproszczona typologia podejść ewaluacyjnych, łącząca powyższe logiki. Inspiracją mogą być dwa źródła: Alkin, Christie, Rose (2006) oraz Dahler-Larsen (2005), starające się połączyć omawiane powyżej typy i podejścia do badania w kontekście praktyki programów publicznych.

Alkin ze współpracownikami (2006: 398–401) proponują pięć typów. Pierwszy to ewaluacja jako badanie naukowe (ang. *evaluation as research*) obejmujące te podejścia, które opierają się na metodach *quasi*-eksperymentalnych. Koncentrują się raczej na poszerzaniu wiedzy, wyjaśnianiu zjawisk społecznych, weryfikowaniu teorii. Odbiorcami są decydenci tworzący programy, liderzy, społeczność akademicka. Typ drugi to ewaluacje zorientowane na cele (ang. *objectives-oriented*), czyli koncentrujące się na specyfikacji celów programu i pomiarze efektów w odniesieniu do tychże celów. Ich odbiorcami są głównie liderzy i politycy decydujący o całych strategiach i politykach. Typ trzeci to ewaluacje zorientowane na zarządzanie (ang. *management-oriented*), analizujące procesy i działania programu mające dostarczyć użytecznych, technicznych informacji decydentom i menedżerom programu. Typ czwarty – ewaluacje zorientowane na ekspertyzę (ang. *expetise-oriented*) – zajmują się oceną wartości danego programu (koncentrując się na efektach) w oparciu o doświadczenie i wiedzę ewaluatora. Ewaluatorzy poszukują również czynników, które wpływają i wzmacniają zaobserwowane efekty. Używane tu metody to obserwacja, wizyty terenowe, wywiady. Odbiorcami są decydenci i liderzy rozwijający dany program, ale również szersza opinia pub-

[15] Te cztery kroki badawcze ujmowane są w formie modeli logicznych, np. modelu CMOC (ang. *Context-Mechanizm-Outcomes-Pattern Configuration*) (Pawson, Tilley, 2005: 365), prościej określanym jako kontekst + mechanizm = efekty (szerzej o modelach logicznych w ewaluacji patrz: Frechtling, 2007).

liczna. Wreszcie piąty wyróżniony przez autorów typ podejścia badawczego to zorientowanie na uczestników (ang. *participant-oriented*) mające odpowiadać na konkretne potrzeby informacyjne interesariuszy. Nacisk jest położony na interakcje i uczestnictwo we wszystkich etapach badania. Jego odbiorcami są kierujący programem, personel i uczestnicy.

Dahler-Larsen (2005: 622–636) natomiast wyróżnia trzy podejścia do ewaluacji programów publicznych: zorientowane na cele, oparte na teoriach oraz podejście responsywne i partycypatywne. Podaje też ich ograniczenia.

Podejścia zorientowane na cele (ang. *goal-oriented*) są analogiczne jak ujęcia w typologii Alkina. Kryterium oceny wartości programu jest to, na ile wypełnił on zapisane cele. Ewaluator koncentruje się na zbadaniu procesu osiągania celów (od nakładów, przez działania do efektów) i stopnia ich wypełnienia (ignorując efekty uboczne). Ograniczenia tego podejścia to: słabe uzasadnienie relacji przyczynowo-skutkowych między interwencją a efektami (są one zakładane nieco na wyrost), raczej ograniczona formatywna funkcja (bez wchodzenia w procesy programu, jego kontekst i efekty uboczne trudno jest dawać praktyczne rekomendacje), ignorowanie efektów ubocznych (nieplanowanych), zbytnie przywiązanie do wskaźników efektów, które są często ograniczające przy niektórych tzw. miękkich polach interwencji (np. innowacje, pomoc społeczna, zdrowie, etc).

Podejścia oparte na teoriach (ang. *theory-based*) koncentrują się na zidentyfikowaniu, jakie założenia i teorie leżą u podstaw danego programu (mogą być to jasne lub nieuświadomione założenia)[16], a następnie sprawdzaniu czy rzeczywiście takie relacje zachodzą, czy też mogą się pojawić inne ciągi przyczynowo-skutkowe[17]. Opierają się więc na krytycznym myśleniu o programach, ich założeniach, celach i efektach (również ubocznych). Te podejścia koncentrują się na tym, jak i dlaczego działa program, dają więc bazę do praktycznych rekomendacji. Co więcej, pomagają kwantyfikować i wartościować efekty (w tym te nieplanowane), pokazać kontekst, zweryfikować relację przyczynowo-skutkową, kumulować wiedzę. Ich ograniczeniem jest założenie o jednej dominującej teorii programu. Tymczasem założenia na temat działania programu mogą być niespójne i odmienne wśród różnych interesariuszy. Są też ściśle przywiązane do kontekstu programu (co może być też uznawane za zaletę). Zwraca się też uwagę, że takie „teorie" są zasadniczo zredukowanym i uproszczonym opisem praktyki, a nie ujęciem jej natury.

Wreszcie podejścia responsywne i partycypatywne (ang. *responsive & participatory*) zakładają, że nie ma zewnętrznie obiektywnego punktu odniesienia do oceny programu i, że ocena jest warunkowana lokalnymi czynnikami oraz postrzeganiem uczestników[18]. Ci, którzy uczestniczą w programie (na różnych

---

[16] Założenia, czyli przeświadczenie, jakim kierują się decydenci, operatorzy i interesariusze programu, że dane działania przyniosą określone efekty (a więc przeświadczenie o określonej relacji przyczynowo-skutkowej).

[17] Przykładem jest tu opisywana wcześniej ewaluacja realistyczna – prace Pawsona, Tilleya, Chena i Rossiego.

[18] Jako przykłady takich ujęć autor wymienia omawiane wcześniej w tej pracy podejście partycypatywne, podeście uprawniające i podejście zorientowane na użycie badania.

szczeblach), są współodpowiedzialni za jego ocenę. Interesariusze są więc włączani w proces badania – formułowania celów, analiz i ostatecznych ocen. Akcent jest stawiany na produkcję użytecznej w danym kontekście wiedzy, adresowanej do konkretnych aktywnych odbiorców. Głównym zarzutem stawianym temu podejściu jest rozmycie obiektywizmu niezależności ewaluatora, zatracenie charakteru ewaluacji jako niezależnej oceny na koszt interaktywnego procesu negocjacji, animacji, dyskusji. Trudnym do uniknięcia jest zdominowanie kierunku i wniosków badania opiniami kluczowych aktorów programu. A nawet, jak zauważają niektórzy, utrata kontroli ewaluatora nad całym procesem badawczym. Słabością może być nacisk na praktykę w odniesieniu do wybranych odbiorców – to może prowadzić do zbytniego skoncentrowania się na rzeczach drobnych kosztem szerszej oceny strategii programu.

Łącząc te dwie typologie (a także odwołując się do prezentowanych poprzednio podejść) na potrzeby funduszy strukturalnych, można zaproponować następujący podział:

- Podejście naukowe (ang. *evaluation as research*) – takie ewaluacje koncentrują się na wyjaśnianiu zjawisk społecznych i założeń interwencji poprzez weryfikowanie relacji przyczynowo-skutkowych. Robią to głównie w oparciu o metody eksperymentalne i *quasi*-eksperymentalne. To podejście może często ignorować cele programu, po to by uchwycić jego rzeczywiste efekty (w tym efekty uboczne) i oddzielić je od zmian kontekstowych (tzw. *goal free evaluation*). Silny akcent jest stawiany na rzetelność projektu i procesu badawczego (trafność, dobór próby, kontrolowanie zmiennych, etc.). Przeważają tu strategie i metody ilościowe. Podstawą filozoficzną tego podejścia jest pozytywizm.
- Ewaluacja oparta na teorii (ang. *theory-based*) koncentruje się na identyfikacji, zweryfikowaniu i ocenie założeń (czy to jawnych, czy nieuświadomionych) na temat mechanizmu interwencji, relacji między działaniami a efektami, które leżą u podstaw danego programu, a którymi kierują się aktorzy realizujący dany program. Założenia te są oceniane w odniesieniu do rzeczywistych efektów programu. Takie podejście sięga więc po metody podobne do podejścia naukowego, tyle że traktuje je bardziej swobodnie. Wnioskowanie opiera w równej mierze na *quasi*-eksperymentach, pomiarach własnych, jak i na wnioskach z wcześniejszych, analogicznych badań (adaptując je do kontekstu badanego programu). W procesie analitycznym używane są, na etapie formułowania hipotez, przeglądy literatury (poszukiwanie analogicznych studiów interwencji), przegląd dokumentów, modele logiczne, zaś na etapie weryfikacji teorii i tworzenia rekomendacji – badania ilościowe, analizy przyczynowo-skutkowe, ale także wywiady, ankiety, studia przypadków. Tym, co także odróżnia ewaluację opartą na teorii od podejścia naukowego jest to, że opiera się ona na realizmie (podczas gdy podejście naukowe na pozytywizmie).
- Ewaluacja zorientowana na cele (ang. *objectives-oriented*) to badania koncentrujące się wyłącznie na weryfikacji stopnia osiągania zakładanych celów programu w planowanym procesie jego wdrażania. Wyłączają więc często z analizy kwestie kontekstowe i koszty efektów ubocznych. Podstawowy-

mi źródłami są tu dane z monitoringu (w tym pomiary stanu wyjściowego programu), wywiady i ankiety z uczestnikami procesu, a dla programów o większej skali badania ilościowe i modele ekonometryczne. Zestaw metodologiczny może być analogiczny do podejścia naukowego, tyle że użycie poszczególnych metod nie jest tak restrykcyjne.

- Podejście zorientowane na zarządzanie (ang. *management-oriented*) koncentruje się na analizie procesów wdrażania i ewentualnych pojawiających się blokad, problemów. Aby dostarczyć menedżerom programu konkretnych i użytecznych informacji (głównie technicznych), badania te używają głownie metod miękkich – analizy dokumentów, procesów, wywiadów, ale także badań ilościowych (w formie badań ankietowych z reguły na ograniczonej próbie) czy grup fokusowych.
- Podejście responsywne i uczestniczące (ang. *responsive & participatory*) ocenia program przez pryzmat lokalnych czynników i opinii uczestników programu. Główny nacisk kładzie na konkretne bieżące potrzeby informacyjne interesariuszy, zaś w procesie badawczym najważniejszymi są dialog, interakcje i aktywizowanie użytkowników badania (przede wszystkim kierujących programem, personelu, najważniejszych beneficjentów). Działania te mają poszerzyć punkt widzenia poszczególnych uczestników programu i wzmocnić współpracę między nimi. Najbardziej odpowiednimi są podejścia interaktywne – panele dyskusyjne, fokusy, wywiady, warsztaty oraz cała gama technik interaktywnych.

Powyższe typy podejść można wpisać pomiędzy dwa bieguny orientacji badań ewaluacyjnych. Podobnie jak w przypadku funkcji, zaznacza się tu dychotomia i wybór „coś za coś". Tym razem uwidaczniają się jednak jej głębsze korzenie teoretyczne. Podstawą filozoficzną tego wyboru są dwa odmienne, opisane już paradygmay badawcze. Z jednej strony jest to pozytywizm. Współgra on dobrze z orientacją na obiektywną ocenę efektów interwencji. Na drugim biegunie jest paradygmat konstruktywizmu – akcentujący mnogość perspektyw, wielość głosów interesariuszy i współgrający z nią zestaw metod miękkich.

Kończąc przegląd typologii ewaluacji należy podkreślić raz jeszcze, że nie są one rozłączne. Typy są rozciągnięte pomiędzy dwoma biegunami, ale typy z odmiennych typologii (wg. przedmiotu, czasu i podejść) współwystępują ze sobą. W Tabeli 4 zestawiono zidentyfikowane typologie, zaznaczając, które z typów ewaluacji najczęściej współwystępują ze sobą.

Tabela 4. Zestawienie typologii ewaluacji

Typ ewaluacji	Charakterystyka	Współwystępowanie
Podział ze względu na przedmiot ewaluacji		
Ewaluacja projektu	Jej przedmiotem jest pojedynczy projekt – najczęściej o charakterze innowacyjnym, eksperymentalnym.	*Ex post*, zewnętrzna, naukowa, zorientowana na cele lub oparta na teorii
Ewaluacja programu	Jej przedmiotem jest program (złożony z szeregu interwencji, projektów, z reguły o kilku celach strategicznych i kilkunastu celach operacyjnych).	Wszystkie typy czasowe, zwykle zewnętrzna, możliwe wszystkie typy podejścia badawczego
Ewaluacja polityki	Obejmuje analizę funkcjonowania i efektów całej polityki (np. rozwoju regionalnego).	Wszystkie typy czasowe, zewnętrzna, strategiczna, zorientowana na cele lub oparta na teoriach
Ewaluacja tematyczna	Koncentruje się na wybranej kwestii funkcjonowania i analizuje ją przekrojowo dla kilkunastu programów (np. analiza zasady partnerstwa).	*Ex post*/bieżąca, strategiczna, zewnętrzna, podejścia zorientowane na zarządzanie lub cele
Metaewaluacja	Jej przedmiotem są wcześniejsze ewaluacje pojedynczych programów. Podsumowuje je, wyciąga wspólne elementy i porównuje ich jakość.	Możliwe wszystkie typy czasowe, strategiczna, zewnętrzna, naukowa/ zorientowana na cele/ oparta na teorii
Podział ze względu na cel ewaluacji		
Ewaluacja formatywna – operacyjna	Ocenia formę, plan oraz proces wdrażania danej interwencji. Koncentruje się na bieżącej pomocy operacyjnej dla zarządzających interwencją.	*Ex ante*/bieżąca, zewnętrzna, ale możliwa wewnętrzna, zorientowana na cele/ zorientowana na zarządzanie/ partycypatywna/ oparta na teorii
Ewaluacja podsumowująca – strategiczna	Ma charakter opisowy i wyjaśniający efekty interwencji. Jest adresowana do szerszej publiki, ma pomóc w podjęciu decyzji strategicznych – np. o kontynuacji czy zatrzymaniu programu.	*ex post*, możliwe bieżące (gdy program jest zaawansowany), zewnętrzna, podejście naukowe/zorientowane na cele/ oparte na teoriach
Podział ze względu na moment podjęcia badania		
Ewaluacja *ex ante*	Prowadzona przed rozpoczęciem interwencji. Ocenia jej logikę.	Formatywna, zewnętrzna/wewnętrzna, łączy podejścia zorientowane na zarządzanie i na cele, podejścia oparte na teoriach
Ewaluacja średniookresowa i jej uaktualnienie	Wykonywana jest równo w połowie trwania wieloletniego programu i powtarzana (w formie aktualizacji) w 3/4 programu. Ocenia postępy we wdrażaniu i identyfikuje ewentualne korekty w strategii.	Formatywna/podsumowująca (gdy program jest zaawansowany) zewnętrzna lub wewnętrzna, możliwe wszystkie podejścia

Typ ewaluacji	Charakterystyka	Współwystępowanie
Ewaluacja *bieżąca*	Wykonywana jest w dowolnym momencie implementacji programu. Może być badaniem punktowym albo składać się z szeregu badań. Prowadzona jest na małą skalę, w odniesieniu do konkretnego problemu wdrażania.	Formatywna/podsumowująca (gdy program jest zaawansowany) zewnętrzna lub wewnętrzna, możliwe wszystkie podejścia (zależne od stanu programu)
Ewaluacja *ex post*	Jest wykonywana po zakończeniu interwencji (zwykle w 2 lata po zakończeniu wszystkich działań). Skupia się na ocenie rezultatów i długookresowych efektów.	Podsumowująca, zewnętrzna, podejścia oparte na teoriach, zorientowane na cele lub naukowe
Podział ze względu na administrację badania		
Ewaluacja wewnętrzna	Wykonywana przez instytucję bezpośrednio zaangażowaną we wdrażanie danej interwencji.	Formatywna, *ex ante*, bieżąca
Ewaluacja zewnętrzna	Badanie zlecane poza instytucję prowadzącą dany program – wykonywane przez zespoły niezależnych ekspertów.	Formatywna/podsumowująca, wszystkie typy czasowe, wszystkie podejścia
Podział ze względu na podejście badawcze		
Ewaluacja jako badanie naukowe	Poszerza wiedzę, wyjaśnia zjawiska społeczne, weryfikuje szersze teorie naukowe, relacje przyczynowo-skutkowe działań i efektów programu.	Podsumowująca, zewnętrzne, *ex post*, średniookresowa lub *on-going* (przy zaawansowanych programach)
Ewaluacja zorientowana na cele	Weryfikuje stopień osiągania zakładanych celów programu w planowanym procesie jego wdrażania.	Formatywna/podsumowująca, wszystkie typy czasowe, zewnętrzne
Ewaluacja oparta na teorii	Identyfikuje, weryfikuje i ocenia założenia na temat mechanizmu interwencji, relacji między działaniami a efektami, które leżą u podstaw danego programu, a którymi kierują się aktorzy realizujący dany program. Założenia te są oceniane w odniesieniu do rzeczywistych efektów programu.	Formatywna/podsumowująca, wszystkie typy czasowe, zewnętrzne
Ewaluacja zorientowana na zarządzanie	Analizuje procesy wdrażania i ewentualne blokady, problemy. Dostarcza menedżerom programu konkretnych i użytecznych informacji technicznych.	Formatywna/podsumowująca, wszystkie typy czasowe bez *ex post*, zewnętrzna lub wewnętrzna
Ewaluacja partycypatywna	Ocenia program przez pryzmat lokalnych czynników i opinii uczestników programu. Główny nacisk kładzie na konkretne potrzeby informacyjne interesariuszy, najważniejszymi są dialog, interakcje i uczestnictwo potencjalnych użytkowników badania.	Formatywna/podsumowująca, *ex ante*, bieżąca, zewnętrzna

Źródło: opracowanie własne.
Uwaga: w tabeli zaznaczano tylko ewaluację bieżącą, traktując ją równorzędnie z ewaluacją średniookresową.

2.3. KONCEPCJE WYKORZYSTANIA EWALUACJI

Dyskusja, a wraz z nią badania użycia ewaluacji pojawiły się w latach 60. w USA. Impulsem było rozczarowanie znikomym wpływem wielu prowadzonych badań ewaluacyjnych na rzeczywistość ówczesnych programów publicznych. Większość z nich została wykorzystana politycznie lub przeszła zupełnie bez echa (Patton, 2004: 338; Preskill, Torres, 2000). To zmusiło ewaluatorów do postawienia pytań o mechanizmy użycia efektów ich badań, sens i cel ewaluacji w sektorze publicznym, rolę ewaluatorów w rozpowszechnianiu efektów badania. Debata toczona w USA od połowy lat 80. stała się centralnym elementem różnicowania koncepcji ewaluacji, ujęć badawczych, a także ważnym elementem nowych podejść do zarządzania sektorem publicznym (porównaj rozdział 3). Od połowy lat 80. koncepcje wykorzystania ewaluacji zaczęły czerpać głównie z podejść do procesów uczenia się organizacji, zarządzania zmianą, tworzenia wiedzy (Scriven, 1991: 369). Należy jednak podkreślić odmienność tematyki użycia ewaluacji (ang. *evaluation use, utilization*) od kwestii użycia badań (ang. *research utilization, applied research*). W badaniach (włączając w to badania społeczne) informacje i tworzona na ich podstawie wiedza podlega generalizacji – jest wykorzystywana powszechnie, stanowi wkład do ogólnego zasobu wiedzy w danej dziedzinie, pozostając aktualna cały czas (aż do kolejnej weryfikacji) dla wszystkich osób i grup, które zechcą po nią sięgnąć. Natomiast ewaluacja jest związana kontekstem – wytworzona wiedza odnosi się do danego kontekstu, w którym powstała, ma służyć określonej grupie użytkowników w danym czasie (Alkin, Taut, 2003: 3; Cousins, Shulha, 2006: 267).

2.3.1. Rodzaje wykorzystania ewaluacji

Użycie czy wykorzystanie ewaluacji (ang. *evaluation utilization, evaluation use*) odnosi się, najszerzej rzecz ujmując, do rzeczywistej roli, jaką spełnia ewaluacja w sektorze publicznym (wypełnienia jej funkcji w praktyce). Tak szeroka definicja wymaga jednak rozpatrywania jej w przynajmniej czterech wymiarach – tego, co jest używane (ang. *source of use*), przez kogo, w jaki sposób i z jaką świadomością (ang. *way and intention*), w jakiej perspektywie czasowej. Wszystkie te cztery kwestie zostaną omówione poniżej.

W precyzowaniu definicji użycia ewaluacji kwestią pierwszą jest pytanie o to, co jest używane. Termin „użycie ewaluacji" może obejmować (Alkin, Taut, 2003: 6; Kirkhart, 2000):

- Użycie jej wniosków, rezultatów badania prezentowanych najczęściej w formie raportu ewaluacyjnego (ang. *findings use*);
- Użycie samego procesu badania – interakcji i dyskusji, które mają miejsce podczas prowadzenia badania między ewaluatorami, operatorami programu, instytucjami i grupami osób zaangażowanymi w program (ang. *process use*).

Drugim zagadnieniem jest pytanie, kto wykorzystuje ewaluację. Można to opisać w dwóch perspektywach. Pierwszą jest skala użycia – ewaluacja może

być używana przez jednostki, grupy, i/lub całe organizacje, instytucje (Cousins, 2003).

Drugą optyką spojrzenia na to, kto używa ewaluacji, jest stopień powiązania użytkowników z ewaluandą (tzn. badanym programem). Adaptując do potrzeb europejskich i nieco rozszerzając typologię Weiss (2005–295) można zidentyfikować następujących użytkowników ewaluacji:

- personel i zarządzający daną interwencją;
- interesariusze programu – aplikujący o projekty, beneficjenci, beneficjenci końcowi;
- szersza publika europejska – personel i zarządzający analogicznymi interwencjami;
- grupy interesu (ang. *interest groups*) – stowarzyszenia, związki, organizacje firm, klientów, regionów, gmin związane z tematyką danej interwencji;
- sieci tematyczne danej kwestii (ang. *issue networks*) – eksperci, osoby, instytucje badawcze, *think tanki* znające daną tematykę, śledzące jej ewolucję i zainteresowane wpływaniem na kształt danej polityki[19];
- główni interesariusze danej polityki – instytucje tworzące daną politykę – zestawy interwencji, a konkretnie personel tychże instytucji – odpowiednich dyrekcji Komisji Europejskiej, krajowych ministerstw (decydenci wyższej rangi i personel przygotowujący sprawozdania);
- mass-media – dziennikarze prasowi i telewizyjni, internetowe serwisy informacyjne;
- szeroka opinia publiczna.

Zagadnieniem trzecim jest pytanie, „jak wykorzystywana jest ewaluacja", czyli jakie formy mogą się kryć pod pojęciem „użycia ewaluacji". W literaturze przedmiotu klasycznym jest trójpodział na użycie instrumentalne, koncepcyjne i symboliczne (Alkin, Taut, 2003: 5; Leviton, Hughes, 1981; Mark, Henry, 2004: 36; Rich, 1977; Shulha, Cousins, 1997: 197; Weiss et al., 2005: 13–14):

- Użycie instrumentalne (ang. *instrumenal use*) to proste do zidentyfikowania i udokumentowania przełożenie wyników ewaluacji na decyzje, działania, sytuację odbiorców ewaluacji (np. wpływ rekomendacji z ewaluacji na podjęcie decyzji co do zmiany kierunku programu, wstrzymania lub rozszerzenia interwencji);
- Użycie koncepcyjne (ang. *conceptual use*) to wpływ ewaluacji na zmianę sposobu myślenia jej odbiorców o istocie realizowanych programów czy procesów społecznych[20];
- Użycie symboliczne (ang. *symbolic use*) to użycie ewaluacji do uzasadnienia de facto podjętych już decyzji; ewaluacja jest więc jedynie merytoryczną podkładką do decyzji powziętych w oparciu o intuicję, doświadczenie

[19] W przypadku UE mogą mieć one charakter formalny – tzw. programów parasoli (*umbrella programme*) – służących wymianie doświadczeń, budowie *know-how* danej polityki. W przypadku INTERREG przykładem może być Program INTERACT.

[20] Używany jest czasami termin „oświecenia" (ang. *enlightment*) odbiorców ewaluacji czy „uczenia się" w oparciu o ewaluację. Terminy te są jednak, jak słusznie zauważa Weiss (2005), całkowicie ze sobą zbieżne.

zawodowe decydentów, ze względów politycznych, społecznych, instytucjonalnych, itp.[21]

Dwa pierwsze z powyższych rodzajów użycia można odnieść do popularnych koncepcji „*single & double-loop learning*" wypracowanych przez Chris Argyrisa i Donalda Schona na temat uczenia się jednostek i organizacji. *Single loop learning* to szybka, techniczna odpowiedź organizacji na zidentyfikowany problem (odpowiadałaby więc użyciu instrumentalnemu). *Double-loop learning* to reakcja organizacji czy jednostek, która pociąga za sobą fundamentalną zmianę w myśleniu – przede wszystkim na temat norm, głównych celów, założeń polityki, którą kieruje się dana organizacja (tak głęboka zmiana odpowiadałaby użyciu koncepcyjnemu) (Smith, 2005). Warto tu też podkreślić, za Weiss (2005) i Cronbachem (Cronbach, Associates, 1981), że użycie koncepcyjne jest najczęściej procesem kumulatywnym – dane badanie ewaluacyjne jest tylko elementem, kamykiem w lawinie zmieniającej myślenie o całym programie czy polityce[22].

Niektórzy autorzy rozszerzają powyżej przedstawiony trójpodział o „użycie procesowe", definiując je jako zmiany (w myśleniu lub zachowaniu) wywołane prowadzeniem badania ewaluacyjnego/uczestniczeniem w procesie ewaluacji (porównaj: Balthasar, 2006; Johnson, 1998; Stern, 2005). Głównymi efektami użycia procesowego miałyby być: uczenie się, rozwijanie sieci, budowanie wspólnego rozumienia (programu, projektu), wzmocnienie projektu lub wzmacnianie morale (Forss et al., 2002: 33–37)[23]. Jak zauważa Weiss (2005: 14), taka rozszerzona kategoryzacja miesza nieprzystające do siebie typy. Trzy tradycyjnie wyróżniane rodzaje użycia opisują efekt użycia – formę zmiany, jaka nastąpiła, podczas gdy „użycie procesowe" mówi o tym „co" (proces – interakcje) i „w jaki sposób" wywołało zmianę. Identyfikowane efekty użycia procesowego (w dużej mierze szeroko pojęty proces uczenia się) można tymczasem włączyć w „użycie koncepcyjne" (wspólne rozumienie programu) lub „instrumentalne" (zbudowane sieci współpracy, wzmocniony projekt).

Przy sposobie wykorzystania ewaluacji Kirkhart (2000: 11–13) zwraca również uwagę na intencjonalność procesu. Ewaluacja może być wykorzystywana świadomie – jej użycie może być celowe, zaplanowane i wspierane przez

---

[21] Często podkreślany jest pejoratywny charakter tego typu użycia ewaluacji (nazywanego również „politycznym"). Tymczasem Weiss (2005) zauważa, iż takie polityczne wykorzystanie ewaluacji nie musi być negatywne: „nie ma nic złego w użyciu ewaluacji jako dodatkowego argumentu we wspieraniu decyzji i wyborów, które już mają silne wsparcie". Problemem może być manipulowanie danymi z ewaluacji, zniekształcanie jej wyników (kłócących się z daną decyzją), tak, by dać poparcie danej decyzji – dotyczy to jednak innego typu wykorzystania ewaluacji (dyskutowanego w dalszej części tekstu).

[22] Pisząc o użyciu ewaluacji w procesie politycznym Weiss podaje ciekawą, aczkolwiek niespójną (mieszającą szerokie funkcje z technicznymi sposobami i efektami użycia) kategoryzację użycia ewaluacji przez aktorów procesu politycznego. Badanie może być sygnałem ostrzegawczym (że coś nie działa), może być wskazówką w poprawianiu polityk i programów, może pomagać w przeformułowaniu interwencji lub też mobilizować do wsparcia – dać impuls do budowy koalicji wokół programu (Weiss, 2005: 296–297).

[23] Angielskie terminy to: *learning to learn, developing networks, creating shared understanding, strengthening the project, boosting morale.*

ewaluatorów (ang. *intended use*), lub też może być w dużej mierze przypadkowe, nieplanowane – tak przez zlecających, jak i wykonawców.

Odnosząc trzy zidentyfikowane kategorie form użycia ewaluacji do wcześniej wymienionych przedmiotów użycia można stwierdzić, że przedmiotem użycia instrumentalnego i symbolicznego jest najczęściej raport – wnioski z badania. Natomiast zmiany w myśleniu (użycie koncepcyjne) wywołane są procesem uczenia się zachodzącym podczas realizacji badania – w interakcjach między ewaluatorami a interesariuszami. Użycie instrumentalne jest w dużej mierze przewidywalne i możliwe do zaplanowania, przy wykorzystaniu koncepcyjnym zaś wpływ samego badania – cech i komentarzy nim wywołanych – jest w dużej mierze nieprzewidywalny.

Rozszerzając dyskusję nad sposobem wykorzystania ewaluacji, należy zwrócić również uwagę na możliwość braku jej użycia, błędnego użycia czy też nadużyć.

Tak więc prócz trzech wyróżnionych już form użycia (instrumentalnej, koncepcyjnej i symbolicznej), które szerzej można określić jako zasadne (ang. *legitimate*), wyróżniany jest brak użycia (ang. *non-use*) oraz cały szereg błędnych, niewłaściwych form wykorzystania ewaluacji (ang. *misuse*).

Alkin ze współpracownikami (Alkin, Coyle, 1988: 336–337; Christie, Alkin, 1999) wstępnie różnicują błędne użycie na trzy typy, odnosząc je do przedmiotu, którego dotyczy. I tak nadużyciem może być sama decyzja o podjęciu badania (a więc samo zlecenie ewaluacji), na przykład kiedy zostaje zlecona tylko w celu przyciągnięcia uwagi lub wydania pieniędzy. Po drugie, niewłaściwe może być użycie procesu badawczego – gdy stosuje się go, by opóźnić podjęcie decyzji, działania, uniknąć odpowiedzialności za decyzje, przerzucić ją – choć częściowo – na aktorów zewnętrznych. Wreszcie, po trzecie, źle użyte mogą być same rezultaty (wnioski) z badania. Do takich należy: nadmierne upraszczanie wniosków, przedwczesne prezentowanie fragmentarycznych wniosków, silne eksponowanie dowodów, które przez ewaluatorów uznane zostały za słabo poparte, wybiórcze prezentowanie wyników i wniosków zmieniające wydźwięk całości badania, zmienianie sformułowań w skrótach konkluzji, przypisywanie badaniu wyników i znaczenia innego niż rzeczywiste, rażący brak użycia (unikanie upublicznienia raportu). Te złe użycia raportów są w zależności od intencji użytkownika określane jako błędne użycie lub rażące nadużycie (ang. *blatant non-use, abuse*).

Analizując złe użycie ewaluacji, autorzy ci wprowadzają kryterium odnoszące się do samej jakości ewaluacji – rozgraniczają między użyciem dobrej i złej (słabej, niespełniającej standardów naukowych) ewaluacji. Używanie słabych merytorycznie ewaluacji jest również kategoryzowane jako błędne użycie, może być równie, jeśli nie bardziej destruktywne niż manipulowanie dobrą ewaluacją. W takim wypadku brak użycia jest rzeczą pozytywną i określa się go jako uzasadniony (ang. *justified non-use*).

Najnowszą próbą syntetycznego ujęcia kwestii form użycia ewaluacji (w tym niewłaściwego użycia i nadużyć w jej wykorzystaniu) jest schemat Cousinsa (2004; Cousins, Shulha, 2006):

Schemat 1. Rodzaje użycia ewaluacji

Użycie (*use*)

Niewłaściwe użycie (*misuse*)	**2. Niewłaściwe użycie (*misuse*)** Błędne użycie (*mistaken use*) Nadużycie – manipulacja (*mischievous use*)	**1. Idealne użycie (*ideal use*)** Instrumentalne użycie (*instrumental use*) Koncepcyjne użycie (*conceptual use*) Symboliczne użycie (*symbolic use*)	**Zasadne użycie (*legitimate use*)**
	3. Nieuzasadniony brak użycia (*unjustified non-use*) Zatajenie (*abuse, supperssion*)	**4. Uzasadniony brak użycia (*justified non-use*)** Racjonalny brak użycia (*rational non-use*) Polityczny brak użycia (*political non-use*)	

Brak użycia (*non-use*)

Źródło: (Cousins, 2004: 392).

Ostatnim (obok przedmiotu, rodzaju użytkowników i ich intencji, a także form użycia) koniecznym do omówienia aspektem wykorzystania ewaluacji jest perspektywa czasowa, czyli odpowiedź na pytanie, „kiedy następuje wykorzystanie ewaluacji". Efekty ewaluacji mogą nastąpić bezpośrednio po jej zakończeniu i wpłynąć na badany program lub też pojawić się z czasem – oddziałując na kolejne edycje danego programu, albo przyszłe inicjatywy o analogicznym charakterze (Alkin, 2005; Shulha, Cousins, 1997)[24].

Analizując perspektywę czasową, niektórzy autorzy (Owen, 2007; Stern, 2005: 7–9) stawiają użycie instrumentalne (propagowane przede wszystkim przez Michela Q. Pattona) w kontraście do użycia koncepcyjnego (badanego przez Carol H. Weiss), określając to pierwsze jako bieżące i natychmiastowe, natomiast drugiemu przypisując cechy długoterminowe. Takie podejście wydaje się ograniczające. Jak zauważa sama Weiss (Weiss et al., 2005), instrumentalne użycie – podjęcie decyzji – również może być przesunięte w czasie. Zaś zmiany w myśleniu o programie jego użytkowników czy szerszej grupy interesariuszy mogą nastąpić natychmiast, nie przekładając się na jasne decyzje.

W ostatnich latach kilku autorów zaczęło podkreślać długoterminowe, szersze efekty ewaluacji (Henry, 2000; Henry, Mark, 2003; Kirkhart, 2000). Zwracają oni uwagę na to, że obecne koncentrowanie się na „użyciu ewaluacji" jako celu samym w sobie (a więc głównym celu ewaluacji) jest niewystarczającą motywacją dla jej prowadzenia, pozbawia ewaluatorów właściwego punktu odniesienia („kompasu moralnego") i redukuje sens ewaluacji do kwestii technicznych.

Autorzy ci postulują więc, by wąski termin „użycia ewaluacji" (ang. *evaluation use*), który definiuje się jako „zamierzone wykorzystanie przez konkretnych użytkowników" (Patton, 1997: 20) zastąpić szerszym terminem „oddzia-

[24] Kirkhart (2000: 14–17) wyróżnia trzy możliwości czasowe: natychmiastowe oddziaływanie (ang. *immediate influence*), użycie na zakończenie cyklu programu (ang. *end-of-cycle use*) oraz wykorzystanie długoterminowe (ang. *long-term influence*) – wykraczające poza trwanie zbadanego programu.

ływania ewaluacji" (ang. *evaluation influence*). „Oddziaływanie ewaluacji" oznacza zdolność do wpływania na osoby czy instytucje w sposób pośredni, obejmuje więc wieloaspektowe, kumulatywne i niezamierzone efekty badań ewaluacyjnych (Kirkhart, 2000: 7). Jako cel nadrzędny prowadzenia ewaluacji wskazywane jest nie jej użycie, ale przyczynianie się do podnoszenia społecznego dobrobytu (ang. *social betterment*) (Henry, 2000: 86; Henry, Julnes, 1998)[25]. Przyjmując taką szeroką interpretację, studia nad wykorzystaniem ewaluacji winny się koncentrować na tym, na ile konsekwencje i rezultaty ewaluacji w sposób przekonujący przyczyniają się (lub nie) do podnoszenia dobrobytu społecznego (Henry, Mark, 2003: 295). Natomiast ewaluatorzy winni myśleć w kategoriach procesów, na które ich badanie może wpływać. Powinni się zastanawiać, jak te procesy ukierunkować i wspierać, by przyczyniały się do osiągnięcia szerszego celu. Tym samym dyskusja o wykorzystaniu ewaluacji zostaje umieszczona na powrót w szerszym kontekście roli ewaluacji w sektorze publicznym, jej szerokich funkcji we współczesnych społeczeństwach demokratycznych.

Jak zauważają Alkin i Taut (2003) czy Cousins i Shula (2006), ta rozszerzona interpretacja wcale nie stoi w sprzeczności z dotychczasowym, węższym pojmowaniem użycia ewaluacji. Rozszerza je jedynie, wzbogacając je, integrując kwestie uczenia się, użycia koncepcyjnego i długookresowej perspektywy czasowej. Odnoszenie jej do praktyki ewaluacji stwarza jednak pewne problemy. Oddziaływanie można zaobserwować dopiero po pewnym czasie, a zależy ono od wielu czynników zewnętrznych. Ewaluatorzy mają więc znikomy, najczęściej jedynie pośredni wpływ na te procesy, nie mogą ich kontrolować. Ponadto analiza tych zjawisk – użycia ewaluacji, które często jest nieuświadomione – stwarza problemy natury operacyjnej. Stąd wskazanym jest, by ewaluatorzy koncentrowali się na maksymalizowaniu użycia ewaluacji (a więc kwestii, na które mają wpływ), zachowując jednak świadomość szerszego kontekstu i mechanizmów jej oddziaływania.

2.3.2. Czynniki wpływające na użycie ewaluacji

W literaturze przedmiotu podawanych jest szereg czynników wpływających na wykorzystanie ewaluacji. Patton wymienia kilkadziesiąt czynników kontekstowych, które należy uwzględnić przy analizie wykorzystania ewaluacji (od kontekstu i formy programu, poprzez moment podjęcia badania, cechy ewaluatorów i zamawiających, metody, formę i przebieg badania, po formę wyników i rekomendacji, sytuację polityczną i strategię rozpowszechniania) (Patton, 1997: 437–439). Huberman i Thurler podają około pięćdziesiąt czynników (1991 za Balthazar 2006), Shulha i Cousins (1997: 196) identyfikują dziesięć szerszych kategorii czynników[26], zaś Leviton i Hughes (1981) przytaczają podobną listę – przystawanie badania do potrzeb, dobrą komunikację

[25] *Social betterment* jest definiowane bardzo szeroko jako poprawa warunków społecznych, redukcja problemów społecznych i łagodzenie ludzkich cierpień (Henry, Julnes 1998).

[26] Są to: przystawanie studium do potrzeb zlecających, wiarygodność, zaangażowanie użytkowników, efektywność w komunikacji, potencjał informacyjny, potrzeby informacyjne klienta,

między badaczami a użytkownikami, dobre przełożenie wniosków na implikacje dla polityki i programu, wiarygodność wyników i zaufanie pokładane w ewaluacji, zaangażowanie użytkowników i siła ich wsparcia. Owen znacznie upraszcza ten temat, podając trzy czynniki użycia ewaluacji: strategię badania, cechy odbiorców i charakterystykę organizacji (jej zdolność zmiany, otwartość na krytykę, szybkość wewnętrznego przepływu informacji) (Owen, 2007: 106).

Wreszcie w kontekście programów UE, przy badaniu wykorzystania raportów z ewaluacji w samej Komisji Europejskiej (EPEC, 2005) jako kluczowe czynniki wzmacniające wykorzystanie zidentyfikowano: moment podjęcia badania i jego orientację (cel), wsparcie ze strony kadry kierowniczej i menedżerów wyższej rangi, jakość samego procesu badania i raportu końcowego, monitorowanie wdrażania rekomendacji (tzw. *follow up*). Za nieco mniej wpływowe uznano zaangażowanie potencjalnych odbiorców, charakterystykę odbiorców w KE i sposoby rozpowszechniania wyników (co tłumaczyć można specyfiką Komisji Europejskiej – silnie zhierarchizowanej struktury biurokratycznej).

Poszczególni autorzy eksponują różne czynniki jako kluczowe dla sukcesu wykorzystania ewaluacji. Weiss zauważa, że głównym czynnikiem wpływającym na ewaluację (szczególnie jej koncepcyjne użycie) jest polityka i sprzeczne interesy organizacji (Weiss, 1998b). Jednak jako czynnik możliwy do kontrolowania przez ewaluatora podaje solidność i jakość metodologii badania oraz wiarygodność zebranych i analizowanych danych (Weiss, Bucuvalas, 1980a: 256). Cronbach i Shapiro (1982) akcentują zewnętrzną zasadność badania (ang. *external validity*), czyli możliwość generalizacji wniosków i obserwacji z danego badania w stosunku do szerszej populacji programów. Cousins i Leithwood (1986), opierając się na przeglądzie badań czynników wykorzystania ewaluacji, podają jako trzy kluczowe: jakość badania, kompleksowość metodologii oraz jej wyrafinowanie. Jako główne czynniki techniczne podkreślana jest oczywiście jakość (komunikatywność) raportu końcowego i rekomendacji (Morriset et al., 1987; Rossi et al., 1999) oraz strategia rozpowszechniania wyników (European Commission, 1999b).

Część autorów, szczególnie ci promujący instrumentalne wykorzystanie ewaluacji, akcentują rolę interakcji z potencjalnymi użytkownikami i umiejętne prowadzenie badania w danym kontekście organizacyjnym. Czołowy autor tego nurtu, M.Q. Patton, w swojej koncepcji „ewaluacji zorientowanej na użycie" pisze o czynniku personalnym (zidentyfikowaniu i włączeniu kluczowych w danej organizacji osób, które zaczną się identyfikować z prowadzoną ewaluacją) jako krytycznym dla użycia badania. Cousins określa to jako *get buy in*[27]. Jako kolejne dwa czynniki Patton wymienia aktywne włączanie interesariuszy i elastyczność samego ewaluatora (zdolność do szybkiej reakcji i adaptacji) (Patton, 1997). Włączanie interesariuszy i regularne kontakty badaczy z bada-

---

zakładany stopień zmiany programu, postrzegana wartość danej ewaluacji jako narzędzia zarządzania, jakość procesu ewaluacyjnego, kontekst decyzyjny i polityczny.

[27] Weiss zauważa krytycznie, że taka strategia ma ograniczone zastosowanie w przypadku dużych interwencji, gdyż rzadko istnieje jeden ośrodek decyzyjny. Decyzje są wypadkową interesów, a także pewnej inercji danej organizacji biurokratycznej (Weiss, 2005: 288).

nymi i zlecającymi badanie – szczególnie na etapie planowania ewaluacji i negocjacji metod oceny – podkreślają również Owen i Rogers (1999: 65) oraz Stufflebeam (2001b). Ten ostatni dodatkowo eksponuje wagę jasno wypracowanego kontraktu badawczego (Stufflebeam, 2000b). Wreszcie King (1998), opisując podejście partycypatywne, jako główne czynniki sukcesu podaje dobre rozpoznanie struktury decyzyjnej (układu sił) oraz efektywną budowę zaufania między badaczami a uczestnikami-odbiorcami ewaluacji.

W literaturze przedmiotu podejmowano próby pogrupowania miriad zarysowanych powyżej czynników wykorzystania ewaluacji. Najnowsze, najbardziej obszerne podsumowanie szerokiej literatury na ten temat przedstawił Owen (2007: 113–122). Wyróżnia on dwie duże grupy czynników: charakterystykę środowiska (ang. *characteristics of the settings*) i charakterystykę ewaluacji (ang. *characteristics of the evaluation*). Do pierwszej włącza:

- Zaangażowanie danej grupy odbiorców badania w proces ewaluacji, intensywność dialogu i negocjacji między zamawiającym a ewaluatorami (w założeniu im więcej interakcji i mocniejszy konsensus wokół celów ewaluacji, metod, kryteriów oceny, tym bardziej użyteczne będą rekomendacje), ich rolę w ewaluacji programu, a także determinację w zmienianiu polityki/programu;
- Potrzeby informacyjne i typ informacji, jakie odbiorcy danego badania potrzebują do podejmowania decyzji i rozwiązywania problemów, tu również mieści się kwestia tzw. „konkurencyjnych informacji", to znaczy dodatkowych źródeł wiedzy o programie, które mogą wzmacniać wnioski z ewaluacji lub też kolidować z nimi;
- Cechy osobowe, czyli stosunek do ewaluacji pojedynczych, kluczowych osób z danej organizacji, ich doświadczenie i siła wpływu, pozycja w danej instytucji, cechy przywódcze, determinacja, entuzjazm, interesy; to kryterium powtarza postulaty Pattona o identyfikowaniu i przekonywaniu do ewaluacji osób kluczowych w danych instytucjach, gdyż one później będą promotorem rekomendacji i zmian[28];
- Proces decyzyjny (ang. *decision-making*), czyli sposób w jaki decyzje na temat programu są podejmowane, charakter decyzji i to, czego mają dotyczyć;
- Klimat polityczny odnosi się do orientacji politycznej dominującej w danej organizacji; gdy ewaluacja jest zbieżna z polityczną rzeczywistością, jest używana;
- Kwestie finansowe to aktualny poziom wsparcia finansowego danego programu i ewentualnych w nim zmian postulowanych ewaluacją; oczywistym jest, że rekomendowanie kosztownych zmian programowych ma małe szanse wdrożenia.

[28] Kryteria identyfikacji tych kluczowych osób to: osoby, które mogą użyć informacji z badania, osoby, dla których dana informacja czyni różnicę, osoby, na których pytania ewaluacja odpowiada, osoby, które są zainteresowane i chcą dzielić odpowiedzialność za ewaluację i jej użycie. Warto nadmienić, że te kluczowe osoby niekoniecznie muszą pochodzić z departamentu, który zlecił ewaluację.

Do kategorii „charakterystyki ewaluacji" Owen zalicza natomiast:

- właściwą orientację badania (ang. *relevance*) czyli stopień, w jakim badanie zostało skoncentrowane na kwestiach ważnych dla odbiorców, w praktyce oznacza to konieczność zidentyfikowania grup odbiorców i określenia ich potrzeb, kluczowych pytań;
- wiarygodność (ang. *credibility*), która odnosi się do tego, na ile ewaluator jest godny zaufania – jako człowiek i ekspert; jako główne cechy wymienia się dotychczasowe doświadczenie, brak uprzedzeń, pole specjalizacji i umiejętności zawodowe, elastyczność w kwestiach finansowych, elastyczność w doborze metod i podejść specyficznych dla danych programów. Na tych cechach ewaluator buduje swój autorytet wobec klienta i przede wszystkim zaufanie w relacjach z interesariuszami;
- jakość prowadzenia badania, szczególnie kwestie doboru metod, zbierania i analizy danych, czyli poprawność metodologiczna warsztatu badawczego;
- siłę wniosków i implikacje dla procesu decyzyjnego;
- komunikację – jakość i częstotliwość komunikacji podczas prowadzenia badania, konieczne wyjaśnienia, wsparcie informacyjne i nagłośnienie rezultatów oraz rekomendacji;
- stosowny moment raportowania (ang. *timeline of reporting*), czyli dostarczenie wniosków i produktów ewaluacji w odpowiednim momencie – gdy są najbardziej potrzebne użytkownikom do podejmowania decyzji.

Owen podkreśla również, że wykorzystanie ewaluacji jest kontekstowe – to, co działa w jednym środowisku, nie musi się sprawdzać w innym. W związku z tym kluczowe z perspektywy ewaluatorów dążących do maksymalizacji użycia są uważna analiza i dopasowanie danego badania do kontekstu, w którym ma być użyte. I tu raz jeszcze powraca waga interaktywności i elastyczności.

Owen idzie też o krok dalej i formułuje, opierając się na swoim doświadczeniu, główne reguły maksymalizacji wykorzystania ewaluacji. Według niego wykorzystanie wyników ewaluacji będzie tym bardziej prawdopodobne i większe (Owen, 2007: 121–122):

- im więcej ewaluator będzie konsultował się z użytkownikami podczas fazy planowania,
- im mocniej ewaluator skoncentruje się na odpowiedzi na pytania kluczowe dla odbiorców,
- im bardziej interaktywny jest proces komunikacji,
- im mniej zróżnicowana jest grupa odbiorców raportu,
- im bliżej ewaluator współpracuje z odbiorcami podczas badania,
- im więcej pomocy ewaluator zaoferuje podczas procesu wdrażania rekomendacji.

Autor zauważa też ironicznie, że ewaluator myślący, iż jego badanie będzie miało wyłączny wpływ na zmianę programu, zwyczajnie cierpi na urojenia.

2.3.3. Ewaluacja w kontekście teorii efektywnego uczenia się

Rola interakcji i dialogu w procesie badania ewaluacyjnego łączy dyskusję nad czynnikami użycia ewaluacji z szerszą dyskusją o determinantach efektywnego uczenia się (jednostek i organizacji) i zarządzania zmianą[29]. Torres i Preskill (2001) utrzymują, że nowoczesne badania ewaluacyjne koncentrują się na procesie uczenia się – jednostek, zespołów, organizacji. Warunkiem efektywnej nauki jest współpraca, dialog i zorientowanie na działania – ludzie i organizacje uczą się szybciej poprzez działania (*learning by doing*) (Preskill, Torres, 2000: 25). Leviton (2003) charakteryzuje użycie ewaluacji jako proces „produkcji wiedzy".

Dyskusja ta w dużej mierze powtarza czynniki przedstawiane w poprzednich akapitach. Jednak stosując optykę „ewaluacji jako procesu efektywnego uczenia" łatwiej jest uporządkować elementy wokół trzech głównych aspektów tegoż:

- warunków kontekstowych, w jakich miałaby zachodzić efektywna nauka i zmiana,
- cech, roli i relacji między uczestnikami procesu uczenia się,
- cech przekazanej wiedzy.

Warunki kontekstowe procesu uczenia się

Alkin i Taut (2003) jako warunki kontekstowe (ang. *context factors*) wymieniają dość wąskie cechy: ramy prawne i ograniczenia finansowo-organizacyjne dotyczące danego kontraktu badawczego, wcześniejsze badania, stabilność środowiska, w którym ma być prowadzona ewaluacja. Ray Rist (Leeuwet al., 1994 za Stern 2005) podchodzi do tego szerzej, pisząc o wpasowaniu się badania w cykl polityki (programu)[30]. Balthasar (2006: 355–356) zwraca szczególną uwagę właśnie na ten czynnik – tzw. szansę badania (ang. *window of opportunity*) – właściwy politycznie i organizacyjnie moment, w który wpiszą się wyniki ewaluacji.

Do tego warto dodać szerszy kontekst kultury ewaluacyjnej w danej administracji, a więc tradycja ewaluacji i zastosowania badań społecznych w zarządzaniu publicznym, responsywność, otwartość na zmiany, elastyczność (szerzej o tym w rozdziale na temat ewaluacji w administracjach krajów UE). Valovirta (2002: 75) opisuje natomiast kontekst ewaluacji w dwóch wymiarach: konsensusu lub konfliktu wokół programu oraz presji do zmian lub braku zainteresowania zmianami w programie.

---

[29] Konkretnie chodzi o następujący ciąg złożeń: a) im bardziej interaktywny jest proces analizy i dochodzenia do wniosków badania oraz b) im częściej interesariusze angażowani są aktywnie w proces analizy i oceny, tym łatwiej utożsamiać się będą się z wnioskami i szybciej będą się uczyć.

[30] Poza tym warunkiem podają jeszcze dwa: sposób rozpowszechniania i filtrowania informacji w danej organizacji (uznany przeze mnie jako cechy uczestników) oraz wiarygodność źródła (dyskutowany przeze mnie przy okazji cech przekazywanej wiedzy).

Charakterystyka i relacje uczestników procesu

Tak jak sugeruje tytuł, dyskusja wokół tego aspektu koncentruje się na dwóch elementach – cechach uczestników i relacjach między nimi.

Cechy uczestników określa się zwykle mianem czynnika ludzkiego (ang. *human factors*), do którego zaliczana jest charakterystyka ewaluatorów (kim są, jakie mają doświadczenie), ale przede wszystkim cechy użytkowników ewaluacji (ich zainteresowanie samą ewaluacją, determinacja w użyciu jej wyników, układ sił, podział kompetencji w organizacji, wiedza jednostek) (Alkin, Taut, 2003). Boyle i Lemaire (1998 za Stern, 2005) twierdzą, że odbiorcy ewaluacji, aby efektywnie się uczyć, powinni być przygotowani do odbioru i użycia wiedzy (być otwartymi, mieć świadomość czym jest ewaluacja, mieć sprecyzowane oczekiwania). Ważną cechą odbiorców ewaluacji jest to, w jaki sposób postrzegają program na wstępie (ang. *user knowledge construction*) – jakie mają stereotypy, utwierdzone opinie, uświadomione bądź nieuświadomione założenia (ang. *flawed assumptions*). Aspekt ten wiąże się z kontekstem badania, na przykład z kwestią, w jaki subiektywny sposób uczestnicy postrzegają kontekst programu[31]. Nutley zauważa, że najważniejszym jest to, na ile skutecznie dane badanie kwestionuje te założenia i utarte poglądy odbiorców (Nutley et al., 2003). Kolejną kwestią są potrzeby informacyjne odbiorców, omawiane już wcześniej, tu przedstawiane jednak jako poziom „niepewności" (ang. *uncertainity*) (Leviton, 2003). Skuteczna ewaluacja z jednej strony redukuje poziom niepewności i niezgodności (konflikty) wynikający z niewystarczających informacji, z drugiej zaś przełamuje stereotypy, dostarcza dowodów potwierdzających bądź obalających powszechne przypuszczenia, poszerzając tym samym zakres rozumienia interwencji.

Drugim szerokim tematem są relacje między uczestnikami procesu uczenia się. Zwykle przedstawiane są w ujęciu dynamicznym jako proces interakcji ewaluator-interesariusze i często łączą się z dyskusją nad podejściami badawczymi (interaktywnością procesów, zastosowanymi technikami poznawania i uczenia się) (Cousins, Shulha, 2006).

Wprowadzanie nowej wiedzy, często przełamującej utarte poglądy użytkowników, może odbywać się jedynie w atmosferze zaufania (Nutley et al., 2003). Liczy się więc zmniejszenie dystansu pomiędzy ewaluatorem a badanymi. Obecnie powszechnie przyjmuje się, że zbudowanie zaufania i poczucia współwłasności wokół ewaluacji oraz wysoka interaktywność procesu gwarantują dużo wyższy stopień użycia wyników[32].

---

[31] Leviton (2003: 528–29) proponuje, by identyfikować założenia interesariuszy (dotyczące programu, sposobu działania interwencji, tego, jak postrzegają innych aktorów) za pomocą wywiadów i ankiet, a następnie ujmować je w formie map mentalnych (w oparciu o: Morgan et al., 2002). Dysponując takimi mapami mentalnymi różnych interesariuszy, ewaluatorzy mogliby szybko zidentyfikować różnice zdań, niejasności, błędne czy niepoparte dowodami wyobrażenia, stereotypy i ukierunkować badanie na te niedobory wiedzy.

[32] Warto zwrócić uwagę, że jest to potwierdzenie – znanej od wieków i z powodzeniem wykorzystywanej w edukacji – mądrości płynącej ze spostrzeżenia Konfucjusza: „Powiedz mi, a zapomnę, pokaż, a zapamiętam, zaangażuj mnie, a zrozumiem".

Podstawowe relacje „producentów wiedzy" i jej odbiorców mogą mieć trzy formy (nie tylko w kontekście ewaluacji, ale także badań stosowanych): nacisk ze strony użytkowników na uczenie się (ang. *user-pull*), nacisk twórców wiedzy (ang. *producers-push*), relacja interaktywna na zasadzie wymiany (Lavis et al., 2003). Te trzy rodzaje słabo opisują jednak specyfikę ewaluacji, gdyż składa się ona z wielu faz i w każdej z nich role i relacje mogą być nieco odmienne.

Rolę uczących – a więc ewaluatorów – podsumowuje Morabito (2002), identyfikując cztery możliwe strategie: edukatora, konsultanta, animatora i doradcy. We wszystkich kluczowym elementem jest zbudowanie klimatu zaufania. Natomiast Valovirta opisuje rolę ewaluatora przez wspomniany już pryzmat dwóch wymiarów kontekstowych (konsensusu-konfliktu, presji do zmiany lub jej braku) i w rezultacie wyróżnia cztery możliwe role ewaluatora (2002: 75): mediatora (konflikt i brak presji zmian), arbitra rozstrzygającego konflikt (gdy ma miejsce konflikt i silna presja zmian), eksperta budzącego świadomość (konsensus i brak potrzeby zmian), reformatora (konsensus i potrzeba zmian)[33].

Podsumowując rozważania dotyczące aspektu ról i relacji, należy zauważyć, że mogą być one różne w kolejnych fazach ewaluacji. I tak w fazie strukturalizacji, na początku badania, gdy precyzowane są pytania badawcze i podejście teoretyczne, a także gdy nawiązywane są kontakty z interesariuszami, wyróżnić można trzy rodzaje relacji. Pierwsza reprezentuje podejście sztywne (ang. *rigid approch*) – gdy zlecający w sposób szczegółowy, często już na etapie przetargu, określa plan badawczy i metodologię. Ewaluator nawiązuje kontakt tylko ze wskazanymi przez zleceniodawcę interesariuszami, a jego rola ogranicza się do uszczegółowienia kwestii czysto technicznych (wskazanie źródeł danych, etc.). Podejście drugie, statyczne (ang. *static approach*), ma miejsce, gdy ewaluator tylko na bazie wymogów przetargu (bez dalszych konsultacji) uszczegóławia plan badania i prezentuje gotowy pakiet zlecającemu, w formie raportu przedstawianego na końcu fazy strukturalizacji. I tu nawiązanie kontaktów z interesariuszami ograniczone jest do minimum. Trzecie to podejście interaktywne (ang. *interactive approach*), które zakłada dyskusję i negocjacje podczas tworzenia modelu badawczego między ewaluatorami a zlecającymi (np. negocjacje nad priorytetowymi pytaniami, konsultacje w sprawie podejścia badawczego, etc.). Dyskusję ze zlecającym i ewentualnie innymi interesariuszami prowadzą i animują ewaluatorzy. Oni też mają decydujący głos w doborze narzędzi. Co więcej, oni identyfikują głównych interesariuszy, kluczowych aktorów i budują z nimi sieć powiązań.

Natomiast w kolejnych fazach (obserwacji, analizy i oceny) relacje mogą mieć następujące formy. Po pierwsze, może być to izolacja (ang. *isolation*) – gdy ewaluator wykonuje swoje badania, ograniczając kontakt ze zlecającym do okresowego raportowania, zaś z badanymi – do zbierania danych (ankiety, wywiady). Po drugie, może być to ograniczony kontakt (ang. *limited conta-*

[33] Angielskie terminy to: *conciliator/legitimizer, referee/confict manager, revisor/awakener, reform agent*.

cts), gdy relacje są w dużej mierze jednostronne, ograniczane przez ewaluatora do niezbędnego minimum (np. prezentacja wstępnego raportu i rekomendacji na warsztacie beneficjentów). Po trzecie, mogą być to interakcje (ang. *interactions*), gdy proces jest dwukierunkowy – ewaluatorzy rozmawiają z interesariuszami i zlecającymi, ci ostatni kontaktują się z ewaluatorami i współpracują z nimi, na co dzień są włączani tak w analizę, jak i w ocenę. W przypadku fazy obserwacji może wystąpić jeszcze szeroka interakcja (*wider networking*) – gdy ewaluatorzy nawiążą kontakty europejskie z badaczami analogicznych programów, szerszą siecią tematyczną danej polityki.

Rozległość i intensywność kontaktów (zbudowane zaufanie) ma znaczenie w momencie zakończenia badania, gdyż usprawnia proces wdrażania rekomendacji. Możliwe trzy typy to nawiązane kontakty ze zlecającym, z interesariuszami, z szerszym środowiskiem (spoza programu). Definiując stopień ich intensywności, można wyróżnić kontakty ze zlecającym – formalne i ograniczone (są nietrwałe, ustają w momencie oddania raportu, nie mają wpływu na proces wdrażania) lub dobre (raport zyskuje promotorów). Przy kontaktach z interesariuszami: takie, które nie przyniosą efektów, kontakty, które dadzą trwałą sieć współpracy i wymiany informacji między interesariuszami, albo nawet kontakty, które owocują szerszą koalicją wspierającą rekomendacje badania. Wreszcie można, lub też nie, nawiązać kontakty europejskie (możliwość promocji raportu, metod, wniosków na szerszym, europejskim forum – np. europejskich sieci tematycznych).

Uwypuklając natomiast rolę ewaluatorów w procesie uczenia się, można wyróżnić trzy role. Pierwsza to rola ewaluatora jako naukowca – w tym wypadku nie ma żadnych interakcji, a zlecający i inni interesariusze są traktowani wyłącznie jako przedmiot badania. Ewaluator bada i ocenia program *ex cathedra* – z perspektywy naukowej, bazując wyłącznie na „obiektywnych" danych. W tej roli ewaluator ma tendencje do odnoszenia programu do szerszych teorii i procesów. Akcent w badaniu jest stawiany na warsztat naukowy. W przypadku rozważania kontekstu programu (konfliktów i presji zmian wokół programu) taka rola dobrze współgra, z uwagi na posiadany autorytet (ale i brak elastyczności), z rolą arbitra lub eksperta.

Rola druga to ewaluator jako konsultant. Ogranicza się on do wytycznych kontraktowych (jego klientem, głównym adresatem i punktem odniesienia podejmowanych działań jest więc zlecający). Metodologia również zależy od oczekiwań klienta – główny akcent jest stawiany na koszty metod w porównaniu do ich efektów i szybkości użycia, a nie ich naukową poprawność i stopień skomplikowania. Gdy konieczne jest uwzględnienie kontekstu programu (konflikt i presja zmian) ta rola łączy się dobrze z rolą mediatora i eksperta (w zależności od posiadanych kompetencji i doświadczenia).

Wreszcie rola animatora-edukatora wymaga od ewaluatora otwartości na różne punkty widzenia, ciągłych interakcji, dyplomatycznego uwzględniania tak opinii zlecającego, jak i innych interesariuszy programu. Ewaluator używa metod interaktywnych, tworząc komfortowe warunki współuczestnictwa aktorów programu w procesie ewaluacji. W kontekście konfliktów i presji wokół programu ta rola łączy się dobrze z rolą mediatora i reformatora, gdyż ma dobre narzędzia budowy zaufania i dialogu.

Cechy przekazywanej wiedzy

Wreszcie trzeci, ostatni aspekt ewaluacji jako procesu uczenia się to cechy przekazywanej wiedzy. I tu dyskusja koncentruje się wokół zakresu tematycznego oferowanej wiedzy i jej jakości, a konkretnie jakości produktu badania – raportu ewaluacyjnego.

W przypadku zakresu tematycznego wiedzy Nutley i współpracownicy (2003: 128) (za: Ekblom, 2002) proponują ciekawą i prostą kategoryzację typów wiedzy (odnoszą to przy tym nie tylko do ewaluacji, ale również do badań stosowanych): *Know-about problems* – wiedza o problemach i procesach społecznych, *Know-what works* – wiedza o tym, które rozwiązania przynoszą oczekiwane efekty, *Know-how* – wiedza o tym, w jaki sposób efektywnie wdrożyć te działania, *Know-who* – czyli kogo włączyć w proces oraz *Know-why* – szersza wiedza i świadomość, dlaczego dane działania są niezbędne, jak odnoszą się do zasad, norm teorii.

Należy jednak podkreślić, że ta kategoryzacja ma nieco mniej egzotyczne źródła (publikacja Ekbloma dotyczy transferu *know-how* w zapobieganiu przestępczości), gdyż jej korzeni można się doszukać w literaturze na temat „gospodarki uczącej się" (ang. *learning economy*) – klasycznym już podziale zaproponowanym przez Lundvalla i Johnsona (Lundvall, Johnson, 1994 cytowane w: Widmaier, 2000: 42–43)[34]. W kontekście ewaluacji należy dodać jeszcze do tych kategorii moment (ang. *know-when*) – czyli wiedzę o tym, kiedy działać, w jakiej perspektywie czasowej.

Oferowaną wiedzę można określić pod kątem jej odkrywczości – świeżości dla odbiorców (kwestie nowe lub nowe spojrzenie na znane już problemy i zagadnienia danej interwencji) lub trywialności (kwestie powszechnie znane i wiadome dla osób zaangażowanych w interwencję).

Co do jakości ewaluacji, standardy „dobrej ewaluacji" bazują w dużej mierze na standardach naukowych. Najczęściej jako kluczowe podawanych jest pięć kryteriów[35]:

- Trafność wewnętrzna (ang. *internal validity*) odnosi się do poprawności projektu ewaluacji, określa, na ile relacja między badanymi czynnikami (interwencją a jej efektami) jest przyczynowo-skutkowa, a na ile może być przypisana czynnikom zewnętrznym – czyli na ile zmienna niezależna, a nie inne czynniki, są odpowiedzialne za przekształcenia zmiennej zależnej (Berk, Rossi, 1999: 19–22; Bingham, Felbinger, 2002: 21; Frankfort-Nachmias, Nachmias, 2001: 611; Langbein, Felbinger, 2006: 34; Shadish et al., 2001: 37);

[34] Kategoryzacja Lundvalla i Johnsona (1994) odnosi się do tworzenia informacji i wiedzy jako zasobu rynkowego. Wyróżniają oni *know-what* – wiedzę powszechnie dostępną, informacje, *know-why* – wiedzę skodyfikowaną, głównie o charakterze naukowym, *know-how* – wiedzę nieformalną, nieskodyfikowaną, ang. *tacit knowledge*, związaną z konkretnymi osobami i ich doświadczeniem, oraz *know-who* – wiedzę o tym, kto co wie, kto posiada informacje, z kim szukać kontaktu.

[35] Tłumaczenie terminów angielskich za polskim wydaniem (Frankfort-Nachmias, Nachmias, 2001: 605–611).

- Trafność zewnętrzna (ang. *external validity*) wiąże się z poprawnością projektu ewaluacji i określa, do jakiego stopnia obserwacje danej relacji z danego badania mogą być generalizowane na inne osoby, czynniki, programy, populacje, lokacje (Berk, Rossi, 1999: 22–26; Bingham, Felbinger, 2002: 25; Frankfort-Nachmias, Nachmias, 2001: 611; Langbein, Felbinger, 2006: 34; Shadish et al., 2001: 38)[36];
- Trafność teoretyczna (ang. *construct validity*) odnosi się do poprawności pomiaru i określa, na ile wykorzystywane narzędzie pomiarowe mierzy badany koncept (czy jest ono właściwie powiązane z wykorzystywanymi koncepcjami, teoriami) (Berk, Rossi, 1999: 16; Bingham, Felbinger, 2002: 37; Frankfort-Nachmias, Nachmias, 2001: 611; Shadish et al., 2001: 506)[37];
- Rzetelność (ang. *reliability*) – to kryterium odnosi się do precyzyjności pomiaru, określa, na ile wyniki i obserwacje otrzymane przy zastosowaniu danego narzędzia pomiaru są powtarzalne, na ile użyte narzędzie gwarantuje powtarzalność wyników i obserwacji (Bingham, Felbinger, 2002: 38; Langbein, Felbinger, 2006: 35; Rossi et al., 1999: 447);
- Trafność konkluzji statystycznych (ang. *statistical conclusion validity*) to poprawność wnioskowania z zaobserwowanej kowariancji między czynnikami (Berk, Rossi, 1999: 26–27; Shadish et al., 2001: 42).

Każde z czterech kryteriów jest podatne na szereg zagrożeń, które powinny być minimalizowane przy zastosowaniu odpowiednich technik i zachowaniu dyscypliny badawczej[38].

Należy zaznaczyć, że powyższe kryteria są szczegółowo przestrzegane przede wszystkim w podejściach *quasi*-eksperymentalnych. Ewaluatorzy z innych nurtów (partycypatywnego, realistycznego) podchodzą do tych kwestii dużo mniej rygorystycznie. Pawson (2006: 181–182) zauważa, że te ścisłe standardy opierają się na podstawowym założeniu przejrzystości i powtarzalności procedury przeniesionym z nauk przyrodniczych (standaryzacja procedury daje gwarancję uzyskania tych samych efektów – nie ma znaczenia jaki zespół ludzi wykona badanie, tak długo jak będzie stosować tę samą procedurę). Podnosi dwie uwagi krytyczne. Po pierwsze, stwierdza, że całkowita przejrzystość procesu ewaluacji jest niemożliwa, ponieważ nie sposób zapisywać (analogicznie do badań laboratoryjnych) każdą z wykonanych czynności – wstępnie odrzucone źródła danych, wszystkie, nawet pobieżnie przej-

[36] Rossi ze współpracownikami (1999: 447) opisując stopień, do jakiego wnioski z badania mogą zostać powtórzone przez innych badaczy, używa terminu powtarzalności (ang. *replicability*). Natomiast do określenia powtarzalności samych obserwacji – pomiarów – używa terminu rzetelności (ang. *reliability*).

[37] Rossi ze współpracownikami (1999: 445) oraz Langbein i Felbinger (2006: 35) zastępują ten termin terminem trafności pomiaru (ang. *measurement validity*). Frankfurt-Nachmias (2001: 605) wyróżnia jeszcze odpowiedniość (ang. *congruence*) – na ile istnieje zgodność między definicjami pojęciowymi a operacyjnymi. Zaś Bingham i Felbinger (2002: 37) dodają trafność kryterium (ang. *criterion validity*), to znaczy, na ile użyta miara mierząca dany koncept jest zbieżna z innymi, analogicznymi badaniami i miarami mierzącymi ten sam koncept.

[38] Szerzej o tych zagrożeniach i technikach ich minimalizacji patrz Bryman (2004: 36–37) oraz Shadish et al. (2001) o trafności wewnętrznej (s. 55–63), o trafności zewnętrznej (s. 86–93), o trafności teoretycznej (s. 73–81), o trafności konkluzji statystycznych (s. 45–53), zaś o rzetelności Bingham i Felbinger (2002: 38–39).

rzane dokumenty (często odrzucone na podstawie własnego doświadczenia, roztropności badawczej), pobieżnie konsultowane osoby, uwzględnić wcześniejsze doświadczenie wyniesione z poprzednich projektów. Po drugie, stwierdza, przenosząc dyskusję na grunt bardziej filozoficzny, że obiektywizm w nauce płynie nie ze standaryzacji, a z krytycyzmu, nie z powtarzania, a obalania teorii. Według niego ewaluator powinien jasno przedstawić cel, podejście, konstrukcję modelu i sposób rozumowania, tak by jego teoria mogła być poddana krytyce.

W praktyce badań ewaluacyjnych użytkownicy ewaluacji – decydenci – stosują do oceny jakości badania podwójny, intuicyjny test (Rossi et al. 1999: 435 za Weiss, Bucuvalas, 1980b):

- Test prawdy (ang. *truth tes*t) – „Czy badanie jest godne zaufania?", „Czy mogę na nim polegać?", „Czy zawiera mocne argumenty i dobrą podstawę do obrony?". Chodzi o jakość badawczą (czy badanie jest poprawne metodologicznie, trzyma się standardów naukowych), na ile jest zgodne z oczekiwaniami i doświadczeniem użytkowników, a na ile kwestionuje istniejącą już wiedzę, podważa *status quo*. Oczywiście im bardziej kontrowersyjne wnioski, tym silniejsze musi być poparcie w warsztacie naukowym i przejrzystość metodologii;
- Test użyteczności (ang. *utility test*) – „Czy badanie daje mi jasne wskazówki?", „Czy ma rekomendacje działań lub alternatywnych podejść?", „Czy jest mi pomocne?". Konkretnie chodzi o praktyczną orientację działań (czy raport daje odbiorcy konkretne wskazówki prowadzące do realistycznych zmian) i czy istnieje związek między tematem badania a zakresem kompetencji danej osoby.

Warto tu jednak zauważyć, że gdy decydenci mają podjąć trudną decyzję, wolą opierać się na badaniach ilościowych, *quasi*-eksperymentalnych, na dużych próbach badawczych. Gdy jednak interesuje ich proces, chcą zrozumieć zachodzące zjawiska, sięgają po badania oparte na studiach przypadków, badania jakościowe, antropologiczne (Christie, 2007)[39]. Co ciekawe, w przypadku studiów, które podważają w zasadniczy sposób istniejące *status quo*, praktyczne rekomendacje są mniej ważne, liczy się wywód – przejście od analizy do wniosków (Owen, 2007).

W praktyce ewaluacji programów publicznych instytucje przyjmujące raporty wypracowały szablony (ang. *checklists*) weryfikacji jakości raportów ewaluacyjnych[40]. W przypadku oceny raportów z ewaluacji programów UE podejście jest bardzo techniczne. Komisja Europejska wypracowała szablon oceny raportu (przedstawiono go w poniższej tabeli), który jest powszechnie stosowany – przez wykonawców jako punkt odniesienia, zaś przez zlecających jako lista weryfikująca.

---

[39] Takie wybory korespondują wprost z „mitem racjonalności" współczesnych społeczeństw, o którym pisze Morgan (Morgan, 2006: 141). Liczby, podobnie jak kiedyś magia, pozwalają stworzyć ramy do interpretacji, wyjaśnić, oswoić i uporządkować świat.

[40] Na przykład amerykański Joint Committee on Standards for Educational Evaluation zatwierdził listę 30 kryteriów jakości ewaluacji programów. Są one pogrupowane w czterech kategoriach: *utility, feasibility, propriety, accuracy*, porównaj: http://www.wmich.edu/evalctr/jc/.

Tabela 5. Formularz oceny raportu ewaluacyjnego

Kwestia	Nie do przyjęcia	Słaby	Dobry	Bardzo dobry
Spełnione oczekiwania: Czy ewaluacja odpowiada na wszystkie oczekiwania informacyjne zlecającego? Czy odpowiada wymaganiom sprecyzowanym w Zakresie Zadań (SIWZ)?				
Właściwy zakres: Czy racjonalność programu, jego produkty, rezultaty, oddziaływania, interakcje z innymi politykami i nieoczekiwane efekty zostały objęte analizą?				
Odpowiedni plan badawczy: Czy wybrane podejście badawcze i użyte narzędzia były właściwe do postawionych pytań (uwzględniając istniejące ograniczenia metodologiczne)?				
Poprawność doboru danych: Czy użyte dane pierwotne i wtórne były odpowiednie do przedmiotu analizy? Czy są wystarczająco wiarygodne?				
Solidna analiza: Czy informacje jakościowe i ilościowe były poprawnie analizowane, zgodnie z bieżącą metodologiczną wiedzą i standardami? Czy były kompletne i wystarczające, by odpowiedzieć na zadane pytania?				
Wiarygodne wyniki: Czy wyniki badania są logiczne, wypływają z przeprowadzonej analizy i mają poparcie w zebranych danych oraz ich interpretacji opartej na sformułowanych hipotezach wyjaśniających?				
Ważne konkluzje: Czy raport dostarcza jasnych konkluzji? Czy są one poparte wiarygodnymi dowodami?				
Bezstronne i użyteczne rekomendacje: Czy rekomendacje są uczciwe, niezniekształcone przez osobiste uprzedzenia zlecającego lub któregokolwiek z operatorów programu? Czy są wystarczająco szczegółowe, by móc je wdrożyć?				
Przejrzysty raport: Czy raport opisuje kontekst i cele, organizację i rezultaty analizowanego programu w taki sposób, że przedstawiane informacje są łatwe do zrozumienia i przystępne?				
Biorąc pod uwagę ograniczenia kontekstowe badania ewaluacyjnego, ogólnie raport można uznać jako:				

Źródło: (European Commission, 1999b: vol. 1 s. 180; 2008).

Jak widać w powyższej tabeli, kryteria wybrane przez Komisję łączą część standardów naukowych (poprawność doboru danych, obiektywizm analizy, kryteria trafności wewnętrznej, konkluzji statystycznej) z praktycznymi wymogami (przystępność raportu, utylitaryzm badania).

Podsumowując i łącząc dotychczasowe rozważania na temat cech przekazywanej wiedzy, można podzielić je na dwie grupy:

Pierwszym są cechy raportu ewaluacyjnego – jego jakość. Tu mieściłyby się zarówno kwestie poprawności metodycznej raportu (test prawdy), jak i jakość jego strony technicznej, konkretnie zaś:

- Poprawny zakres badania – zbieżność ze zleconym zakresem zadania, całościowe podejście do programu i jego efektów;
- Poprawny plan badawczy (w tym kwestie trafności wewnętrznej i trafności teoretycznej zrozumienia specyfiki programu);
- Poprawność doboru i użycia danych (w tym rzetelność i triangulacja);
- Przejrzystość i poprawność analizy (w tym trafność konkluzji statystycznych);
- Wiarygodność wyników – poparcie w danych;
- Jakość rekomendacji – ich bezstronność, konkretność, operacyjność;
- Przystępność i przejrzystość raportu.

Drugim aspektem jest charakter przekazanej wiedzy zawartej w raporcie, czyli jego wniosków i rekomendacji (ang. *utility test*). Kluczowym jest tu kwestia, na ile jest to wiedza nowa dla użytkowników. Mogą oni otrzymać różne rodzaje wiedzy:

- *Know-about* (wiedzieć o…) – raport dostarcza nowej wiedzy o problemach, wyzwaniach, konfliktach, potencjalnych zagrożeniach, ale i szansach, które pojawiają się w szerszym kontekście, otoczeniu programu;
- *Know-what* (wiedzieć co…) – raport zawiera informacje, co sprawnie funkcjonuje, które działania i rozwiązania przynoszą oczekiwane efekty lub nieoczekiwane sukcesy, a które są opóźnione, sprawiają problem, nie działają;
- *Know-why* (wiedzieć dlaczego…) – raport opisuje i wyjaśnia źródła oraz przyczyny tak problemów, jak i sukcesów, interpretuje dane procesy, odnosi program do szerszej wiedzy z danej tematyki, ukazuje go na tle teorii, norm, założeń polityk, trendów międzynarodowych;
- *Know-how & when* (wiedzieć jak i kiedy…) – raport zawiera konkretną wiedzę praktyczną, w jaki sposób i w jakim momencie poprawić lub zmienić interwencję, efektywnie wdrożyć usprawnienia, może również zawierać alternatywne rozwiązania lub przykłady *know-how* z innych programów;
- *Know-who* (wiedzieć kto…) – raport wskazuje, kto (jakie instytucje, aktorzy) powinien być odpowiedzialny za podjęcie danych działań naprawczych lub zaradczych oraz kogo należy lub warto zaangażować w dane procesy.

2.3.4. Próby ujęcia systemowego

Kompleksowość tematu wykorzystania ewaluacji i czynników ją warunkujących sprawiły, że w badaniach wykorzystania ewaluacji pojawiła się nowa tendencja. Wiąże się ona z próbami zintegrowanego i modelowego ujęcia tematu, takimi jak identyfikowanie i łączenie czynników wpływających na wykorzystanie ewaluacji z różnymi rodzajami jej użycia (ang. *integrated theory of use*) (Stern, 2005: 10).

Pierwszą kompleksową próbę tego rodzaju podjęli Cousins i Leithwood (1986). Przeanalizowali oni sześćdziesiąt pięć studiów (z lat 1971–1985) na te-

mat wykorzystania ewaluacji w programach z zakresu edukacji, zdrowia psychicznego i usług publicznych. W każdym przypadku opisali zastosowaną metodologię, zidentyfikowane zmienne zależne i niezależne, założony kierunek relacji. Na tej podstawie stworzyli model (patrz: Aneks 6.2, schemat 1). Umieścili w nim dwanaście czynników warunkujących użycie ewaluacji (zmiennych niezależnych). Połowa z nich związanych jest z charakterystyką samej ewaluacji, druga zaś z procesem decyzyjnym i kontekstem polityki/programu. Jako zmienne zależne autorzy wyróżnili zaś trzy typy użycia – podjęcie decyzji (ang. *decision*), uczenie się (ang. *education*) oraz zmiany w osobistym postrzeganiu sytuacji programu przez decydentów (ang. *processing*)[41]. Analizując wnioski z sześćdziesięciu pięciu studiów, autorzy jako dominujące czynniki dla użycia decyzyjnego i osobistego zidentyfikowali w kolejności:

1. jakość ewaluacji (obejmująca zastosowane podejście badawcze, stopień wyrafinowania metod, rygor metodologiczny, charakterystykę procesu);
2. wnioski z ewaluacji (ich odkrywczość, przystawalność do powszechnych opinii lub ładunek krytyczny, jaki ze sobą niosły);
3. zaangażowanie użytkowników – ich nastawienie i aktywizowanie ich przez ewaluatorów (ang. *commitement/receptiveness*);
4. przystawalność ewaluacji (zarówno do potrzeb informacyjnych, jak i do instytucjonalne go umiejscowienia ewaluatora).

Natomiast w przypadku „uczenia się z ewaluacji" kolejność rang była częściowo odwrócona. Kluczowe okazały się jakość ewaluacji i nastawienie – otwartość użytkowników, zaś kolejne, równoprawne trzy czynniki to klimat polityczny (orientacja polityczna zlecających, uzależnienie od zewnętrznych sponsorów, rywalizacja wewnątrz lub między instytucjami, walki o budżet), konkurencyjne informacje (ang. *competing information* – brane dosłownie jako źródła podające wnioski stojące w sprzeczności z wnioskami z ewaluacji) i cechy osobiste (pozycję głównych decydentów w organizacji, doświadczenie organizacyjne, charakterystykę społeczną etc.).

Niewątpliwą zaletą modelu przedstawionego przez Cousins i Leithwood jest, po pierwsze, powiązanie czynników wpływających na ewaluację z typami wpływu, po drugie zaś określenie siły czy raczej rangi tych zależności. Mocną stroną jest również praktyczny charakter modelu – poparcie wniosków w badaniach empirycznych. Model dotyczył głównie studiów ewaluacji programów edukacyjnych, które stosowały metody *quasi*-eksperymentalne. Dzięki temu w analizie tych studiów można było zastosować metody ilościowe oraz analizę statystyczną korelacji między czynnikami. Ciekawym w omawianym modelu jest też wyróżnienie w „użyciu koncepcyjnym" aspektu jednostkowego i organizacyjnego.

Za główne ograniczenia modelu uznać można natomiast dwie sprawy. Po pierwsze, definicje zidentyfikowanych czynników wpływu są bardzo szerokie i niespójne, co wynika z łączenia różnych podejść użytych w sześćdziecięciu pięciu studiach. Na przykład kategoria *relevance* zawiera zarówno kwestie ja-

[41] Odnosząc te trzy typy użycia do typów wykorzystania ewaluacji dyskutowanych wcześniej, „podjęcie decyzji" odpowiada instrumentalnemu użyciu ewaluacji, „uczenie się" jest rodzajem koncepcyjnego użycia w skali organizacji, zaś „zmiany w osobistym postrzeganiu sytuacji programu przez decydentów" to koncepcyjne użycie na poziomie jednostek.

kościowe zależne od ewaluatora (zakres meritum ewaluacji), jak i kwestie instytucjonalne (umiejscowienie ewaluatora – zewnętrzny vs. wewnętrzny). Drugim ograniczeniem jest brak perspektywy czasowej – sekwencji zdarzeń, czyli próby określenia, kiedy w grę wchodzą poszczególne czynniki. Z perspektywy niniejszej pracy, słabością jest też fakt, że *gros* z wykorzystanych studiów dotyczyło programów edukacji.

Ten wyjściowy model był następnie rozwijany przez jego autorów. W publikacji z 1993 roku (Cousins, Leithwood, 1993), bazując na dotychczasowych kategoriach zmiennych zależnych i niezależnych, zaprezentowali oni nowy model (patrz: Aneks 6.2, schemat 2). Kluczową różnicą jest w nim wydzielenie elementów interakcji i procesu (takich jak zaangażowanie, ciągłość kontaktów, socjalne przyswajanie informacji) jako oddzielnego etapu. Nowy model miał już więc wyraźną sekwencję zdarzeń. Ponadto w opisie silniej podkreślono użycie koncepcyjne na poziomie jednostki (zmiany w myśleniu decydentów) jako etap owocujący decyzjami lub szerszym procesem uczenia się. Należy jednak podkreślić, że omawiany model z 1993 roku nie odnosił się już tylko do tematu wykorzystania ewaluacji, ale także do strategii uczenia się jednostek i organizacji. Dotyczy to również specyficznego przypadku – udoskonaleń wprowadzanych w programach szkolnych na poziomie szkół podstawowych (studium opierało się na szerokim badaniu ankietowym 233 nauczycieli i 155 pracowników kuratorium).

Kolejnym modelem jest „teoretyczny model wykorzystania ewaluacji" wypracowany przez Burke'a Johnsona (1998) (patrz: Aneks 6.2, schemat 3). Zastosował on wypracowane przez siebie podejście metamodelowe mające charakter indukcyjny i teoretyczny – polega ono na budowaniu modelu na podstawie syntezy wcześniejszych modeli, wniosków z badań i koncepcji funkcjonujących w literaturze tematu. Johnson oparł się na ponad osiemnastu wcześniejszych modelach – część z nich domniemanych (stworzonych przez niego w oparciu o przegląd publikacji i koncepcji danych autorów), część zaś jawnych (schematów zaczerpniętych z literatury przedmiotu). Część z modeli wyjściowych użytych przez Johnsona bazowała na badaniach ilościowych, inne były wynikiem badań jakościowych, jeszcze inne miały zaś charakter czysto teoretyczny i spekulatywny (Johnson, 1998: 94–95). Łącząc te różne koncepcje, zidentyfikowane grupy czynników i kierunki relacji Johnson uzupełniał je o dodatkowe elementy, które uznał za brakujące. Główna „fabuła" modelu jest następująca:

„Wykorzystanie ewaluacji to ciągły i szeroki proces, który jest uzależniony od lokalnych czynników kontekstowych, organizacyjnych i politycznych. Kluczowymi czynnikami są uczestnictwo interesariuszy, szeroki proces rozpowszechniania wyników i informacji, dobra dwukierunkowa komunikacja między ewaluatorami a użytkownikami tak przed, jak i po programie. Te kwestie gwarantują lepsze użycie ewaluacji także poprzez zapewnienie jej lepszego przystawania do potrzeb programu oraz poprzez zbudowanie poczucia współodpowiedzialności wśród interesariuszy" (Johnson, 1998: 104). Natomiast praktyczny wniosek płynący z modelu jest taki, iż ewaluatorzy we współpracy z zarządzającymi programem i z kluczowymi interesariuszami powinni stoso-

wać podejścia i metody organizacyjne, które zintensyfikują współpracę i proces uczenia się.

Zaletą omawianego modelu jest jego eklektyczność – łączy on wiele źródeł i koncepcji (właściwie uwzględnia wszystkich kluczowych teoretyków ewaluacji i główne badania empiryczne z ostatnich piętnastu lat). Ponadto w założeniu przystaje on do różnych typów ewaluacji (zewnętrznej/wewnętrznej, formatywnej/podsumowującej). Lista wyróżnionych czynników wpływających na wykorzystanie ewaluacji jest bardzo szeroka. Szczególną uwagę zwraca wyróżnienie organizacyjnej, indywidualnej charakterystyki użytkowników oraz charakterystyki ewaluatora. Ponadto ciekawym, porządkującym podziałem jest wyróżnienie grupy czynników bazowych/kontekstowych (ang. *background variables*), procesowych (ang. *interactional variables*) oraz czynników wykorzystania (ang. *utilization variables*). Wreszcie bardzo użytecznym jest wprowadzenie dwóch koncepcji – sieci interakcji tzw. organizacji sieciowej (ang. *network organisation*) oraz uczestnictwa (ang. *participation*). Nie zostały one jednak uwzględnione w graficznej stronie modelu – są obecne wyłącznie w opisie. W koncepcji „organizacji sieciowej" ewaluator zewnętrzny wchodzi ze swoim badaniem w układ zależności i interakcji wewnątrz i międzyorganizacyjnych. Wokół prowadzonego badania tworzy się więc system społeczny – interesów, interakcji, osób. Ewaluator jest ogniwem tej sieci i powinien właściwie sterować tymi powiązaniami, zwracając szczególną uwagę na osoby kluczowe – te, które mogą nauczyć się najwięcej w procesie ewaluacji, potrzebują związanych z nią informacji, mają wpływ na kierunek programu. W przypadku drugiego pojęcia „uczestnictwa" nie jest ono *novum* w stosunku do poprzednich kategoryzacji i modeli. Johnson porządkuje jednak zależność między skalą uczestnictwa a czynnikami kontekstowymi. Po pierwsze uznaje, że uczestnictwo jest bardziej intensywne w organizacjach o strukturze organicznej, niż w tych o strukturze hierarchicznej – mechanistycznej. Po drugie, uczestnictwo jest częstsze w przypadku osób „uczących się", otwartych w porównaniu do osób o typowo biurokratycznym usposobieniu. I wreszcie, im bardziej ewaluator jest zorientowany na praktykę (a mniej na czyste badania naukowe), tym silniejszy kładzie akcent na uczestnictwo.

Kompleksowość jest jednak również pewną słabością modelu. Trudne do uchwycenia są kierunki relacji (właściwie wszystkie elementy są przedstawione jako współgrające ze sobą), co utrudnia przyłożenie modelu do ewentualnej sytuacji rzeczywistego programu. Ponadto czynniki środowiskowe i interakcje ze środowiskiem zewnętrznym zostały dla programu uwzględnione w modelu tylko w małym stopniu (ang. *input from/output into external environment*), przy jednoczesnym podkreślaniu ich wagi w opisie. W samym modelu brak jest wielu kategorii (uznanych za autora jako kluczowe przy opisie modelu) – np. poziomu współuczestnictwa, kultury organizacyjnej czy omawianej już sieci interakcji. Zastanawia również brak całego segmentu rozważań teoretycznych i praktycznych badań dotyczących błędnego używania i nadużywania ewaluacji (które również często mogą mieć miejsce), jak i długookresowej perspektywy użycia (tzw. oddziaływania ewaluacji). To sprawia, że opisany proces wydaje się niepełny – pokazuje jedynie jedną z możliwych

ścieżek. Wreszcie same kategorie użycia (ang. *cognitive use & behavioral use*) są – choć teoretycznie interesujące – to przy ewentualnej próbie praktycznego zastosowania modelu mało operacyjne.

Najnowszym podejściem modelowym jest „Schemat teorii oddziaływania ewaluacji" wypracowany przez Malvina Marka i Gary'ego Henry'ego (2004). Jest to model teoretyczny, do pewnego stopnia metamodel, gdyż integruje wcześniejsze badania i typologie, jak również artykuły samych autorów. Pomysł Marka i Henry'ego bazuje silnie na „Modelu logiki ewaluacji" B. Cousinsa (2003) stosowanym do ewaluacji uczestniczącej. Cousins zastosował do przedstawienia logiki działania ewaluacji taki sam ciąg logiczny, jaki ewaluatorzy stosują do opisu i analizy logiki programów (tzw. CIPP model – *Context – Input – Process – Product* (szerzej patrz: Stufflebeam, 2005)[42]. Idąc za nim Mark i Henry wyróżnili cztery etapy procesu ewaluacji (patrz: Aneks 6.2, schemat 4):

- Wkład/nakłady ewaluacji (ang. *evaluation input*) obejmujące zarówno czynniki kontekstowe odnoszące się do ewaluacji (czas, zasoby, doświadczenie, role, wytyczne, zakładane kanały komunikacji), jak i kontekst decyzyjno-polityczny (wsparcie administracyjne, gry interesów, kultura organizacji, potrzeby informacyjne, etc.);
- Działania badawcze (ang. *evaluation activities*) do których przypisane zostały: wybór i włączanie interesariuszy, projektowanie ewaluacji, zbieranie danych i ich analiza, wypracowywanie konkluzji i rekomendacji, przygotowanie raportu, rozpowszechnianie informacji;
- Produkty badania (ang. *evaluation products*) opisujące cechy wiedzy wytworzonej w procesie ewaluacji, jej chłonność, wiarygodność, wyrafinowanie, komunikatywność, wpasowanie w moment decyzyjny);
- Wykorzystanie ewaluacji (ang. *evaluation utilization*) opisywane dwoma grupami kategorii – użyciem procesu (przez jednostki, grupy, organizacje) i użyciem wyników (koncepcyjnym, instrumentalnym lub symbolicznym).

W kolejnym kroku Mark i Henry czerpiąc z psychologii, teorii uczenia się i przyswajania zmian, skoncentrowali się na mechanizmach psychologiczno-społecznych (na trzech poziomach – jednostkowym, interpersonalnym, zbiorowym), które prowadzą do wykorzystywania ewaluacji. Znacznie rozbudowali model, adaptując swoje wcześniejsze prace z zakresu przyswajania wiedzy i zmiany (porównaj: Henry, Mark, 2003: 297–305):

- Poszerzyli segment produktów ewaluacji o listę ogólnych czynników motywujących (zarówno jednostki, jak i grupy);
- Segment wykorzystania ewaluacji przemianowali na średnio i długookresowe efekty ewaluacji i umieścili w nim cały zestaw mechanizmów i motywacji psychologiczno-społecznych, pogrupowanych na poznawcze, motywujące, behawioralne;
- Zwieńczyli cały proces logiczny ewaluacji celem nadrzędnym, jakim według nich jest przyczynianie się do dobrobytu społecznego;

[42] CIPP jest modelem amerykańskim. W odniesieniu do programów UE stosowany jest dłuższy ciąg logiczny: kontekst (*context*) – nakłady (*input*) – działania (*actions*) – produkty (*outputs*) – rezultaty (*results*) – oddziaływanie (*impact*).

- Wzbogacili model o czynniki środowiskowe – kwestie niezależne i zewnętrzne, takie jak istnienie konkurencyjnych w stosunku do danej ewaluacji informacji i badań, czynników sprzyjających lub blokujących ewaluację;
- Zaznaczyli w modelu wzajemne powiązania i interakcje elementów z poszczególnych etapów „produkcji ewaluacji".

Największą zaletą tych obydwu modeli jest pokazanie jasnej relacji przyczynowo-skutkowej oraz logicznej sekwencji zdarzeń. Ewaluacja i jej wykorzystanie zostało przedstawione jako proces – analogiczny do procesów wdrażania programów/projektów i obserwowania ich efektów. To otwiera możliwości łatwiejszego operacjonalizowania badań wykorzystania ewaluacji. Po drugie, oba modele (szczególnie drugi) wychodzą poza użycie ewaluacji, dają szerszą perspektywę, w jasny sposób włączają w analizę (znów w sekwencji czasowej) kwestie długoterminowego, opóźnionego użycia, oddziaływania na zewnętrzne polityki i programy. Co więcej, umieszczają ewaluację w szerszym kontekście jej naczelnej funkcji – definiowanej przez autorów jako przyczynianie się do dobrobytu społecznego. Kolejną zaletą jest zidentyfikowanie i opisanie mechanizmów psychologiczno-społecznych stojących za procesem akceptowania wiedzy i zmian.

Przedstawione kategorie motywacji wykorzystywania ewaluacji (czy raczej podatności na oddziaływanie ewaluacji) są wieloaspektowe (poziom jednostek, interpersonalny, organizacji) i skomplikowane. Te, które odnoszą się do aspektów psychologicznych jednostek, mogą sprawiać trudność w operacjonalizacji (na przykład *elaboration* – czyli jak często jednostka myśli o danej kwestii, czy *priming* – na ile ta kwestia jest uważana za kluczową i wpływa na postrzeganie i interpretowanie innych zdarzeń). Ważne choć trudno mierzalne są też takie elementy jak na przykład zmiany w postrzeganiu, proces nabywania nowych umiejętności w wyniku zetknięcia się z ewaluacją, psychospołeczny mechanizm skutecznej presji mniejszości na zmiany. Inne wyróżnione elementy, jak np. *change agent* – czyli obecność osoby zdeterminowanej do przeprowadzenia zmian, czy *agenda setting* – efektywne umieszczanie wyników z ewaluacji na wokandzie publicznej, są bardziej praktyczne[43].

W przypadku drugiego modelu próba pokazania wszystkich interakcji pionowych i poziomych między poszczególnymi sekwencjami wykorzystania ewaluacji zaciemnia obraz i utrudnia śledzenie ciągu logicznego, jak również formułowanie hipotez o relacjach przyczynowo-skutkowych.

Wydaje się więc, że drugi model ma małe szanse operacjonalizacji w całości i może być raczej traktowany jako koncepcyjna, kompleksowa mapa oddziaływania ewaluacji, z której wybierane będą elementy dalszych, empirycznych badań.

Analizując modele teoretyczne, warto się odnieść do empirii. Otóż badania praktyczne wykorzystania ewaluacji zostały zapoczątkowane w USA, na przełomie lat 60. i 70. (pierwsze publikacje Carol Weiss). Impulsem było roz-

[43] Szczegółowe opisy zidentyfikowanych mechanizmów wpływających na oddziaływanie ewaluacji zostały podsumowane w Mark i Henry (2004: 41), zaś scharakteryzowane w Henry i Mark (2003: 297–305).

czarowanie samych ewaluatorów słabszym niż się spodziewano przełożeniem wyników z ewaluacji na rzeczywistość programów publicznych[44]. Zainteresowanie tą kwestią, a z nią liczba publikacji, wzrosło szczególnie w latach 80.[45] Dotychczasowa amerykańska praktyka badań w tym polu jest bardzo bogata. Z uwagi na zakres tej pracy nie będzie tu szczegółowo omawiana. Warto tylko wspomnieć, że stosowane są bardzo różne podejścia i plany badawcze – zarówno ilościowe, jak i jakościowe, *quasi*-eksperymentalne, retrospektywne, porównań podłużnych, a także symulacje, studia przypadku czy nawet badania etnograficzno-antropologiczne (Cousins, Leithwood, 1986).

W Unii Europejskiej badania wykorzystania ewaluacji prezentowane w literaturze są incydentalne. Można je uznać za przyczynkowe, zważywszy na fakt, że w ostatnich dziesięciu latach pojawiły się tylko dwa takie badania. Obydwa dotyczą wykorzystania badań o ewaluacjach przez dyrekcje Komisji Europejskiej – właściwie można je uznać za kolejne edycje tego samego badania. W edycji pierwszej zastosowano wywiady i badania ankietowe z pracownikami Komisji, studia przypadku, a także jako punkt wyjścia przegląd literatury na temat użycia ewaluacji (Williamset al., 2002: 1). W drugiej edycji badania sięgnięto po analogiczne narzędzia: badania ankietowe personelu poszczególnych dyrekcji, śledzenie użycia (analiza cytatów i treści) raportów ewaluacyjnych w procesie legislacyjnym UE (dokumentach Rady, Komisji Europejskiej, Parlamentu), a także studia przypadku programów. *Novum* było zastosowanie w studiach przypadku odmiennego przedmiotu analizy. Zamiast użycia pojedynczego badania analizowano wykorzystanie ewaluacji w jednostkach Budżetowania Opartego na Działaniach (ang. *Activity Based Budgeting*), a więc w zestawach działań skupionych wokół jednego, przewodniego tematu polityki (EPEC, 2005: Annex B). Słabością obydwu badań europejskich jest to, że w sposób ograniczony odnoszą się do naukowej literatury przedmiotu. Ponadto koncentrują się głównie na instrumentalnym użyciu. Mają mocno techniczną orientację, wnioski są kierowane głównie do Komisji Europejskiej, a w znikomym stopniu odwołują się do poszerzenia rozważań teoretycznych. W sumie mają charakter konsultingowy (choć oparty na naukowym warsztacie), a nie naukowy.

Prowadzone były również pewne badania skali wykonywanych ewaluacji i ich użycia. Przykładem może być badanie ewaluacji bieżących w programach funduszy strukturalnych w Irlandii (CSF Evaluation UNIT, 1998), czy roczne podsumowania praktyk ewaluacyjnych w Komisji (np. European Commission, 2002a, 2003, 2005a, 2006). Nie mają one jednak charakteru badań naukowych. Są raczej ćwiczeniem administracyjnym, sprawdzającym efektywność narzędzia (według prostych wskaźników – liczba wykonanych badań, opinia zlecających o ich przydatności, liczba wdrożonych rekomendacji).

Najnowszym badaniem empirycznym, które pojawiało się w literaturze europejskiej jest artykuł Andreasa Balthazara (2006). Dotyczy ono jednak przy-

---

[44] *Gros* wniosków z ewaluacji programów *Great Society* z lat 60. została albo upolityczniona, albo zwyczajnie zignorowana (Caralcelli, 2000).

[45] Wtedy też została stworzona grupa dyskusyjna ds. użycia ewaluacji (*Topical Interest Group In Evaluation Use*) w ramach Amerykańskiego Towarzystwa Ewaluacyjnego.

padków spoza UE – wykorzystania ewaluacji w Szwajcarii na poziomie administracji federacyjnej. Celem badania było zgłębienie kwestii, na ile kontekst instytucjonalny (a konkretnie relacja – dystans – między ewaluatorem a zlecającym) wpływa na użycie. Zastosowanym planem i metodą badawczą były studia przypadku z jakościową analizą porównawczą (*Qualtitative Comparative Analysis*) umożliwiającą zidentyfikowanie kombinacji czynników wspólnych dla różnych przypadków.

Podsumowując, należy zauważyć, że koncepcje wykorzystania ewaluacji i empiryczne badania z tej dziedziny są rozwijane głównie w USA. Używają zróżnicowanej metodologii, choć nie pozbawione są ograniczeń (trudności w oszacowaniu koncepcyjnego i długookresowego wykorzystania, uwzględnianie zwykle tylko jednej perspektywy, bardzo ograniczona triangulacja, a co za tym idzie słaby obiektywizm, rzadko analizowany kontekst). Badania w Unii Europejskiej w odniesieniu do – tak przecież powszechnych – ewaluacji programów UE są całkowicie incydentalne, ograniczone do analiz wykorzystania ewaluacji w jednej tylko instytucji – Komisji Europejskiej. Jest to tym bardziej zaskakujące, że pole badań jest bardzo szerokie. Co więcej, specyfika UE – zróżnicowane administracje krajowe wdrażające programy według zestandaryzowanych procedur i w wieloletnich pakietach – daje możliwość postawienia wielu ciekawych pytań badawczych (o rolę kultury w zarządzaniu, uczenie się organizacji, efekty, w tym uboczne, europejskiej standaryzacji) i prowadzenia analiz porównawczych.

3. PRAKTYKA EWALUACJI W SEKTORZE PUBLICZNYM

3.1. WSTĘP

Niniejszy rozdział osadza tematykę ewaluacji w praktyce zarządzania sektorem publicznym i w praktyce polityk Unii Europejskiej. W jego pierwszej części geneza ewaluacji pokazana została na tle szerszych procesów i zmian myślenia o domenie publicznej. Udowadniam, iż ewaluacja jest integralnym elementem kolejnych reform zarządzania sektorem publicznym – od klasycznej administracji poprzez menedżeryzm i dominujący w latach 90. New Public Management, aż po obecnie tworzącą się szkołę New Public Governance.

Druga część tego rozdziału podsumowuje dotychczasową praktykę ewaluacji w Unii Europejskiej – przede wszystkim w odniesieniu do programów współfinansowanych z unijnego budżetu. Omówiona w niej zostaje zarówno geneza zjawiska ewaluacji w UE, sposób podejścia do unijnych regulacji, jak i wypracowywane od lat 80. praktyki badań ewaluacyjnych.

3.2. EWALUACJA W REFORMACH ZARZĄDZANIA SEKTOREM PUBLICZNYM

Ewaluacja jest jednym z elementów ewolucji administracji i zarządzania sektorem publicznym (ang. *Public Administration & Management*). W niniejszym podrozdziale prześledzono genezę ewaluacji w szerszym kontekście przemian sektora publicznego – od tradycyjnego administrowania sektorem publicznym, poprzez reformy lat 60. koncentrujące się na planowaniu, Menedżeryzm w latach 80. i dominujący od lat 90. nurt Nowego Zarządzania Sektorem Publicznym (przesuwający akcent na zarządzanie), aż po najnowsze trendy w administracji i zarządzaniu sektorem publicznym na początku nowego milenium[1].

3.2.1. Administracja i zarządzanie w sektorze publicznym – uwagi wstępne

W tradycyjnej literaturze administracja publiczna była definiowana jako struktury umożliwiające rządowi egzekwowanie woli społeczeństwa. Nowsze ujęcia dodały rozumienie „zarządzania sektorem publicznym" jako procesu – rzemiosła praktykowanego przez urzędników publicznych oraz zarządza-

[1] Należy podkreślić, że zastosowany podział na etapy ewolucji sektora publicznego jest pewnym uproszczeniem. Przedstawione charakterystyczne elementy danych okresów często współwystępowały ze sobą, mieszając się w okresach przejściowych. Również daty są w dużej mierze orientacyjne – w poszczególnych krajach reformy pojawiały się w różnym okresie, część z administracji przyjmowała później rozwiązania, które od dawana dominowały już w krajach ościennych.

nia publicznego jako zinstytucjonalizowanych praktyk kierowania i wdrażania polityk publicznych. Tak więc współcześnie zarządzanie w sektorze publicznym jest definiowane jako (Lynn, 2003):

- struktura – czyli konstytucyjnie sformalizowany system działań i kompetencji do zarządzania i podejmowania decyzji, mający na celu umożliwienie rządowi egzekwowania woli obywateli;
- rzemiosło – praktyka i umiejętności jednostek wykonujących role menedżerów sektora publicznego;
- instytucja – system wartości, kodeks odpowiedzialnej praktyki zarządzania w sektorze publicznym, gwarantujące legitymizację i poprawność działań menedżerów sektora publicznego.

Wszystkie te trzy elementy są uwzględnione w dalszej dyskusji reform zarządzania sektorem publicznym.

Kwestią kluczową w definiowaniu zarządzania w sektorze publicznym jest pytanie o to, czym różni się ono od zarządzania w sektorze prywatnym. Mimo wielu podobieństw (codziennego zlewania się instytucji, podobnych kompetencji i analogii w strategiach) dwie podstawowe różnice to (Lynn, 2003: 17):

- odmienna logika działania oparta na służeniu odmiennym interesom społecznym – sektor prywatny kieruje się zyskiem, wypracowywanym z korzyścią dla określonej, z reguły wąskiej grupy, sektor publiczny na celu ma dobro wspólnoty i równość;
- wynikające z tego inne umiejętności, system zasad i hierarchia wartości.

Fundamentalne różnice w logice sektora prywatnego i publicznego podsumowują Stewart i Ranson (patrz poniższa tabela).

Tabela 6. Odmienne filozofie sektora prywatnego i publicznego

Logika sektora prywatnego	Logika sektora publicznego
Indywidualny wybór na rynku	Kolektywny wybór w ramach polityki
Potrzeba zysku	Potrzeba zasobów
Zamknięcie na inne działania prywatne	Otwarcie na działania publiczne
Równość rynku	Równość potrzeb
Poszukiwanie zaspokojenia potrzeb rynkowych	Poszukiwanie sprawiedliwości
Niezależność klienta	Obywatelstwo
Konkurencja jako instrument rynku	Kolektywne działania jako instrument polityki
Wypadnięcie z rynku stymulatorem działań	Zagwarantowanie udziału (głosu) warunkiem działań

Źródło: (Stewart, Ranson, 1994).

Różnice między organizacjami publicznymi a prywatnymi, szczególnie w ostatnich dwudziestu latach, coraz częściej się zacierają (funkcjonowanie *quangos* – zadaniowych agencji rządowych operujących na rynku, firmy zależne i częściowo kontrolowane przez państwo, przenoszenie technik działania z sektora prywatnego do publicznego, stosowanie w agencjach publicznych zarządzania opartego na wynikach, a co za tym idzie kryterium zysku).

Niemniej, opierając się na badaniach, można wyróżnić zestaw cech specyficznych dla administracji publicznych. Można je podzielić na trzy grupy (szerzej patrz: Rainey, Chun, 2005: 92–95):

- specyficzne warunki środowiskowe – ograniczony rynek gospodarczy dla tworzonych produktów, opieranie się na politycznym oszacowywaniu alokacji finansowych (a co za tym idzie, słabsze bodźce do oszczędności i racjonalności w gospodarowaniu zasobami), kontrola zewnętrzna ze strony polityków, silne wpływy polityczne;
- organizacja (otoczenie transakcji) – uwzględnianie kosztów zewnętrznych, dopuszczalny monopol, szersze, często symboliczne oddziaływanie (kwestie opinii publicznej, kwestie narodowe), presja publiczna oraz kwestie etosu pracy – przejrzystości dla opinii publicznej, otwartości, poszanowania sprawiedliwości i równości;
- role organizacyjne, struktury i procesy – zróżnicowanie i wieloaspektowość celów, kryteriów wyboru opcji, silniejsze konflikty i konieczność łączenia bardzo zróżnicowanych opinii, więcej negocjacji, poszukiwania równowagi miedzy zewnętrznymi naciskami a wewnętrznym zarządzaniem, mniejsze możliwości decyzyjne, mniejszy autorytet i znaczenie menedżerów, mniejsza elastyczność w podejmowaniu decyzji, na wyższych szczeblach tendencja do silnej kontroli działań niższego szczebla, do wydłużania i formalizowania procedur, tendencja do budowania bardziej zbiurokratyzowanych organizacji opartych na stałych procedurach (ang. *red tape*), słabsza relacja między osiąganymi efektami w pracy a zachętami dla menedżerów publicznych, niższy poziom satysfakcji z pracy.

3.2.2. Początki ewaluacji w Klasycznej Administracji Publicznej

Punktem wyjścia ewolucji współczesnej administracji i zarządzania sektorem publicznym było tradycyjne administrowanie sektorem publicznym (ang. *Public Administration*), które dominowało od końca XIX wieku aż do połowy lat 60. wieku XX. Kluczowymi elementami Klasycznej Administracji Publicznej były (Gruening, 2001: 3; Osborne, 2006: 378):

- kierowanie się zasadą prawa i zestawem standardowych procedur, reguł i wytycznych administracyjnych;
- zasada jednego ośrodka decyzyjnego i hierarchii w odniesieniu do delegacji kompetencji (kluczem do sukcesu jest stabilna, trwała struktura administracyjna);
- centralna rola biurokracji w formułowaniu i wdrażaniu polityk i dostarczaniu usług oraz relatywne oddzielenie jej od polityków;
- zasada podziału pracy, specjalizacji organizacji biurokratycznych i obiektywnego podejścia naukowego;
- przyrostowe budżetowanie (zwiększanie wydatków z roku na rok budżetowy).

Nieco upraszczając, administratorzy oddzieleni od polityków mieli w profesjonalny sposób, kierując się wypracowanymi procedurami: planować, organizować działania publiczne, prowadzić politykę etatową, kierować, koordynować, raportować i prowadzić budżetowanie (ang. *posdcorb*; Gulick, 1937).

Teoretycznych podstaw temu nurtowi dostarczyły politologia (z wizją państwa unitarnego, gdzie polityka i system wdrażania są zintegrowane pionowo w ramach rządu; Wilson, 1887) oraz klasyczne teorie organizacji – zarządzanie naukowe (Fredericka Taylora) i, nieco późniejsze, zarządzanie administracyjne rozwijane głównie w USA.

Z perspektywy rozwoju ewaluacji ważnym elementem ówczesnego paradygmatu było założenie, iż polityki publiczne wdrażane przez państwo i jego aparat biurokratyczny powinny kierować się naukowym obiektywizmem. Wraz z wprowadzaniem programu *New Deal* w latach 30. w USA, zaangażowanie państwa znacznie wzrosło, a tym samym pojawiła się silniejsza potrzeba testowania i weryfikowania skuteczności i racjonalności tych działań w oparciu o obiektywne, jak sądzono, podejście naukowe. Wtedy to pojawiły się pierwsze badania naukowe stosowane w odniesieniu do programów publicznych. Były to głównie badania jakości programów edukacyjnych – skuteczności rozwijanych *curriculów* (Worthen, 1990)[2] i programów szkoleniowych (np. badania efektywności programów wzmacniania morale amerykańskiej armii w latach 40. XX wieku; Cronbach, Associates, 1981: 30)[3]. Guba i Lincoln (1989: 22–28) charakteryzują ówczesne badania jako skoncentrowane na pomiarze i opisie. Ich słabością był nadmierny rygoryzm w metodologii badań społecznych (wzorowanie się na naukach przyrodniczych w prowadzeniu eksperymentów i ocenie ich ważności), a przede wszystkim pomijanie elementu oceny. Ewaluatorzy dostarczali opisu relacji przyczynowo-skutkowych, ale nie formułowali ostatecznej oceny na temat wartości danej interwencji.

3.2.3. Ewaluacja w Neoklasycznej Administracji Publicznej

W latach 60. amerykańska administracja prezydencka rozpoczęła inicjatywę *War on Poverty* składającą się z całego pakietu programów sektora publicznego (później nazwane planem *Great Society*), z czym związany był lawinowy wzrost zaangażowania państwa i wzrost wydatków publicznych w szeregu sektorów ówczesnej gospodarki[4]. Przy tak znaczącym wzroście zaangażowania administracji publicznej zaczęto poszukiwać dalszych metod racjonalizowania zarządzania w sektorze publicznym. Serią reform wprowadzono do powszechnego użycia w administracji amerykańskiej „Planowanie, Programowanie, System Budżetowy" (ang. *PPBS – Planning, Programming, Budgeting System*)

---

[2] Worthen wymienia jako pierwsze badanie jeszcze sprzed okresu *New Deal* studium porównawcze umiejętności literowania przeprowadzone w latach 1897–1898 wśród 33 tys. uczniów szkół amerykańskich. Guba i Lincoln (1989: 27) przytaczają ośmioletnie badanie R.W. Tylera na temat skuteczności programów autorskich dla gimnazjalistów.

[3] Dodatkowym impulsem i inspiracją był rozwój badań eksperymentalnych i obserwacji uczestniczącej w badaniach społecznych. Najbardziej znany przykład, przytaczany jako inspiracja późniejszych ewaluacji, to eksperyment Hawthorne (Cronbach and Associates 1981; Shafritz 2003).

[4] Według obliczeń Shadisha i współpracowników (1991: 22) wydatki na programy publiczne w USA wzrosły z 23,5 mld USD w latach 50. do 428,3 mld USD pod koniec lat 70. Gdy uwzględnimy inflację, okazuje się, że był to wzrost o 600%! Trend ten – budowanie państwa socjalnego (welfare state) narastał też w innych rozwiniętych krajach demokracji zachodnich.

wykorzystujące mikroekonomiczne techniki decyzyjne. Były one oparte na silnym przeświadczeniu, że centralne planowanie i administracja publiczna prowadzić będą do zoptymalizowania decyzji społeczno-gospodarczych (Gruening, 2001). Neoklasyczna Administracja Publiczna opierała się na tych samych podstawach filozoficznych co administracja tradycyjna z poprzedniego okresu – zwiększono jedynie nacisk na naukowe zarządzanie, szczególnie na element cyklicznego, ciągłego planowania. W praktyce polityk publicznych oznaczało to (Gruening, 2001; Heinrich, 2003):

- zastosowanie logiki procesowej do analizy interwencji publicznych, tzn. planowanie i ocena wszystkich działań według ciągu: nakładów, przepływów, produktów, rezultatów, programów i alternatywnych rozwiązań;
- strukturalizowanie działań oraz funkcji instytucji publicznych w formie mierzalnych celów;
- tworzenie wieloletnich planów budżetowych (konstruowanych na zasadzie zadaniowej);
- obiektywną analizę wdrażanych programów opartą na badaniach i ewaluacji.

Założenie było proste – każde działanie miało służyć konkretnemu celowi, do niego był przyporządkowany budżet, zaś weryfikowanie postępów z wdrażania opierało się na sprawdzaniu zmian w mierzalnych wskaźnikach. To nowe podejście szybko zaczęły stosować również administracje krajów Europy zachodniej (Wielka Brytania, Francja, Holandia, Niemcy) (Toulemonde, 2000).

Wprowadzenie PPBS i intensyfikacja działań publicznych było momentem przełomowym dla rozwoju ewaluacji. Badania ewaluacyjne rozszerzono na inne pola – z programów głównie edukacyjnych na programy infrastrukturalne, zasobów ludzkich, etc. Ewaluacja w tamtym okresie koncentrowała się na ustaleniu relacji przyczynowo-skutkowej między nakładami zaangażowanymi w program a jego bezpośrednimi i natychmiastowymi efektami. Korzystano z modeli i metod eksperymentalnych na dużych próbach (grupy kontrolne i metody ilościowe). Praktyka kolejnych lat ujawniła jednak słabości i ograniczenia tego podejścia. Ewaluacje nic nie mówiły o rzeczywistości programów – o tym, jakie procesy stały za pojawieniem się lub brakiem oczekiwanych efektów. Tworzone modele często okazywały się zbyt abstrakcyjne – operowały zbyt dużym stopniem agregacji informacji, pomijały lokalną specyfikę programów (późniejsza krytyka nazwała to metodą czarnej skrzynki) (Greene, 2003; Herman et al., 1987).

3.2.4. Menedżeryzm i użycie ewaluacji

W końcu lat 70. kryzysy paliwowe zmusiły administracje wielu rozwiniętych krajów do refleksji nad kosztami państwa dobrobytu, racjonalnością rozbudowanych programów publicznych i efektywnością sektora publicznego (Pollitt, Bouckaert, 2003: 21). Szerszym impulsem było przechodzenie gospodarek rozwiniętych od fordyzmu do postfordyzmu i związana z tym nowa organizacja rynku, metod pracy i procesów produkcji oraz roli państwa (Gorzelak, 1995: 18–21)[5]. Administracja publiczna musiała się określić w nowej,

[5] W sferze produkcji postfordyzm charakteryzował się ekonomią jakości, produkcją elastyczną zależną od popytu, *quasi*-pionową integracją i zlecaniem prac podwykonawcom. W sfe-

postfordystowskiej rzeczywistości. Zaowocowało to tzw. menedżeryzmem (ang. *New Managerialism*), który jednocześnie zaczęła wprowadzać administracja brytyjska (pod rządami M. Thatcher) i amerykańska (prezydentura R. Regana), a za nimi głównie administracje krajów anglosaskich – nowozelandzka, australijska, kanadyjska.

Podstawą teoretyczną dla nowego ruchu była ekonomia neoklasyczna, szczególnie teorie racjonalnego wyboru/wyboru publicznego, w tym zarządzanie systemami policentrycznymi (Ostrom, Ostrom, 1971) i szereg „racjonalnych technik zarządzania" zapożyczonych z sektora prywatnego. Rdzeniem filozofii nowego paradygmatu było przestawienie się z administrowania na zarządzanie sektorem publicznym (od podejścia hierarchicznego „rozkazuj i kontroluj" do rynkowych zachęt i konkurencji) i hasło efektywnego (tańszego) zarządzania pieniędzmi publicznymi (*Value for Money*). Jako kluczowe elementy menedżeryzmu można wyróżnić (Gruening, 2001; Kettl, 1997; Shafritz, Russell, 2003):

- nową filozofię planowania budżetowego – uzależnienie wielkości alokacji od planowanych produktów (czyli, jak to obrazowo ujmuje Kettl (1997: 449), od pytania o nakłady – „Ile powinniśmy wydać?", do pytania o efekt – „Jakie produkty przynoszą nasze nakłady?")[6];
- wprowadzenie do sektora publicznego technik i narzędzi racjonalnego zarządzania przeniesionych – często w całości i bez zmian – z sektora prywatnego: Strategicznego Planowania i Zarządzania przez Cele (*Management by Objectives* – MBO) opartych na ciągu logicznym: analiza stanu i perspektyw – wizja – cele strategii – nakłady – działania – efekty – ich ocena;
- wprowadzenie zapożyczonego z sektora prywatnego Zarządzania opartego na Wynikach (ang. *Performance Management*) i rachunkowości polegających na ocenie efektywności działań na podstawie osiąganych wskaźników, stopnia realizacji celów założonych w danej strategii;
- wprowadzenie do stabilnego, sztywnego sektora publicznego konkurencji rynkowej – rynkowego podejścia do usług publicznych, zlecania na zewnątrz i prywatyzacji części zadań;
- przesunięcie akcentu z procedur i struktur na procesy, rozbicie funkcji i działań publicznych na szereg oddzielnych i łatwych w zarządzaniu procesów (analizowanych i prowadzonych jak w sektorze prywatnym) (Kettl, 2006: 316) oraz odejście od sztywnej hierarchii, pionowej delegacji kompetencji na rzecz wyspecjalizowanych działów i agencji (konkurujących ze sobą) elastycznych działań, płaskiego zarządzania organizacją, grup zadaniowych;

---

rze pracy zwiększono nacisk na współodpowiedzialność pracowników, ich ciągłe szkolenie, poziomą koordynację i elastyczność zadań. Ponadto zdywersyfikowały się rynki pracy, a nowym trendem stała się konsumpcja zindywidualizowana (szerzej patrz: Gorzelak, 1995: 18–25).

[6] W praktyce najpopularniejszym podejściem była technika „*Zero-based budgeting*". ZBB domagało się uzasadnienia racjonalności dla całego planu budżetowego i poszczególnych jego pozycji, podczas gdy dotychczasowe tradycyjne, przyrostowe budżetowanie tworzyło budżet w oparciu o rezultaty decyzji i zaszłości budżetowe lat poprzednich (tzw. bazy budżetowej), koncentrując się wyłącznie na różnicach pomiędzy jednym a drugim rokiem budżetowym (Shafritz, 2003: 258).

- zmianę mentalności urzędniczej i relacji biurokratów z obywatelami – przeistoczenie urzędników w przedsiębiorczych menedżerów, zaś petentów w urzędach w klientów (a więc zorientowanie na usługi i akcent na ich jakość, na stopień zadowolenia obywateli-klientów), w parze z tym szła promocja i marketing sektora publicznego.

Reformy menedżeryzmu były wprowadzane w sposób szybki i radykalny[7], choć w literaturze wyróżniane są dwa nurty o różnej sile reform (Kettl, 1997; Pollitt, Bouckaert, 2003). Pierwszy, charakteryzowany hasłem „pozwólmy menedżerom zarządzać" (*let the managers manage*), zakładał, że menedżerowie sektora publicznego wiedzą, co robić, a tym, co krępuje ich przedsiębiorczość, są istniejące procedury i hierarchie. Znosząc te bariery menedżerowie będą mogli pracować lepiej i wdrażać efektywnie swoje wizje i strategie. Drugi, bardziej radykalny nurt (zastosowany głównie w Wielkiej Brytanii, USA i Nowej Zelandii), to „zmuśmy menedżerów do zarządzania" (*make the managers manage*) – zakładający, że tym, co zmusi menażerów sektora publicznego do lepszej i tańszej pracy jest presja rynkowa, konkurencja, akcent na rozliczenie z efektów (oceniane w odniesieniu do stopnia satysfakcji klientów). Nurt pierwszy koncentrował więc reformy na poprawach struktur i upraszczaniu procedur, nurt drugi na prywatyzacji, kontroli produktów i ich jakości.

W polu ewaluacji rozczarowanie szerokimi programami (np. USA – *Great Society*) opartymi na inżynierii społecznej i PPBS, przemiany społeczne lat 70. (kryzysy polityczne i rozwijający się pluralizm społeczny) oraz ujawnione ograniczenia metod eksperymentalnych i ilościowych wywołały dyskusję nad zmianą podejścia (Greene, 2003)[8]. Akcent postawiono na użycie metod jakościowych jako bardziej odpowiednie narzędzie uchwycenia zróżnicowanych opinii i punktów widzenia różnych aktorów społecznych. Ta kolejna fala ewaluacji koncentrowała się na identyfikowaniu i adresowaniu wyjątkowych potrzeb różnych aktorów-klientów zaangażowanych w lokalne programy oraz opisywaniu specyfiki procesów danego programu (z perspektywy wdrażania) w celu poprawy efektywności. Menedżeryzm potrzebował narzędzia kontroli postępów i jakości – nowa ewaluacja, obok audytu, dawała takie możliwości. Od lat 80. zaczęto więc w ewaluacji rozwijać metody i działania, które zapewniałyby rzeczywiste przełożenie wniosków z ewaluacji na praktykę programów i polityk (Herman et al., 1987). Rozwijano utylitarne podejście – ewaluacja dostarczała informacji menedżerom, oceniała efektywność działań (zestawiając nakłady i procesy z produktami), ułatwiała proces uczenia się danej organizacji, weryfikowała efektywność działań, dawała konkretne rekomendacje na temat poprawy procesów i kontekstu wdrażania danego programu.

---

[7] Głównym terminem ówcześnie używanym na określenie tak radykalnej zmiany był „*Reenginering*", czyli ponowne budowanie – start od podstaw. Dla praktyki organizacji publicznych oznaczało to trzy kroki: 1. analizę stanu danej organizacji – tego, jakich i w jaki sposób dostarcza usług, jaką ma strukturę, jak prowadzone są procesy; 2. ocenę potrzeb klientów, grupy docelowej; 3. zbudowanie wizji i strategii realizacji, włączając w to sposoby działań (zaprojektowanie efektywnych procesów dostarczania usług, produktów) (Shafritz 2003: 262).

[8] Dyskusja ta miała burzliwy przebieg i jest czasem określana Okresem Wojny Paradygmatów (*The Paradigm Wars*) między tradycyjnym ilościowo-pozytywistycznym i eksperymentalnym podejściem a konstrukcjonizmem opartym głównie o metody jakościowe.

Nowe ewaluacje miały jednak również swoje ograniczenia – trudno było agregować spostrzeżenia z badań i uogólniać wnioski. Serie szczegółowych badań i studiów przypadku zorientowane na procesy miały więc małe znaczenie praktyczne dla polityków i co za tym idzie minimalną moc sprawczą w odniesieniu do budowania szerszych wizji, strategii, weryfikowania założeń stojących za całymi politykami, wyciągania lekcji na przyszłość dla nowych generacji programów. Pojawiło się rozczarowanie samymi efektami tych czysto technicznych ewaluacji.

3.2.5. Rola ewaluacji w Nowym Zarządzaniu Sektorem Publicznym (NPM)[9]

Budując na doświadczeniach menedżeryzmu w latach 90. pojawiła się, dominująca obecnie, doktryna Nowego Zarządzania Sektorem Publicznym (ang. *New Public Management – NPM*)[10]. W literaturze przedmiotu podawane są dwie publikacje, które dały impuls nowej doktrynie (porównaj: Gruening, 2001: 11; Jones, Thompson, 1997: 15). Pierwszą była *In a Search of Excellence* (Peters et al., 1992), której autorzy dowodzili, iż najbardziej efektywne amerykańskie korporacje odchodzą od racjonalnych technik zarządzania na rzecz struktur organicznych, humanistycznego podejścia do zasobów ludzkich, kulturowo elastycznych strategii. Ta dyskusja została przeniesiona do sektora publicznego, który od czasów menedżeryzmu czerpał wzorce z biznesu. Drugą publikacją było *Reinventing Government* (Osborne, Gaebler, 1992), której autorzy postulowali, by sektor publiczny skoncentrował się na misji – koordynacji i sterowaniu – zamiast na codziennym zarządzaniu (winno być zlecane na zewnątrz), poszukiwaniu rynkowych nieadministracyjnych rozwiązań, pomiarze sukcesu poprzez stopień zadowolenia klientów-obywateli. Pojawiały się też głosy, że słabością menedżeryzmu jest kopiowanie często przestarzałych już wzorów biznesowych do sektora publicznego, bez należytego uwzględnienia specyfiki tegoż ostatniego (Hood, Jackson, 1991; Metcalfe, Richards, 1990). Szerszy impuls do dalszych zmian dały też w latach 90. przyspieszające procesy globalizacji, w tym przemiany technologiczne, które odbierały możliwość pełnej administracyjnej kontroli, wymuszając elastyczność i koordynację.

Zestawiając nowe postulaty z menedżeryzmem, charakterystyczne elementy reform Nowego Zarządzania Sektorem Publicznym można podsumo-

---

[9] W polskiej literaturze przedmiotu spotyka się tłumaczenie: „Nowe zarządzanie publiczne" (porównaj: Mazur, 2006). Zdecydowałem się jednak na tłumaczenie rozszerzające, które jasno zaznacza pole zarządzania – „sektor publiczny".

[10] Niektórzy autorzy traktują łącznie menedżeryzm i NPM – licząc NPM już od przełomu lat 70. i 80. (Gruening, 2001; Kettl, 1997; Osborne, 2006). W pracy zastosowano jednak rozróżnienie między menedżeryzmem a NPM (porównaj Shafritz, 2003: 255–300), przyjmując za moment przełomowy artykuł Hooda (1991), gdzie po raz pierwszy użyto terminu NPM, charakteryzując całościowo nową doktrynę. Obecnie termin NPM funkcjonuje co najmniej w trzech kontekstach, co może prowadzić do pewnej konsternacji. Po pierwsze, oznacza doktrynę administracji publicznej, po drugie, cały zestaw technik zarządzania w sektorze publicznym, po trzecie wreszcie służy określaniu międzynarodowej tendencji, ruchu reformatorskiego w administracjach krajowych (głównie z kręgu kultury anglosaskiej).

wać w sześciu punktach (Gruening, 2001; Heinrich, 2003; Hood, 1991; Jones, Thompson, 1997):

- silniejsze zorientowanie na jakość usług – zaadaptowanie do potrzeb sektora publicznego zarządzania jakością (ang. *Total Quality Management*) (SWISS, 1992);
- kontynuowanie logiki strategicznego myślenia, podejścia procesowego i zarządzania opartego na wynikach tak w planowaniu budżetowym, jak i wdrażaniu interwencji; silniejszy akcent zostaje jednak położony na szersze efekty (ang. *outcomes*), zamiast – jak dotychczas – głównie na produkty (ang. *outputs*); planowanie w okresach wieloletnich (ang. *multi-annual strategies*);
- wprowadzenie rynkowego i *quasi*-biznesowego podejścia do sektora publicznego – prywatyzowanie i zlecanie usług na zewnątrz, jednak tylko do wybranych funkcji, zaś stosowanie *quasi*-rynkowych mechanizmów w polach sektora publicznego, które nie mogą być sprywatyzowane;
- dekoncentracja i decentralizacja (również przestrzenna) zadań i funkcji – przede wszystkim przekazanie ich wyspecjalizowanym agencjom *quasi*-rządowym, tworzenie departamentów zorientowanych wokół usług, w zarządzaniu danymi jednostkami stosowanie elastycznych grup zadaniowych (ang. *task-force*), grup eksperckich powoływanych do rozwiązania danej kwestii;
- wyciąganie lekcji i adaptowanie metod i technik z sektora prywatnego przy poszanowaniu specyfiki działania sektora publicznego, jego odmiennych celów i dłuższej perspektywy czasowej;
- szersze stosowanie badań naukowych, audytu i ewaluacji tak do bieżącej oceny procesów wdrażania, jak i weryfikacji szerszych, długookresowych strategii, uzasadniania odpowiedzialności działań publicznych przed obywatelami.

NPM ugruntował więc rolę ewaluacji jako narzędzia podnoszenia efektywności działań sektora publicznego. Badania przestały jednak koncentrować się wyłącznie na procesach, produktach i krótkoterminowych rezultatach, a zaczęły adresować oddziaływanie inwestycji publicznych. Co więcej, dużo silniej powiązano ewaluację z bieżącymi działaniami danych organizacji i z ich strategicznym planowaniem. Organizacje publiczne zaczęły internalizować zadania ewaluacyjne, rozwijając tzw. ewaluacje wewnętrzne prowadzone na bieżąco przez kadrę własną danych organizacji (Wollmann, 2003: 2). Pomiar efektów i strategiczne użycie ewaluacji stawiało oczywiście nowe wyzwania metodologiczne – szczególnie w sferze określenia relacji przyczynowo-skutkowej między działaniami a ich długoterminowymi efektami, w tym efektami ubocznymi. Nowe metody czerpały zarówno z badań ilościowych jak, i jakościowych, zaczęto w dużo większym stopniu używać myślenia systemowego (modele logiczne) i procesowego. Ewaluacja stała się więc dla sektora publicznego tym, czym mechanizm rynkowy dla sektora prywatnego. Idee NPM, a wraz z nimi użycie ewaluacji, promowane w kolejnych latach przez Bank Światowy i OECD, stały się wzorcem, do których dążą współczesne administracje publiczne.

3.2.6. Ewaluacja w reformach sektora publicznego na początku XXI wieku

Pod koniec XX wieku na bazie opisanych już elementów NPM pojawiły się nowe tendencje w administracji i zarządzaniu sektorem publicznym. Charakteryzując współczesną sytuację, można zidentyfikować sześć głównych trendów (OECD, 2005; Wolf, 2000):

- Otwarty rząd i włączanie obywateli (ang. *Open government & Citizen engagement*). Rządy i administracje otwierają się na obywateli. Po pierwsze, starają się, by ich działania były bardziej przejrzyste (ang. *transparency*), a standardy relacji jasne. Po drugie, widać wyraźny trend zwiększania dostępności administracji i państwa dla obywateli. Jednak w odróżnieniu od ery NPM, gdzie udział obywateli i firm – postrzeganych jako „klientów" – nie jest ograniczany tylko do otrzymywania usług i komentowania ich jakości, ale jest rozszerzany do oceny, czy wręcz współuczestnictwa w formułowaniu strategii, kierunków polityk i tworzenia nowych usług (OECD, 2005; Wolf, 2000). Obywatele i aktorzy życia społecznego z klientów polityk publicznych stają się ich interesariuszami (ang. *stakeholders*)[11].
- Systemowe podejście. Próby całościowego, systemowego podejścia wynikają z uświadomienia sobie współzależności dotychczas izolowanych elementów i sieciowej organizacji zglobalizowanych społeczeństw i gospodarek (Castells, 2000). Próby te przejawiają się w kilku formach. Po pierwsze, jest to odchodzenie od klasycznego rozdziału tworzenia polityki i zarządzania jej wdrażaniem na rzecz podejścia wielowymiarowego – interakcji między politykami, doradcami, służbą cywilną, interesariuszami (Wolf, 2000). Chodzi o „wyjście" poza sektor publiczny i spojrzenie na polityki przy uwzględnieniu sieci powiązań sektora publicznego, rynku i społeczeństwa obywatelskiego – różnych perspektyw, opcji, zależności, konfliktów interesów zróżnicowanych grup interesariuszy (ang. *holistic mapping*). Po drugie, chodzi o przełamywanie sektorowych podziałów i wypracowywanie spójnych polityk, całościowych i długofalowych wizji rozwoju. Po trzecie wreszcie, są to próby zintegrowania dotychczas oddzielanych od siebie czterech komponentów sektora publicznego (polityk, kadr, środków finansowych, organizacji) we współgrający system, w tym także przełamywanie podziałów organizacyjnych (Kettl, 2006: 314).
- Koordynacja i zarządzanie sieciami (ang. *governing by network*) (Goldsmith, Eggers, 2004). Jest ono logiczną konsekwencją podejścia systemowego. Skoro współczesne społeczeństwa są sieciowe, a tworzenie polityk, usług i wiedzy opiera się tak silnie na współzależnościach i współgraniu wielu elementów i aktorów (Castells, Cardoso, 2006), to rolą nowoczesnej administracji jest koordynowanie i strategiczne wpływanie na te procesy. Chodzi więc o współzarządzanie sieciami informacji, organizacji, zasobów ludzkich, koordynowanie działań pomiędzy poziomami rządzenia, sektorami działań, uczestnikami, a także o partnerstwo z sektorem prywatnym i po-

[11] Interesariusz to „każdy, kto «w sposób znaczący wpływa na lub jest dotknięty przez podejmowanie decyzji przez kogoś» (Chevallier) lub, innymi słowy, „każda osoba, która będzie bezpośrednio dotknięta przez rozstrzygnięcie władzy" (Izdebski, 2007: 16)

zarządowym (Goldsmith, Eggers, 2004: 19–23; OECD, 1997; Wolf, 2000: 694).

- Efektywna i odpowiedzialna administracja. Rdzeniem i „jedynie słuszną drogą" pozostaje budżetowanie i zarządzanie oparte na wynikach (*Performance Budgeting & Management*) (Premfors, 2006: 334)[12], połączone z otwieraniem sektora publicznego na konkurencję i dalszym stosowaniem podejść rynkowych czy *quasi*-rynkowych. Dodatkowo, w zglobalizowanym, informacyjnym społeczeństwie gwarancją efektywności jest zdolność szybkiej reakcji i adaptacji, zmiany strategii działania, a więc szeroko pojęta elastyczność (procedur, struktur, strategii działań). Poszukiwanie elastyczności wymusza kolejne restrukturyzacje i reorganizacje. Kluczowym wyzwaniem jest jednak równowaga między elastycznością elementów systemu (organizacji i osób) i otwarciem na rynkową konkurencję a kontrolą gwarantującą odpowiedzialność, rozszerzającą ocenę skuteczności działań publicznych poza kryterium krótkookresowego zysku finansowego, zapobiegającą mieszaniu prywatnych zysków z interesem publicznym. Trend reform biegnie od dotychczasowych analiz *ex ante* i kontroli pojedynczych operacji, ku rozwiązaniom opartym na analizie i kontroli *ex post*, rozwijaniu wewnętrznych kontroli (w tym wzmacnianiu etyki zawodowej służby cywilnej), weryfikowaniu już nie pojedynczych operacji, a całego funkcjonalnego systemu, poszerzeniu oceny oddziaływania o szersze kryteria – środowiskowe, kulturowe, społeczne (Bouckaert, 2006; OECD, 2005; Wolf, 2000: 695).
- Nowoczesne podejście do zasobów ludzkich. W nowoczesnej gospodarce głównym kapitałem jest wiedza – kapitał społeczny i ludzki. Nowe reformy kładą więc szczególny nacisk na zarządzanie zasobami ludzkimi i wiedzą. Nacisk na profesjonalizm pracowników wymuszany jest z jednej strony nowym systemem płac (coraz częściej opieranym, jak w sektorze prywatnym, na osiąganych wynikach), z drugiej strony zachętami w postaci szkoleń, ustawicznego kształcenia czy atrakcyjnego systemu kadrowego (szybkie ścieżki awansu, bardziej płaska struktura, sprzyjająca temu, by przyciągnąć i utrzymać w sektorze publicznym specjalistów). W administracji publicznej coraz częściej stawia się na pobudzanie kreatywności pracowników, niezależne, krytyczne myślenie, współpracę i interakcje (podejście zespołowe, zadaniowe). W skali organizacji akcent przesuwany jest na budowanie pamięci instytucjonalnej, uczenie się organizacji i wymianę doświadczeń.
- Globalna wymiana doświadczeń. Wymiana doświadczeń również uległa globalizacji. Innowacje w sektorze publicznym pojawiają się na skalę globalną – poszczególne administracje krajowe, organizacje szukają i znajdują rozwiązania na całym świecie. Administracje krajów anglosaskich przestały być głównym źródłem inspiracji. Nowe pomysły na upraszczanie usług, do-

[12] PM jest definiowane jako cykl zarządzania, w którym na początku określane są cele interwencji, menedżerowie mają relatywnie wolną rękę w wyborze strategii dochodzenia do tych celów, skuteczność ich działań jest mierzona, badana i raportowana, zaś informacje te służą podejmowaniu decyzji na temat przyszłości danych interwencji publicznych. W skrócie PM to nacisk na zysk, rezultat prowadzonych działań w odniesieniu do celów, którym miały służyć (OECD; 2005: 59).

tarcie do obywateli, organizację procesów powstają we wszystkich krajach rozwiniętych. Główną platformą jest tu OECD, ale również UE. Reformatorzy coraz częściej, w odróżnieniu od czasów NPM, uwzględniają specyfikę kraju – odchodzi się od promowania jednego modelu dobrego dla wszystkich, na rzecz uważnego kulturowo, wrażliwego adaptowania „najlepszych światowych praktyk" we własnych warunkach.

Siła i specyfika tych nowych procesów sprawiła, że w literaturze przedmiotu pojawia się nowy termin – podkreślający, iż administracja publiczna wchodzi w kolejny etap swojej ewolucji (Dunsire, 1995: 17; Osborne, 2006; Wolf, 2000). **Nowe Rządzenie Sektorem Publicznym (ang. *New Public Governance*)** łączy biurokrację i administrację z polityką i zarządzaniem, postuluje wyjście poza wizję biurokratyczno-menedżerską i uwzględnianie interakcji między rządzącymi, maszynerią administracji a społeczeństwem obywatelskim i biznesem. Podstawą NPG jest kapitał relacyjny, uznanie pluralizmu i wielowymiarowości (spojrzeń, perspektyw, interesów, interakcji), zaś w praktyce zorganizowane sieciowo partnerstwa, zorientowane na wysoką efektywność (Kettl, 2006: 316). Podstawy teoretyczne NPG czerpie z socjologii organizacji, teorii sieci, literatury zarządzania strategicznego i zarządzania relacjami (Osborne, 2006: 383–384). Jak zauważa Izdebski (2007: 12–14), podczas gdy *New Public Management* za punkt odniesienia miał rynek i konkurencyjność, *New Public Governance* stawia w centrum społeczeństwo obywatelskie (rozumiane jako sieć organizacji i powiązań społecznych).

W tym nowym paradygmacie ewaluacja może znaleźć nowe, szersze zastosowanie, które można podsumować w pięciu punktach.

Po pierwsze, ewaluacja staje się elementem promowania otwartego i przejrzystego rządzenia. Jest elementem systemu kontroli i wzmacniania odpowiedzialności (ang. *accountability*), o czym szerzej za chwilę. Pokazuje jakość interwencji publicznych, krytycznie analizując ich efekty. Wpisuje się w hasło otwartego rządu i dostępności informacji, gdyż raporty ewaluacyjne są publikowane, a dyskusja nad nimi przyczynia się do poszerzania dialogu publicznego. Wyzwaniem jest tu zachowanie równowagi między obiektywnym i pełnym informowaniem opinii publicznej o wynikach analiz a koniecznością zachowania przystępnej, przejrzystej i skróconej formy (coraz częściej pisze się o rosnącej piramidzie raportów ewaluacyjnych, które pozostają niezauważone i mało przydatne zarówno dla zarządzających, jak i dla szerszej opinii publicznej).

Po drugie, ewaluacja współgra z wielowymiarowym, systemowym podejściem do polityk i zarządzaniem sieciami. Wypracowane formy (szczególnie takie podejścia badawcze, jak ewaluacja uczestnicząca czy realistyczna) lub techniki ewaluacji (triangulacja, dialog ze zlecającym i interesariuszami, multi-sektorowe podejście, analiza powiązań) umożliwiają zidentyfikowanie różnych perspektyw, interakcji i współzależności, wyjście poza sektorowe myślenie o programach publicznych. Co więcej, nowoczesne teorie socjologii i zarządzania stosowane w ewaluacji umożliwiają praktyczną analizę i ocenę jakości sieci, identyfikowania blokad, doradzania ulepszeń. Ewaluacje wydają się więc wartościowym narzędziem tak w momencie formułowania polityk,

jak i ich wdrażania oraz oceny *ex post*. W tym punkcie pojawia się wyzwanie natury metodologicznej – zachowanie rygorów i standardów warsztatu naukowego przy jednoczesnym kompleksowym podejściu, szerokim polu tematycznym i krótkim czasie badań. Narzędzia ilościowej analizy sieci są wciąż słabo rozwinięte i rzadko stosowane w ewaluacjach. Wiele z badań ewaluacyjnych ucieka się wyłącznie do metod jakościowych, jednak poddając się presji czasowej i rynkowej zaniedbuje standardy warsztatu naukowego. W rezultacie część może być uznana od strony metodologicznej jedynie za przyczynki do poważnych analiz. Tymczasem oceny są formułowane na podstawie szczątkowych danych.

Po trzecie, ewaluacja może być przydatna we wzmacnianiu trendu współuczestnictwa obywateli (włączania szerszych grup interesariuszy, angażowania społeczeństwa obywatelskiego w dyskusję nad strategią i uczestniczenie w jej formułowaniu, w tworzenie kierunków polityk i nowych usług) i rozwijania sieci. Szczególnie przydatne są tu podejścia uprawniające i partycypatywne – włączanie interesariuszy w moment dyskusji i oceny programu. Efektywne animowanie dialogu między interesariuszami wokół oceny programu przyczynia się do rozwijania i wzmacniania relacji. Tak więc ewaluacje mogą być wykorzystywane do wzmacniania i poszerzania kapitału relacyjnego skoncentrowanego wokół danych programów i polityk. Wyzwanie stojące w tym polu to zachowanie przez ewaluatorów obiektywizmu i krytycyzmu. Wchodzenie w rolę animatora dyskusji, skracanie dystansu między badaczem a interesariuszami programu wzmacnia dialog i ułatwia wypracowywanie wniosków i rekomendacji. Grozi jednak rozmyciem roli ewaluatora jako oceniającego krytycznie program i poddaniem się merytorycznemu dyktatowi zlecającego.

Po czwarte, ewaluacja pozostaje jedną z kluczowych (obok audytu i monitoringu) metod kontroli i oceny efektywności i wydajności (podstawy *Performance Budgeting & Management*) (OECD, 2005: 23). Jest jednak bardziej wartościowa niż audyt i monitoring. Umożliwia szersze spojrzenie wykraczające poza kontrolę procedur i poprawności finansowo-prawnej pojedynczych działań na rzecz oceny sensu, logiki całego systemu, a nawet całych polityk. Ewaluacja tworzy więc możliwości konsekwentnej oceny polityk publicznych – ich zróżnicowanych i wieloaspektowych efektów. Jest „spoiwem pozwalającym na stałe doskonalenie polityk, programów i projektów zorientowanych na rozwiązanie problemów społecznych" (Górniak, 2007: 11). Ponadto daje szansę przezwyciężenia jednej z głównych słabości PBM – koncentrowania się na krótkookresowej i wąskiej logice wskaźników produktów kosztem szerszej strategii[13]. Ewaluacja dostarcza informacji do dyskusji i powiązania decyzji o alokacjach z osiąganymi efektami. Tak więc ewaluacje jakościowe mogą wyjaśniać efekty polityk, programów i procesy stojące za sukcesem lub porażką danych interwencji. Ewaluacja rozszerza też kryteria oceny, uwzględniając prócz efektywności kosztowej aspekty środowiskowe i społeczne, kulturowe koszty i zyski. Wyzwanie z tym związane jest oczywiste – kryteria oceny

[13] Przeciwdziałanie ewaluacji dysfunkcjom zarządzania opartego na wynikach zostało omówione szerzej w Tabeli 4.

są bardziej miękkie i dużo mniej jednoznaczne niż w audycie czy monitoringu. Kwestia pomiaru i oceny odniesionego sukcesu programu jest po pierwsze bardziej względna (zależna od postrzegania różnych aktorów), a po drugie trudna do przeprowadzenia od strony metodologicznej (pomiar efektów długoterminowych, udowodnienie relacji przyczynowo-skutkowej między interwencją a jej skutkiem).

Wreszcie, analizując perspektywy użycia ewaluacji w nowoczesnym zarządzaniu w sektorze publicznym, należy podkreślić, że ewaluacja jest identyfikowana jako narzędzie uczenia się. Tym samym wpisuje się więc w obecną tendencję reform – poszukiwania innowacyjnych, skuteczniejszych rozwiązań, jak również w postulaty uczenia się organizacji (OECD, 1998). Dobrze prowadzone, jasno komunikowane badania ewaluacyjne dostarczają nie tylko bieżących informacji i rekomendacji, ale również przyczyniają się do budowania pamięci instytucjonalnej, wzmacniania powiązań międzyorganizacyjnych. Wyzwaniem idącym w parze z tą kwestią jest budowanie wiedzy instytucjonalnej instytucji publicznych przy jednoczesnym zlecaniu badań na zewnątrz nich (niezależnym ekspertom, firmom konsultingowym). Badania prowadzone bez interakcji, skoncentrowane wyłącznie na raporcie jako produkcie, wydają się mieć małe przełożenie na uczenie się tak ludzi, jak i instytucji zaangażowanych w oceniany program.

Podsumowując, ewaluacja jest jednym z elementów współczesnego zarządzania w sektorze publicznym. Wraz z ewolucją ujęć teoretycznych i paradygmatów zarządzania zmieniała się również jej rola – od narzędzia obiektywnej naukowej oceny jakości, weryfikowania relacji przyczynowo-skutkowej między interwencją a jej efektem, poprzez instrument poprawy bieżącego zarządzania i weryfikacji efektywności procesów i technikę szerszej analizy racjonalności interwencji publicznych aż po narzędzie wzmacniania sieci współpracy, budowania kapitału relacyjnego oraz uczenia się. Miejsce ewaluacji w ewolucji paradygmatów administracji i zarządzania w sektorze publicznym podsumowano w tabeli 7.

Tabela 7. Ewaluacja w paradygmatach zarządzania sektorem publicznym

Paradygmat	Główne założenia i cechy	Główna rola ewaluacji
Klasyczna Administracja Publiczna	Zasada jednego ośrodka decyzyjnego i hierarchii w podziale kompetencji Centralna rola biurokracji w formułowaniu i wdrażaniu polityk Zasada podziału pracy, specjalizacji Przyrostowe budżetowanie	Wzmocnienie odpowiedzialności Zweryfikowanie stopnia wypełnienia celów
Neoklasyczna Administracja Publiczna	Procesowa logika planowania i analizy interwencji publicznych Strukturalizowanie działań i funkcji w formie mierzalnych celów Tworzenie wieloletnich planów budżetowych Obiektywna analiza wdrażanych operacji oparta na badaniach i ewaluacji	Poprawa planowania Wzmocnienie odpowiedzialności
Menedżeryzm	Uzależnienie wielkości alokacji od planowanych produktów Wprowadzenie technik zarządzania z sektora prywatnego Zarządzanie oparte na wynikach Wprowadzenie konkurencji rynkowej do sektora publicznego Zmiana relacji z urzędnik – petent na menedżer – klient	Poprawa planowania Poprawa bieżącego wdrażania Poprawa wyników wdrażania (akcent na produkty) Wsparcie procesów uczenia się operatorów programu
Nowe Zarządzanie Sektorem Publicznym (*New Public Management*)	Zarządzanie oparte na wynikach Techniki zarządzania z sektora prywatnego i konkurencja rynkowa w sektorze publicznym Zmiana relacji z urzędnik – petent na menedżer – klient Zarządzanie jakością i akcent na jakość usług Dekoncentracja i decentralizacja	Poprawa planowania Poprawa bieżącego wdrażania Poprawa wyników wdrażania (w tym jakości efektów) Wsparcie procesów uczenia się
Nowe Rządzenie ***(New Public Governance)***	Zarządzanie oparte na wynikach Podejście systemowe, włączanie sektora prywatnego i trzeciego sektora (NGOs) Koordynacja i zarządzanie sieciami Aktywne włączanie obywateli Nowy system kadr sektora publicznego (elastyczność i otwartość)	Poprawa planowania i jakości wdrażania Wparcie procesów uczenia się Wzmacnianie partnerstwa Wzmacnianie odpowiedzialności

Źródło: opracowanie własne na podstawie (Dunsire, 1995; Ferlie et al., 2005; OECD, 1995, 2005; Osborne, 2006).

3.3. PRAKTYKA EWALUACJI W UNII EUROPEJSKIEJ[14]

Praktykę ewaluacji na obszarze obecnej Unii Europejskiej można analizować z dwóch perspektyw. Pierwszą są działania ewaluacyjne prowadzone przez administracje krajowe. Perspektywą drugą są ewaluacje prowadzone przez Unię Europejską. Oczywiście jest to podział nieco upraszczający. W praktyce wiele z interwencji unijnych jest realizowanych w krajach członkowskich przez instytucje państw członkowskich UE (przede wszystkim w polu funduszy strukturalnych), a co za tym idzie wymogi unijne ewaluacji są realizowane przez administracje krajowe. To uproszczenie zastosowano jednak w dalszej części tego rozdziału.

3.3.1. Ewaluacja w administracjach krajów członkowskich UE

Analizując doświadczenia i praktykę badań ewaluacyjnych (tzw. potencjał ewaluacyjny) administracji państw Unii Europejskiej, można podzielić je na cztery grupy (Bachtler, 2001; Bachtler et al., 2000; Olejniczak, 2005; Toulemonde, 2000). Do pierwszej grupy należą państwa, które wprowadzały ewaluację systemowo, bazując na swojej kulturze administracyjnej. W tych krajach ewaluacja pojawiła się w latach 70. i od początku była traktowana jako integralne, praktyczne narzędzie codziennej pracy sektora publicznego. Klasycznym przykładem jest Wielka Brytania. Obok niej kraje tej grupy to Holandia, Niemcy, Francja, Szwecja, w mniejszym stopniu Finlandia. Część z tych państw od początku traktowała ewaluację jako unikalne narzędzie. Inne, takie jak na przykład Szwecja, mając silną tradycję audytu, traktowały ewaluację jako uzupełniający, bardziej miękki aspekt kontroli.

Drugą grupę stanowią natomiast kraje, w których wiedza i praktyka ewaluacji miały początkowo słabe bądź też żadne poparcie w kulturze administracyjnej. Zostały zaszczepione z zewnątrz, były budowane często oddolnie (ze szczebla lokalnego i regionalnego na centralny) i dopiero z czasem rozwinęły się w efektywną praktykę krajową. Przykładami są administracje Irlandii, Belgii, Włoch (północnych) i Austrii, do których wymóg ewaluacji przyszedł wraz z realizowanymi tam programami Banku Światowego, a później Funduszy Strukturalnych. Ewaluacja była traktowana początkowo jako kolejna biurokratyczna powinność wobec zewnętrznych darczyńców. Jej użyteczność dla działań krajowych sektor publiczny odkrył z czasem i zinternalizował, skutecznie adaptując ewaluację do działań krajowych (przykład Irlandii jest tu wręcz wzorcem europejskim).

Grupą trzecią są kraje bez tradycji ewaluacyjnej, gdzie ewaluacja jak dotąd się nie przyjęła, mimo że istniała zewnętrzna presja jej stosowania (podobnie jak w przypadku grupy drugiej standard ewaluacji przyszedł z zewnętrznymi środkami pomocowymi). W Hiszpanii, Portugalii, a przede wszystkim Grecji ewaluacja, mimo, iż wykonywana (acz słabo i pobieżnie), była i pozo-

[14] W całej pracy zastosowano ujednolicone nazewnictwo. Dla uproszczenia nazwę Unii Europejskiej zastosowano zarówno do sytuacji sprzed 1993 roku, a więc do istniejących wtedy Wspólnot Europejskich, jak i do sytuacji po wejściu w życie Traktatu z Maastricht.

staje papierową rzeczywistością tworzoną na potrzeby administracji brukselskiej, a ignorowaną przez urzędników krajowych. Co więcej, praktyka ewaluacji nie rozprzestrzeniła się na fundusze i programy krajowe (te niepowiązane z funduszami strukturalnymi).

Ostatnią grupę stanowią nowe kraje członkowskie. Ich administracja dopiero od początku lat 90. zaczęła przestawiać się na nowoczesny sposób zarządzania (wieloletnie planowanie, budżety zadaniowe, strategie), głównie za sprawą pieniędzy pomocowych (unijnych, Banku Światowego, *USAID*). Wraz z programami pomocowymi szły wymagania procedur i specyficznego sposobu zarządzania – wśród nich również ewaluacja. Początkowo stosowana tylko dla pieniędzy pomocowych, wykonywana głównie przez firmy zagraniczne i wyłącznie na potrzeby darczyńców (przyjmująca ich perspektywę), stopniowo i selektywnie zaczęła być przejmowana przez krajowe instytucje publiczne i rodzimych wykonawców badań. Analizując rozwój ewaluacji w nowych krajach członkowskich, należy jednak podkreślić dwie kwestie. Po pierwsze, proces internalizacji był dotychczas bardzo powolny (bo długotrwały jest szerszy proces – zmiany sposobu myślenia w sektorze publicznym). Po drugie, przyspiesza on w ostatnich dwóch latach, czego głównym czynnikiem są fundusze strukturalne.

Znacznie odmienny podział od prezentowanego powyżej przedstawiają wnioski ze studium badania potencjału ewaluacyjnego w UE, przeprowadzonego przez CSEC (Malan, 2004). Według tego badania do grupy krajów o wysokiej kulturze ewaluacji należą: Finlandia, Wielka Brytania, Niemcy, Dania, Holandia, Szwecja, Francja, Włochy, Hiszpania, Polska. Krajami o słabym potencjale są Austria, Belgia, Bułgaria, Estonia, Irlandia, Malta, Słowacja, Słowenia. Zaś kraje bez kultury ewaluacji to: Litwa, Czechy, Węgry i Rumunia. Do tych wyników należy podchodzić jednak wyjątkowo sceptycznie. Całe badanie było oparte wyłącznie na badaniu kwestionariuszowym, wnioski wyciągano na podstawie 107 bardzo obszernych, przeładowanych kwestionariuszy (z czego 59 z UE15, a 48 z nowych krajów członkowskich). W przypadku niektórych krajów odpowiedzi były przyczynkowe (np. 2 ankiety z Irlandii, po 1 z Bułgarii, Słowacji, Malty). Zwrot ankiet ze strony wykonawców – ewaluatorów – był minimalny, a *gros* odpowiedzi (jak np. w przypadku Polski) pochodziła od urzędników sektora publicznego – a więc osób znających tylko jedną stronę praktyki ewaluacyjnej, zainteresowanych korzystnym i nieco idealistycznym przedstawianiem możliwości i praktyki ewaluacyjnej ich własnych organizacji. Wnioski na temat całej UE zostały więc oparte na przyczynkowych, jednostronnych materiałach, bez próby weryfikacji na podstawie innych źródeł, na przykład użycia takich źródeł, jak przegląd jakości krajowych raportów ewaluacyjnych, liczby rzeczywiście opublikowanych i dostępnych raportów on-line, wywiadów w samej Komisji Europejskiej czy przeglądu wcześniejszych badań. Przedstawiały jednostronną, często wręcz personalną i urzędniczą ocenę sytuacji całego kraju. W związku z tym wiarygodność, a co za tym idzie wartość poznawcza tego badania jest znikoma.

W Polsce kultura ewaluacji tworzy się od połowy lat 90., głównie za sprawą wymagań funduszy strukturalnych. Jest to element szerszego, mozolnego

procesu – wprowadzania w sektorze publicznym zarządzania opartego na wynikach[15]. W najbliższych latach intensyfikacji ulegnie prawdopodobnie proces internalizacji ewaluacji w programach krajowych – ewaluacja poprzez fundusze strukturalne będzie się rozprzestrzeniać na rodzime polityki i programy[16].

Obserwując obecną polską sytuację rynku ewaluacji, można zidentyfikować kilka wyzwań. Po stronie popytu dla instytucji publicznych głównymi wyzwaniami są:

- zrozumienie odmienności ewaluacji od audytu, monitoringu i kontroli,
- zachowanie realizmu przy kreśleniu oczekiwań odnośnie badań ewaluacyjnych (powiązanie pytań ewaluacyjnych i problemów do badania z budżetem i z czasem przeznaczonym na badanie),
- otwarcie się na konstruktywną krytykę, chęć współpracy i dialogu z badaczami.

Jeszcze do niedawna podstawową bolączką polskiego rynku było oferowanie znikomych budżetów na badania połączone z używaniem ceny (a nie jakości) jako głównego kryterium wyboru ofert. W ostatnich dwóch latach ten problem powoli znika.

Po stronie podaży (dla wykonawców badań ewaluacyjnych) wyzwaniami są:

- sprostanie standardom naukowym i jednoczesne zapewnienie wysokiej jakości raportów (przede wszystkim ich komunikatywności),
- używanie w badaniach metod interaktywnych,
- wychodzenie poza utarte schematy badawcze i czerpanie z dobrych praktyk europejskich, z uwzględnieniem specyfiki polskiej.

Wciąż widoczny jest także brak ekspertów ewaluacyjnych, wykonywanie badań po kosztach i ograniczanie się do statycznych, jakościowych metodologii.

Patrząc całościowo, podstawowe trzy kwestie, które zaważą na przyszłości ewaluacji w Polsce można zdefiniować następująco. Pierwszą jest brak wiarygodnych i kompletnych danych, przede wszystkim informacji kontekstowych i bazowych. Słabo działający system monitoringu sprawia[17], że dotychczasowe badania ewaluacyjne koncentrowały się na uzupełnianiu danych, a przy analizie korzystały głównie z cząstkowych danych pierwotnych (zebranych w trakcie kontraktów). Przy ewaluacjach zorientowanych na planowanie lub na procesy (które dotychczas przeważały) problem danych nie był tak dotkliwy i wyraźny. Jednak dla nadchodzącej fali ewaluacji *ex post* problem ten okazać się może kwestią zasadniczą – nie do pokonania. Brak danych z początków procesu inwestycji, a więc bazy do porównań, grozi tym, że ewaluacje

---

[15] Szerzej o historii ewaluacji w Polsce i etapach tworzenia się polskiej kultury ewaluacji od lat 90. do 2002 roku patrz: Kierzkowski (2002a); Olejniczak (2002).

[16] Próby takie podejmowano już na początku 2000 roku, gdy ewaluację (ocenę) wprowadzono jako prawny wymóg kontraktów regionalnych. Ponieważ jednak kontrakty miały roczny charakter, a odpowiedzialność za ewaluację i jej zakres były słabo sprecyzowane, pozostały martwym zapisem.

[17] Rdzeń systemu monitoringu europejskiej polityki regionalnej w Polsce – SIMIK – nigdy w pełni nie został uruchomiony.

efektów (rzeczywistych rezultatów i oddziaływania) w praktyce nie przyniosą nic ponad opis produktów. Tym samym nie będziemy mogli podjąć refleksji nad strukturalną absorpcją funduszy unijnych i ich użytecznością dla rozwoju kraju[18].

Drugą bolączką jest brak rodzimej refleksji na temat metod ewaluacyjnych i brak zorientowanej praktycznie literatury naukowej z tego zakresu. Jak dotąd wątpliwą rolę „ram metodologicznych" wypełniały wytyczne Komisji Europejskiej, a z czasem tworzone przez poszczególne ministerstwa tłumaczenia komisyjnych, lub własne poradniki ewaluacyjne. Wszystkie one są jednak dokumentami administracyjnymi i biurokratycznymi, oferującymi podstawowe definicje i uproszczone szablony ewaluacji „krok po kroku". Z tym brakiem refleksji metodologicznej idzie w parze chaos terminologiczny (Szlachta, 2007: 8)[19].

Trzecią kwestią, która wymaga jednak weryfikacji empirycznej, jest rzeczywiste i kumulatywne wykorzystywanie wiedzy z ewaluacji. Patrząc na obecny system, zaskakujące i symptomatyczne jest ograniczone przenoszenie doświadczeń z funduszy przedakcesyjnych do funduszy strukturalnych. Wydaje się, że wiedza zgromadzona dzięki funduszom pomocowym przed przystąpieniem Polski do UE nie znalazła zastosowania po 2004 roku, co więcej, nie zachowała się pamięć o niej (Kozak, 2007). Instytucje odpowiedzialne za fundusze strukturalne w momencie naszej akcesji do UE uczyły się zarządzania, w tym ewaluacji, od zera. Jeśli taki stan rzeczy potwierdziłby się, oznaczałoby to, że tworzona w badaniach ewaluacyjnych wiedza jest wiedzą doraźną i nie tworzy kapitału do szerszej, długookresowej refleksji[20].

Szerszym wyzwaniem jest budowanie krajowego potencjału ewaluacyjnego (Szlachta, 2007b: 86): sięganie do doświadczeń i koncepcji międzynarodowych, podejmowanie tematyki ewaluacji w badaniach naukowych, wspieranie studiów podyplomowych z zakresu ewaluacji („kuźni kadr" zarówno przyszłych zlecających, jak i wykonawców) oraz transfer doświadczeń w ewaluacji i zarządzanie z segmentu programu europejskich na środki stricte krajowe.

Podsumowując przyczyny upowszechniania się ewaluacji w krajach Unii Europejskiej, można wymienić trzy czynniki. Pierwszym i podstawowym jest

---

[18] Termin „absorpcji strukturalnej" wprowadził Hauser (2007). Zdefiniował ją jako zdolność do używania funduszy pomocowych w sposób strategiczny, zorientowany na rozwój. Terminowi temu przeciwstawił pojęcie „absorpcji technicznej" – czyli zdolności zagospodarowania unijnych środków zgodnie z procedurami i według określonego harmonogramu.

[19] Przykładami mogą być dwa różne tłumaczenia poradników zarządzania cyklem projektów (jeden wydany przez UKIE i drugi, rozpowszechniany przez MG) czy pozbawiony wszelkich odniesień bibliograficznych podręcznik KJO (2005). Najbardziej rażący *casusem* chaosu terminologicznego jest, przytaczane już w tej książce, tłumaczenie terminu ewaluacja jako „ocena". Pierwszą próbą przezwyciężenia tego stanu rzeczy jest najnowsza publikacja PARP (Haber, 2007). Jest ona o tyle przełomowa, że nie tylko rozpoczyna dyskusję nad kwestiami metodologii badań ewaluacyjnych, ale wybiega w przyszłość, podejmując temat nadchodzących ewaluacji *ex post*. Szansą rozwoju w polu metodologii może się też okazać rozpoczęty pod koniec 2007 roku projekt UKIE, dotyczący wzmocnienia zdolności ewaluacyjnych w sektorze publicznym. Finansowanych jest z niego kilka podręczników ewaluacyjnych, przygotowywanych przez różne zespoły autorskie.

[20] Pierwszą próbę empirycznego zbadania wykorzystania ewaluacji w polskim sektorze publicznym podjęto ostatnio w ramach grantu „Sprawne Państwo": www.sprawnepanstwo.pl.

oczywiście konieczność racjonalizowania (a w praktyce najczęściej ograniczanie) wydatków publicznych. Drugim czynnikiem jest wytworzenie się pewnych ogólnoświatowych wzorców zarządzania w sektorze publicznym krajów rozwiniętych (silnie promowanych przez OECD i Bank Światowy), których elementem jest ewaluacja. Wreszcie trzecim czynnikiem, najsilniejszym w przypadku nowych krajów członkowskich, jest standaryzacja narzucana w ramach UE, wymagana przez jej programy, polityki i procedury funduszy strukturalnych.

Patrząc na wszystkie administracje państw UE można stwierdzić, że obecnie w sposób bardzo wyraźny ewaluacja staje się standardem nowoczesnej administracji publicznej. Oczywiście proces adaptacji przebiega odmiennie – bo inne są kultury administracyjne państw członkowskich. Wyraźnie jednak widać, że ewaluacja jest internalizowana i rzeczywiście używana w administracjach na północy kontynentu. Na południu wciąż jest często papierową rzeczywistością tworzoną na potrzeby funduszy strukturalnych. Nowe kraje członkowskie stoją na rozdrożu – między „północnym" a „południowym" podejściem do ewaluacji.

3.3.2. Ewaluacja programów współfinansowanych przez UE

W politykach Unii Europejskiej (ówczesnych Wspólnot) ewaluacja pojawiła się na początku lat 80. Co ciekawe (acz typowe dla procesów reformatorskich w UE), pojawiała się najpierw w politykach o stosunkowo najmniejszych budżetach i małej wadze politycznej. Pierwsze po ewaluację sięgnęły w Komisji Europejskiej DG-Stosunki Zewnętrzne (i późniejsze zintegrowane biuro EuropeAid) i DG-Badania. Szukając narzędzi bardziej efektywnego ukierunkowania prowadzonych przez siebie programów, zaczerpnęły z doświadczeń Banku Światowego i OECD. Symptomatyczne, iż w przypadku polityki rolnej ewaluacja pojawiła się dopiero w latach 90.

Analizując genezę ewaluacji w programach Unii Europejskiej, wyróżnić można trzy główne etapy jej wprowadzania do głównego nurtu polityk unijnych. Pierwszą przyczyną był skokowy wzrost wydatków oraz związana z nim konieczność kontroli i bardziej racjonalnego wykorzystania pieniędzy unijnych. Momentem przełomowym była tu reforma budżetowa Delorsa z 1987 roku (tzw. I Pakiet Delorsa), a głównym polem dyskusji polityka regionalna, której budżet został dwukrotnie zwiększony (Dinan, 1999: 433)[21]. Przy tak znaczącym wzroście wydatków państwa płatnicy netto i Komisja Europejska (wiedząc, że *gros* środków będzie skierowana do regionów Włoch, Irlandii, Grecji, Hiszpanii i Portugali) szukały odpowiednich narzędzi, które zapewniłyby kontrolę zasadności tak znaczącej interwencji publicznej oraz wymusiłyby jej skuteczność. Sięgnięto więc po ewaluację (bazując na doświadczeniach krajów północnej Europy) i wpisano ją jako jeden z wymogów systemu zarządzania funduszami strukturalnymi obok spójnego systemu monitoringu, podejścia opartego na programach i wieloletnim planowaniu. Obowiązkiem ewaluacji *ex*

[21] Na lata 1988–1993 przeznaczono na fundusze strukturalne 60 mld ECU. Wydatki te sukcesywnie zwiększano i tak w 1999 roku roczny budżet polityki spójności UE wyniósł ok. 30 mld euro.

ante i *ex post* obarczono Komisję Europejską oraz instytucje państw członkowskich zarządzające danymi programami unijnymi. Ewaluacja po raz pierwszy pojawiła się w kontekście dużych europejskich pieniędzy – funduszy na politykę regionalną (porównaj: Council of the European Economic Community, 1988: art. 6)[22]. Ówczesna praktyka ewaluacji w tamtym okresie była ograniczona z kilku powodów. Po pierwsze, zapisy na temat monitoringu i ewaluacji były bardzo ogólne i przez to różnie interpretowane w poszczególnych państwach (najczęściej w sposób ograniczony). Po drugie, całe podejście do programowania wieloletniego i interwencji regionalnych było *novum* – administracje i Komisja dopiero się uczyły i eksperymentowały. Po trzecie, możliwości badań były ograniczone z uwagi na duże opóźnienia w uruchamianiu programów (Bachtler, Taylor, 1999: 228–229).

Drugim etapem zwiększania roli ewaluacji był początek lat 90. Dobiegał końca pierwszy okres programowania budżetowego, finalizowano dyskusję nad Traktatem z Maastricht mającym otworzyć kolejny etap integracji europejskiej – Unię Europejską. Instytucje Wspólnot potrzebowały spójnej i całościowej informacji na temat wdrażania programów wspólnotowych i ich efektów (Bachtler, Taylor, 1999: 229). W nowym Rozporządzeniu położono więc silniejszy nacisk na monitoring oraz ewaluacje *ex ante* (nazwaną *ex ante apprisal*) i *ex post* (porównaj: Council of the European Economic Community, 1993: art. 26). Praktyka ewaluacji znacznie poprawiała się – administracje zdążyły oswoić się z nowym narzędziem. Wpływ miały też działania Komisji Europejskiej – wprowadziła ona bardziej skoordynowane podejście do ewaluacji, tworząc specjalny departament ewaluacyjny w Dyrekcji Generalnej ds. Polityki Regionalnej (DG-Regio) oraz zaczęła budować europejską kulturę ewaluacji w ramach programu MEANS[23].

Wreszcie trzecim momentem skokowego wzrostu znaczenia ewaluacji był koniec lat 90., a przyczyną konieczność podniesienia wiarygodności instytucji europejskich w oczach obywateli. Po spektakularnym kryzysie Komisji Santera rozpoczęto reformę Komisji Europejskiej. Jej głównymi architektami byli Anglosasi, którzy opierali się na anglosaskim podejściu do zarządzania w sektorze publicznym, a następnie wtłaczali je w typowo francuską strukturę Komisji. Celem reformy było podniesienie przejrzystości działań instytucji unijnych, zapewnienie ich racjonalności, jakości, sprawności, poszanowania wartości pieniądza europejskiego podatnika. Wymóg ewaluacji, obok wzmocnionych procedur audytu i monitoringu oraz strategicznego planowania został

[22] W tym pierwszym rozporządzeniu na temat polityki regionalnej UE ewaluacja jest określana angielskim terminem *assessment* (porównaj art. 6 cytowanego rozporządzenia). Z czasem (w kolejnym Rozporządzeniu Rady 2082/93 na lata 1994–1999) w Komisji zaczęto używać terminu *evaluation*.

[23] Program MEANS (Evaluation Methods for Actions of a Structural Nature) został uruchomiony przez Dyrekcję ds. Polityki Regionalnej (DG-Regio) w 1994 roku. Opierał się na serii konferencji, seminariów oraz publikacji, które służyły koordynacji, wymianie doświadczeń, wypracowywaniu wspólnych standardów dla praktyk ewaluacji programów strukturalnych w różnych państwach członkowskich. Program ten zaowocował specjalnym podręcznikiem wydanym w 1999 roku (MEANS 1999). Kolejne edycje z 2003 i 2006 roku (*Evalsed* i *Evalsed Update*) zostały już opracowane w formie strony internetowej.

przedstawiony jako jeden z kluczowych narzędzi reformy – element *New European Governance* (European Commission, 2000c, d, 2001e, f)[24]. Jednocześnie wszystkie instytucje UE zostały zobowiązane do prowadzenia ewaluacji dla wszystkich inicjatyw współfinansowanych w znaczący sposób ze środków UE (Council of the European Union, 2002: art. 27.4). Wymóg ewaluacji został również powtórzony w kolejnym, trzecim już rozporządzeniu dotyczącym funduszy strukturalnych – tym razem na lata 2000–2006 i z rozszerzeniem o trzeci obowiązkowy typ ewaluacji (obok *ex ante* i *ex post*) – ewaluację średniookresową (Council of the European Union, 1999: art. 40–43). Zaś szczególnym zainteresowaniem Komisji zaczęły się cieszyć ewaluacje dużych projektów przede wszystkim z zakresu infrastruktury i ochrony środowiska)[25]. Regulacje te sprawiły, że liczba badań ewaluacyjnych zaczęła przyrastać w tempie lawinowym, a w literaturze pojawiło się określenie *evaluation boom* (Pollitt, 1998). Na przestrzeni ostatnich dziesięciu lat Komisja Europejska zleciła lub wykonała w ramach swoich departamentów ponad 1500 badań (patrz Tabela 8). Natomiast liczba ewaluacji wykonanych dla programów z funduszy strukturalnych, dla samych krajów UE15, wynosiła ponad 2000 badań (patrz Tabela 9). Do tego należy dodać tzw. ewaluacje bieżące zlecane przez krajowe ministerstwa na zasadzie *ad hoc* (a więc niepodlegające scentralizowanej, unijnej statystyce) dotyczące poszczególnych działań, priorytetów, kwestii tematycznych. Są ich setki. W samej Irlandii, dla celu 1 tylko dla lat 1994–1999 było ich około 150.

Tabela 8. Liczba ewaluacji zlecanych przez Komisję Europejską

Pole działań[26]	1996–1997	1998	1999	2000	2001	2002	2003	2004	2005	Suma
Rolnictwo	9	2	10	4	5	5	2	9	9	55
Fundusze Strukturalne	4	11	6	8	2	5	6	15	3	60
Polityki wewnętrzne	61	26	39	33	36	36	63	140	138	572
Działania zewnętrzne	47	37	25	37	34	22	26	57	21	306
Administracja	0	1	1	0	1	2	7	14	12	38
Inne	0	0	0	0	10	7	13	9	3	42
Suma	121	77	81	82	88	77	117	244	186	1073

Źródło: obliczenia własne na podstawie (European Commission, 1999a, 2000b, 2001a, 2002a, 2003, 2005a, b, 2006).

[24] Co prawda prace nad podnoszeniem jakości zarządzania, przede wszystkim finansowego, w administracji europejskiej rozpoczęto już w połowie lat 90., jednak program SEM 2000 (*Sound and Efficient Managemenent* 2000 – Przejrzysty i skuteczny system zarządzania) stał się priorytetem dopiero po wspomnianym już kryzysie Komisji.

[25] Fakt ten można prosto wytłumaczyć wagą takich projektów (finansową i strategiczną), ale także stosunkową łatwością badania – przy dużych inwestycjach infrastrukturalnych wskaźniki produktu, rezultatu, a nawet oddziaływania są relatywnie jasne, a skala inwestycji sprawia, że można przypisywać nawet szersze zmiany danego układu (transportowego, środowiskowego a nawet społeczno-gospodarczego) wpływowi interwencji unijnej.

Tabela 9. Ewaluacje polityki regionalnej UE w latach 1994–2006 (tylko UE15)

	Cel 1*	Cel 2	Inicjatywy Wspólnotowe**	Suma
Rodzaj ewaluacji	Okres programowania 1994–1999			
Ewaluacje *ex ante*	147	80	299	526
Ewaluacje średniookresowe	140	72	0	212
Ewaluacje *ex post*	18	25	4	47
Suma dla lat 1994–1999	305	177	303	785
Rodzaj ewaluacji	Okres programowania 2000–2006			
Ewaluacja *ex ante*	100	95	232	427
Ewaluacje średniookresowe	100	95	232	427
Ewaluacje *ex post****	100	95	4	199
Uaktualnienie ewaluacji średniookresowej	100	95	232	427
Suma dla lat 2000–2006	400	380	700	1480
Ogółem dla lat 1994–2006	705	557	1003	2265

* W przypadku 1994–1999 również Cel 6; ** INTERREG, Urban, Leader, Equal; *** zaplanowane ewaluacje.
Źródło: obliczenia własne na podstawie dokumentów programowych.

Jeśli uwzględnić w tej całościowej statystyce nowe kraje członkowskie i początek nowego okresu programowania funduszy strukturalnych (2007–1913, ewaluacje *ex ante* wykonywane w 2006), to liczbę ewaluacji programów UE w ostatnich dziesięciu latach można szacować na ponad 4,5–5 tysięcy badań. Ta skala zjawiska sprawia, że jeden z głównych odbiorców badań – Komisja Europejska zaczyna zlecać metaewaluacje (podsumowujące w różnych aspektach wyniki i wnioski ewaluacji robionych na poziomie krajowym), starając się ogarnąć tę rosnącą „piramidę raportowania".

Podsumowując obecną sytuację, należy zwrócić uwagę na następujące kwestie. Ewaluacja jest obecnie prawnie usankcjonowanym wymogiem w odniesieniu do wszystkich interwencji współfinansowanych z budżetu UE. Największym jej polem, z uwagi na zdecentralizowany charakter interwencji, jest polityka regionalna UE – druga pod względem wielkości budżetu, a jedna z najważniejszych strategicznych polityk UE. Tak więc ewaluacje (obowiązkowo *ex ante* i *ex post*, a dodatkowo bieżące) są wymogiem systemu zarządzania funduszami strukturalnymi (Council of the European Union, 2006: art. 47–49).

Warto też wspomnieć o głównych aktorach procesu ewaluacji – instytucjach zlecających i wykonawcach. Na poziomie instytucji UE głównymi zleceniodawcami i odbiorcami studiów ewaluacyjnych jest Komisja Europejska – jej poszczególne Dyrekcje. Szczególną rolę odgrywają trzy dyrekcje. DG-Budget z racji tego, że nadzoruje, koordynuje i promuje działania ewaluacyjne w ramach Komisji (w odniesieniu do badań zlecanych przez KE). Drugą jest DG-Regio opiekująca się polityką regionalną UE, a więc największym polem tematycznym ewaluacji. Wreszcie trzecią instytucją jest biuro EuropeAid (w ramach DG-Stosunki Zewnętrzne) z racji tego, że nadzoruje szeroką pulę

programów pomocowych, tzw. programy pomocy zewnętrznej (program TACIS – dla Ukrainy, Rosji, dawnych republik radzieckich, CARDS – dla Bałkanów, MEDA – dla rejonu Morza Śródziemnego, ECHO – dla Afryki). Oprócz Komisji ewaluację wykorzystuje także (głównie jako „amunicję" do debat politycznych) Parlament Europejski. Trzeba również wspomnieć o badaniach prowadzonych przez Trybunał Audytorów. Mimo iż są to audyty, to jednak w przypadku audytów programów europejskich stawiają one również pytania z zakresu ewaluacji (o skuteczność działań, efektywność procedur, etc.).

Natomiast w odniesieniu do funduszy strukturalnych, głównymi zlecającymi i odbiorcami ewaluacji są administracje krajów członkowskich – konkretne ministerstwa i agencje zarządzające poszczególnymi programami, działaniami, projektami współfinansowanymi z funduszy europejskich. System organizacyjny zależy od kraju, ale coraz bardziej powszechną praktyką w krajach europejskich są Jednostki Ewaluacyjne na poziomie ministerstwa centralnego, które koordynują w danym kraju zlecanie badań ewaluacyjnych, promocję ich wyników, identyfikowanie i promowanie najlepszych praktyk badawczych. W Polsce instytucją taką jest Krajowa Jednostka Oceny[26] w ramach Ministerstwa Rozwoju Regionalnego. Ważną instytucją jest również Polska Agencja Rozwoju Przedsiębiorczości zlecająca szereg badań ewaluacyjnych (głównie w odniesieniu do projektów krajowych) i inicjująca publikacje z zakresu ewaluacji (tak praktyki, jak i podejść metodologicznych). Do pewnego czasu ważną instytucją był również Urząd Komitetu Integracji Europejskiej, zlecający badania ewaluacyjne funduszy pomocowych i przedakcesyjnych, robiący przekłady anglojęzycznych podręcznikików Komisji Europejskiej.

W przypadku gdy badanie ewaluacyjne nie jest wykonywane wewnętrznie (a takie badania są nieliczne, gdyż główną praktyką stanowią ewaluacje zewnętrzne), wykonanie badań jest zlecane w formie przetargów otwartych na usługi. Przetargi, w zależności od skali programu i jego charakteru, są ogłaszane w krajowych serwisach zamówień publicznych, bądź też na stronach europejskich TED (http://ted.publications.eu.int/official/).

Wykonawcami badań są firmy konsultingowe lub instytuty naukowo-badawcze, uniwersytety, a zwykle konsorcja łączące te instytucje. Badania ewaluacyjne dają możliwość badań *quasi*-naukowych na dużą skalę (gdyż budżety przeznaczone na te przedsięwzięcia są duże), dają możliwość badań porównawczych w skali europejskiej, stanowią również znaczącą i atrakcyjną część rynku konsultingowego w Europie.

3.3.3. Praktyka ewaluacji – główne podejścia metodologiczne UE

W praktyce programów UE wykształciło się kilka nieco odmiennych podejść metodologicznych do ewaluacji[27]. Wyróżnić można trzy główne nur-

[26] http://www.funduszestrukturalne.gov.pl/EWALUACJA+funduszy+w+Polsce/Krajowa+Jednostka+Oceny/.

[27] Różnice w podejściu dotyczą najczęściej terminów (co po części wynika z tradycji danej dyrekcji, a po części z niechęci do współpracy z innymi dyrekcjami), rozłożenia akcentów (wynikającego z odmiennej tematyki poddawanej badaniu), formy badania (specyfika poszczególnych

ty. Pierwszym są praktyki DG-Badania i nauka, gdzie ewaluacja oparta jest, z uwagi na specyfikę dziedziny, na tzw. *peer assessment* (czyli ocenie przez panel naukowców – ekspertów uczestniczących w danym programie). Metodologie te są stosowane jednak głównie do wyboru projektów badawczych. Odnosząc je do przyjętej w tej pracy definicji ewaluacji, nie są ewaluacją *sensu stricte*, lecz formą oceny projektów.

Drugim jest podejście EuropeAid oparte na Zarządzaniu Cyklem Projektu (ang. *Project Cycle Management*). Jest stosowane przy programach pomocy zewnętrznej, takich jak Tacis, Meda, Echo, Cards, a do niedawna również Phare. Metodologia ewaluacji promowana przez tę dyrekcję jest prezentowana jako element procesu zarządzania cyklem projektu, stąd główne narzędzia analizy to drzewa problemów i celów, SWOT, matryce logiczne (ang. *logframe*). Promowane podręczniki to kolejne edycje Poradnika ZCP oraz szablony metod na stronie EuropeAid (porównaj: European Commission, 1993, 2001b, c, d, 2004b)[28].

Wreszcie trzecia, wiodąca metodologia to wspominany już MEANS (*Evaluation Methods for Actions of the Structural Nature*), rozwijany od początku lat 90. przez DG-Regio. Początkowo wydawany w formie zeszytów metodycznych (porównaj: European Commission, 1995a, b, c, d, e, f, g), w 1999 roku wydany jako pakiet sześciu podręczników (European Commission, 1999b), a od 2003 roku funkcjonujący pod nazwą *Evalsed* w formie strony internetowej[29]. *Evalsed* zawiera kompleksowy materiał na temat funkcji ewaluacji w programach publicznych, struktury typowego badania z podziałem na etapy badawcze, zasady wyboru wskaźników monitoringu i ewaluacji, a także doboru narzędzi badań, opisy i sposób zastosowania głównych metod i technik (z podziałem na typy ewaluacji i tematykę programów), sposoby uwzględnienia tematów horyzontalnych (ochrona środowiska, zatrudnienie, równe szanse kobiet i mężczyzn). *Evalsed* jest wykorzystywany przy funduszach strukturalnych, a także w działaniach koordynowanych przez DG-Rolnictwo, DG-Zatrudnienie czy DG-Przedsiębiorczość. Prócz poradnika *Evalsed*, DG-Regio wydała też szereg zeszytów metodycznych (tzw. wytycznych) dotyczących prowadzenia różnych typów ewaluacji[30].

Należy zaznaczyć, że podejścia EuropeAid i DG-Regio są zbieżne, mogą być łączone, a często się przenikają. Różnice dotyczą najczęściej terminów (co po części wynika z tradycji danej dyrekcji, a po części z niechęci do współpracy z innymi dyrekcjami), rozłożenia akcentów (wynikającego z odmiennej tematyki poddawanej badaniu), specyfiki i skali pól interwencji (EuropeAid zajmuje się pomocą głównie o charakterze humanitarnym i koncentruje na poziomie projektów, DG-Regio prowadzi pomoc strukturalną i politykę regional-

polityk oczywiście narzuca formę ewaluacji), jak i przewidywanej roli raportów (np. raporty dla EuropeAid mają silne przełożenie praktyczne – program jest kontrolowany, może być wstrzymany, przy funduszach strukturalnych raporty dają raczej obraz sytuacji i są punktem wyjścia do dyskusji: Komisja-administracja krajowa lub do dużych debat politycznych).

[28] Patrz też: http://ec.europa.eu/europeaid/how/evaluation/index_en.htm.

[29] *Evalsed* funkcjonował pod adresem: www.evalsed.com. W 2007 roku stworzono jego nową wersję (uwzględniającą rozszerzenie UE) i przeniesiono na stronę DG-Regio – Inforegio: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/index_en.htm.

[30] Porównaj: http://ec.europa.eu/regional_policy/sources/docoffic/working/sf2000_en.htm.

ną głównie na poziomie programów), przewidywanej roli raportów. Przykładowo raporty dla EuropeAid mają silne przełożenie praktyczne – projekt jest kontrolowany, może być wstrzymany, natomiast przy funduszach strukturalnych raporty dają raczej obraz sytuacji i są punktem wyjścia do dyskusji Komisja–administracja krajowa lub do dużych debat politycznych).

Inne dyrekcje również wydają swoje poradniki metodyczne, pokrywające się w strukturze i ogólnej logice z *Evalsed*, a jedynie uściślające kwestie charakterystyczne dla danego pola interwencji czy sugerujące metodologię dla konkretnych programów[31].

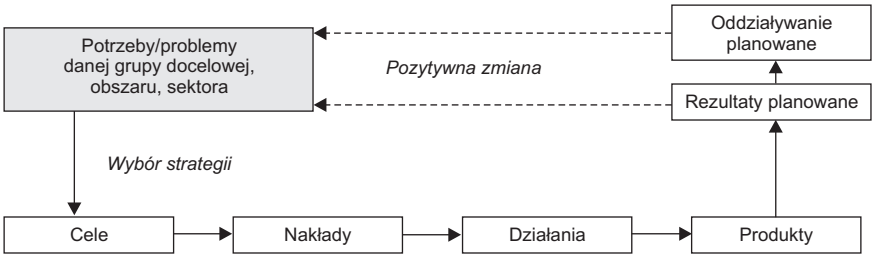
W niniejszej książce skoncentrowano się na podejściu *Means*/*Evalsed*, po pierwsze, jako najbardziej dominującym i rozpowszechnionym w UE, a po drugie, odpowiadającym analizowanym programom (współpraca terytorialna to element polityki regionalnej UE). Należy jednak podkreślić, że wszystkie materiały Komisji Europejskiej mają charakter poradników technicznych, a nie książek akademickich i publikacji *stricte* metodycznych. Nie mogą więc służyć jako jedyne źródło inspiracji badawczych.

Niezależnie od szczegółowego podejścia metodologicznego ewaluacja programów Unii Europejskiej opiera się na trzech kluczowych elementach:

- procesowym podejściu do interwencji publicznej,
- wkomponowaniu ewaluacji w cykl interwencji publicznych,
- użyciu zestawu standardowych kryteriów do oceny programów (tzw. kryteriów ewaluacyjnych).

Pierwszy element to **podejście procesowe**, ma ono swe korzenie w nowoczesnym zarządzaniu sektorem publicznym (porównaj rozdział 3.2). Zakłada, że każdą interwencję możemy zaplanować w całościowy sposób, w ciągu logicznym przedstawionym na poniższym schemacie.

Schemat 2. Logika planowania interwencji UE



Źródło: opracowanie własne na podstawie: (European Commission, 1999b, 2000e, 2001c).

Według tego schematu przebiega również wdrażanie i monitoring. Poszczególnym działaniom i celom przyporządkowane są wskaźniki. Cały ciąg logiczny wraz z definicjami został przedstawiony na kolejnym schemacie. Wszystkie interwencje, niezależnie od skali i tematyki, są układane według takiego schematu logicznego.

[31] Porównaj np.: http://ec.europa.eu/employment_social/evaluation/guidance_en.html DG Employment czy DG-Agriculture: http://ec.europa.eu/agriculture/eval/index2_en.htm#guide.

Schemat 3. Logika wdrażania i monitorowania interwencji UE

Hierarchia celów	Hierarchia efektów	Hierarchia wskaźników
Cel globalny, ogólny (ang. *Global objective*) Cel programu, do którego aplikuje dana interwencja – priorytet, do którego jej realizacja będzie się przyczyniać	**Oddziaływanie (ang. *Impact*)** Długookresowe skutki interwencji oddziaływujące na szerszą populację i szerszy obszar (np. na cały region, jego mieszkańców i środowisko naturalne) Pojawiają się dopiero po pewnym czasie od zamknięcia interwencji (zwykle po 12, 18, 36 miesiącach), ale można je bezpośrednio powiązać z podjętymi działaniami	**Wskaźniki oddziaływania (ang. *Impact indicators*)** Opisują konsekwencje interwencji, które wychodzą poza jej bezpośrednie i natychmiastowe działanie na bezpośrednich beneficjentów Mają zastosowanie w perspektywie średnioterminowej
↑	↑	↑
Cel bezpośredni projektu (ang. *Specific objective*) Sens istnienia tej interwencji Bezpośredni powód, dla którego podjęto interwencję	**Rezultaty (ang. *Results*)** Bezpośredni skutek interwencji (korzyści i ułatwienia) dla jej bezpośrednich odbiorców Pełen rezultat jest widoczny dopiero po zamknięciu wszystkich działań w ramach interwencji	**Wskaźniki rezultatu (ang. *Results indiators*)** Przedstawiają informacje na temat zmian zachodzących w odniesieniu do bezpośrednich odbiorców, beneficjentów
↑	↑	↑
Cele operacyjne (ang. *Operational objectives*) Cele określające produkt (infrastrukturę, rzeczy, wartości), jakie mają zostać wypracowane w czasie interwencji	**Produkty (ang. *Products*)** Wartości, rzeczy, elementy infrastruktury, które zostały wytworzone w czasie projektu	**Wskaźniki produktu (ang. *Products indicators*)** Opisują produkt (w jednostkach fizycznych) działalności osób, instytucji realizujących działania
↑	↑	↑
Nakłady i ich wskaźniki (ang. *Input and input indicators*) Zasoby i środki, udostępnione przez władze finansujące oraz wykorzystane przez instytucje realizujące projekt, potrzebne do dostarczenia produktu i realizacji celu bezpośredniego		
Potrzeby i problemy, do rozwiązania których interwencja ma się przyczynić.		

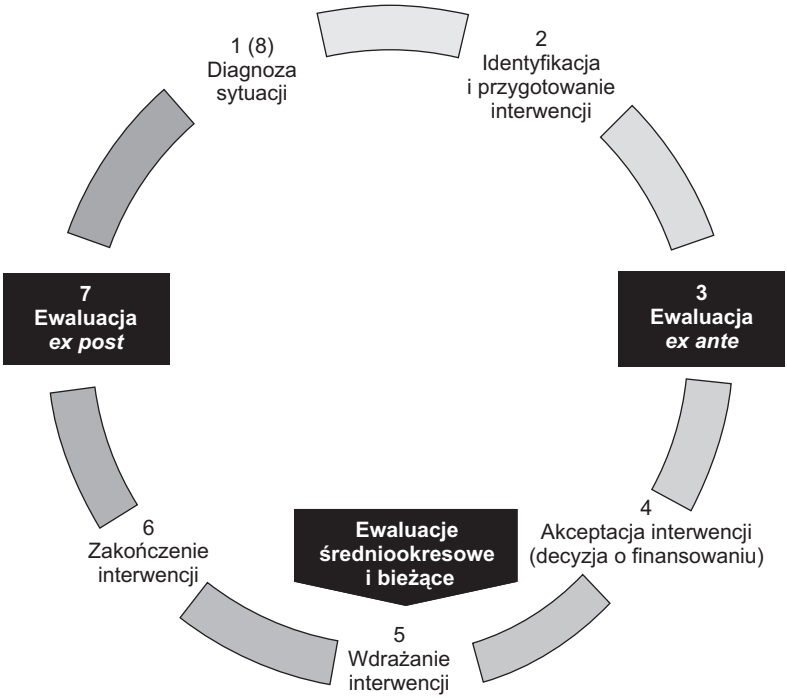
Źródło: opracowanie własne na podstawie European Commission (1999b, 2000e, 2001c)[32].

Drugim kluczowym elementem unijnego podejścia jest wpisanie ewaluacji w **cykl planowania i wdrażania interwencji**, a konkretnie przypisanie każdemu z typów ewaluacji szczególnej funkcji (patrz: schemat poniżej). I tak, w założeniu ewaluacja *ex ante* ma pomóc na początku – w lepszym zaplanowaniu interwencji, ewaluacje średniookresowe i ich uaktualnienie oraz ewaluacje bie-

[32] Niniejsza tabela została użyta w publikacji Gorzelak i współpracownicy (2006).

żące pomagają w trakcie wdrażania interwencji, natomiast ewaluacja *ex post* winna dostarczyć informacji, które pomogą podjąć decyzję o kontynuacji danego programu i pomóc w planowaniu nowego.

Schemat 4. Cykl interwencji UE



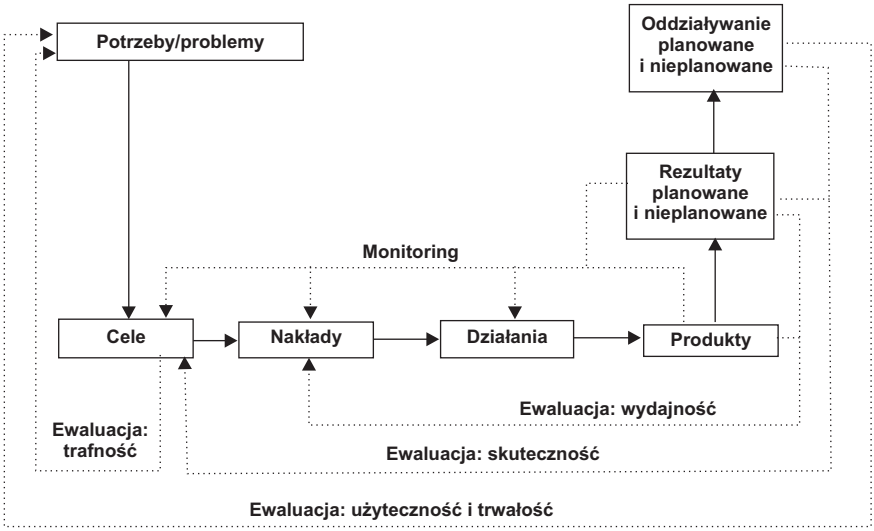
Źródło: opracowanie własne na podstawie European Commission (2004b).

Należy jednak zwrócić uwagę, że jest to podejście czysto teoretyczne. W praktyce ewaluacja – szczególnie *ex post* – znacznie rozmija się z cyklem interwencji, ponieważ programy UE są planowane i wdrażane w sposób ciągły. Na przykład interwencje na lata 2007–2013 były planowane już w 2005 i 2006 roku. Tymczasem ewaluacja *ex post* programów z lat 2000–2006 nastąpi w 2008 roku, a więc w połowie kolejnego okresu programowania. Jej przełożenie na planowanie programów 2007–2013 jest więc znikome. Może służyć pewnymi rekomendacjami w trakcie ich wdrażania, ale i tu pole jest ograniczone, gdyż system wdrażania z reguły jest inny w każdym kolejnym okresie programowym. Ewaluacja średniookresowa natomiast, mimo iż z definicji była podejmowana w połowie wdrażania programów, w rzeczywistości z uwagi na opóźnienia badała programy nie na ich półmetku, a w 1/3 wydatkowania środków (stąd pomysł wprowadzenia uaktualnienia ewaluacji średniookresowej). W praktyce unijnej funkcje formatywne ewaluacji *ex post* są bardzo ograniczone.

Wreszcie trzecim elementem unijnego podejścia do ewaluacji jest **zestaw standardowych kryteriów**, według których jest ona dokonywana. Są to tak zwane kryteria ewaluacji (ang. *evaluation criteria*). Najczęściej stosowanych jest pięć kryteriów ewaluacyjnych: trafność, skuteczność, wydajność, użyteczność i trwałość. Stosowanie kryteriów spełnia trojaką rolę. Po pierwsze pomagają one porządkować pytania ewaluacyjne w bloki (przykłady szczegółowych pytań ewaluacyjnych i ich powiązania kryteriami podano w Tabeli 10). Po drugie kryteria uwypuklają element normatywny ewaluacji (ocena różnych aspektów programu, a nie tylko odpowiedź na pytania opisowe). Po trzecie wreszcie umożliwiają zweryfikowanie przystawania zadanych pytań ewaluacyjnych do cyklu interwencji (czy zadajemy je we właściwym momencie, a co za tym idzie, czy odpowiedź jest możliwa do udzielenia).

Kryteria ewaluacji są ściśle powiązane z przedstawianym schematem logiki interwencji unijnych. W odróżnieniu jednak od monitoringu uwzględniają tak planowane, jak i nieplanowane efekty (patrz poniższy schemat).

Schemat 5. Kryteria ewaluacji a logika interwencji UE



Źródło: opracowanie własne na podstawie European Commission (2000e).

Tabela 10. Główne kryteria ewaluacyjne i pytania badawcze

Kryterium	Opis kryterium	Typ czasowy ewaluacji	Przykłady stawianych pytań
Trafność (ang. *Relevance*)	Bada adekwatność celów i zastosowanych metod do problemów i kwestii społeczno-ekonomicznych, które program miał rozwiązać.	*Ex ante* Bieżąca (dla programu w początkowym stadium wdrażania)	Czy cele programu odpowiadają zidentyfikowanym potrzebom? Czy program odnosi się do głównych sił i słabości danego układu? Czy obrana strategia jest spójna? To znaczy, czy strategia jest spójna wewnętrznie (cele-działania)? Czy strategia jest zbieżna z celami szerszych działań, polityk? Czy logika interwencji jest poprawna (zależności przyczynowo-skutkowe)?
Skuteczność (*Effectiveness*)	Bada stopień realizacji zakładanych celów, skuteczność użytych metod, instytucji oraz wpływ czynników zewnętrznych na ostateczne efekty.	*Ex ante*: trafność, wydajność i skuteczność (*a priori*) Bieżąca *Ex post*	Czy cele programu zostały osiągnięte? Jak silny jest/był wpływ czynników zewnętrznych? Jakie są/były sukcesy programu? Jakie problemy napotyka/napotkał program? Czy wybrane instrumenty i rozwiązania są/były odpowiednie? Czy system zarządzania i wdrażania jest/był skuteczny? Czy instytucje poprawnie wypełniały przypisane im role?
Wydajność (*Efficiency*)	Bada relacje między nakładami, kosztami, zasobami (finansowymi, ludzkimi, administracyjnymi) a osiągniętymi efektami interwencji.	*Ex ante* (szacowanie *a priori*) Bieżąca *Ex post*	Czy podobne efekty można osiągnąć przy wykorzystaniu innych instrumentów? Czy podobne efekty można osiągnąć przy wykorzystaniu niższych nakładów finansowych?
Użyteczność (*Utility*)	Jest powtórzeniem kryterium trafności. Stawia analogiczne pytania, tyle, że w innym momencie czasowym – w trakcie lub po zamknięciu interwencji.	Bieżąca (dla zaawansowanego programu – tylko tych jego działań, które zostały zakończone) *Ex post*	Czy program spełnił oczekiwania adresatów? Czy przyczynił się do rozwiązania problemów? Czy efekty są korzystne dla różnych grup odbiorców? Czy pojawiły się pozytywne efekty uboczne? A konkretnie: jakie środki publiczne zmobilizowane zostały przez środki pomocowe (efekt dodatkowości)? Czy pomoc zmobilizowała jakieś środki prywatne (efekt dźwigni)? Czy zrealizowana inicjatywa pobudziła kolejne inwestycje (efekty impulsu, mnożnikowe)? Czy inwestycja sprzęgła się z innymi działaniami publicznymi – wzmacniające je (efekty synergii)? Czy pojawiły się negatywne efekty uboczne? A konkretnie: czy wykonane inwestycje powstałyby i bez pomocy UE? Jeśli tak, to czy przesunięcie czasowe byłoby znaczące (efekt *deadweight*, tzw. biegu jałowego)? Czy osiągnięte pozytywne efekty nie zostały uzyskane kosztem pogorszenia sytuacji na sąsiednim obszarze (efekt zastępowania)? Czy osiągnięte efekty nie zostały uzyskane kosztem pogorszenia sytuacji osób i instytucji nie objętych programem (efekt substytucji)?
Trwałość (*Sustainability*)	Obejmuje pytania o ciągłość efektów (przede wszystkim pozytywnych) danej interwencji w perspektywie średnio i długookresowej.	Bieżąca (dla zaawansowanego programu – tylko tych jego działań, które zostały zakończone) *Ex post*	Czy efekty działania programu są trwałe, długookresowe? Czy są odczuwalne po zakończeniu interwencji?

Źródło: opracowanie własne na podstawie (European Commission, 1999b, 2004a; Olejniczak, 2005).

3.3.4. Praktyka ewaluacji programów regionalnych – przebieg badań[33]

Niezależnie od typu ewaluacji, obranego podejścia metodologicznego czy tematu, badanie ewaluacyjne jest opisywane według standardowych etapów:

- przygotowania i zlecenia ewaluacji,
- prowadzenia badania ewaluacyjnego,
- przyjęcia i rozpowszechniania wyników badania.

Samo prowadzenie badania podzielić zaś można na cztery fazy: strukturalizacji (sprecyzowania pytań badawczych, podejścia oraz kryteriów, według których program będzie oceniany), obserwacji (zebrania danych wtórnych i pierwotnych), analizy i oceny.

W dalszej części omówiono pokrótce każdy z powyższych bloków działań.

Przygotowanie i zlecanie ewaluacji

W tym etapie wyróżnia się cztery podstawowe fazy: planowanie ewaluacji, jej projektowanie oraz przygotowanie zamówienia na ewaluację i wybór wykonawcy badania.

Zasadniczo wszystkie działania w ramach przygotowania i zlecania ewaluacji dotyczą zlecającego ją podmiotu (Komisji Europejskiej lub jednostki zarządzającej danym programem funduszy strukturalnych). Rola wykonawców badania rozpoczyna się dopiero w momencie ogłoszenia przetargu.

Celem fazy **planowania ewaluacji** jest podjęcie decyzji o badaniu ewaluacyjnym oraz określenie jego ogólnych ram. Decyzja dotycząca podjęcia takiego badania jest najczęściej spowodowana wymogami funduszy strukturalnych (zapisy Rozporządzeń Rady UE), bądź też wynika z potrzeb operatora danego programu. Głównymi decydentami są tutaj instytucja wdrażająca dany program i Komitet Monitorujący. Decyzja dotycząca planu ewaluacji przyjmuje formę krótkiego komentarza w dokumentach programowych, bądź też oddzielnego, krótkiego dokumentu – tak zwanego „mandatu ewaluacyjnego" (ang. *evaluation mandate*). Na tym etapie z reguły określanych jest kilka następujących kwestii:

- Zakres ewaluacji – przedmiotowy (jakiego programu, jakich kwestii horyzontalnych dotyczy ewaluacja), czasowy (jaki przedział czasowy ma objąć ewaluacja), przestrzenny (region, wybrane terytoria objęte programem);
- Cel i forma ewaluacji, czyli ogólne określenie jakim problemem, jaką kwestią ma zająć się badanie – planowaniem, wdrażaniem czy oceną efektów (tu mogą być też postawione wstępne pytania);
- *Momentum* badania ewaluacyjnego (ang. *timing*) i odbiorca, czyli moment w cyklu programu, w którym wnioski z ewaluacji będą najbardziej pożądane i instytucje, które mają używać wyników tego badania. Kwestia ta w sposób oczywisty określa ramy czasowe realizacji badania;

[33] Podrozdział ten w zasadniczej części jest podsumowaniem trzech tekstów źródłowych: European Commission (1999b: vol. 1); Olejniczak (2004); Tavistock Institute i inni (2003). Prezentuje modelowe podejście do procesu ewaluacji. Szerszą dyskusję na temat podejść do ewaluacji interwencji regionalnych Czytelnik może znaleźć w Mazur (2007).

- Kwestie instytucjonalne – wyznaczenie instytucji bezpośrednio odpowiedzialnej za zlecenie i nadzór nad ewaluacją (z reguły dany departament ministerstwa zarządzający programem lub Komitet Monitorujący dany program) oraz decyzja co do tego, kto będzie włączony w proces projektowania kształtu ewaluacji (zakresu badania) i w późniejsze wdrażanie ewaluacji (w tym monitorowanie postępów badania)[34].

Celem fazy **projektowania ewaluacji** jest sprecyzowanie oczekiwań i ich zamiana w wytyczne do realizacji badania ewaluacyjnego. Główna odpowiedzialność na tym etapie spoczywa na instytucji wyznaczonej (na etapie planowania ewaluacji) do nadzoru nad badaniem. W praktyce funduszy strukturalnych jest to zwykle instytucja zarządzająca programem.

Na tym etapie uściślane są decyzje z fazy planowania. W idealnym przypadku zlecający powinien uściślić sześć kwestii (Herman et al., 1987: 14–24).

Po pierwsze, sprecyzować swoje potrzeby informacyjne – czyli dlaczego i do czego potrzebuje ewaluacji. W tym celu wybiera kryteria i pytania, na które odpowiedzi poszukiwać będą ewaluatorzy. Oczywistym jest, że nie na wszystkie wstępnie wybrane pytania i kwestie można odpowiedzieć równie szczegółowo – czas i środki są zawsze ograniczone. Dlatego też wybierane są pytania priorytetowe. W etap wyboru pytań priorytetowych włączani są często członkowie Komitetu Monitorującego, instytucje pośredniczące i wspomagające zarządzanie programem. Decyzja o przesunięciu priorytetów lub ograniczeniu skali badania może się pojawić również w trakcie badania, w ramach konsultacji wykonawca – zlecający.

Po drugie, publika określa odbiorców ewaluacji, a więc grupy osób i instytucji, do których wnioski z badania będą kierowane, którym jego wyniki mają pomóc. Do odbiorców ewaluacji należeć może spektrum aktorów zaangażowanych bezpośrednio lub pośrednio w program tzw. interesariuszy. Chodzi na przykład o sponsorów programu – Komisję Europejską, Ministerstwa, operatorów programu – instytucje bezpośrednio zlecające badanie, instytucje wdrażające program, beneficjentów, lokalną społeczność, a nawet o szeroką opinię publiczną czy polityków. Krąg odbiorców jest w dużej mierze warunkowany rodzajem ewaluacji oraz jej celem.

Po trzecie, określa formę, w jakiej mają być przekazane wyniki badania (forma raportu końcowego czy też prezentacji, serii seminariów, etc.). Zwykle jest to wypadkowa decyzji o strategii użycia ewaluacji, czyli sprecyzowania, w jaki sposób jej wyniki będą rozpowszechniane wśród odbiorców ewaluacji określonych na etapie planowania (np. konferencja podsumowująca, publikacja raportów, publikacja on-line i e-mailowa wysyłka wyników).

Po czwarte, zlecający określa preferowane przez niego podejście badawcze (oczekiwania co do metodologii badania). Zlecający, planując pytania ewaluacyjne, powinien określić swoje oczekiwania co do metod (choćby po to, by uniknąć sytuacji, w której program jest analizowany tylko zza biurka), na przy-

[34] Możliwości są dwie. Cały proces może być opracowany i przeprowadzony przez instytucję do tego zobligowaną (wąskie podejście). Albo też w cały proces zostaną zaangażowane inne podmioty związane z funkcjonowaniem danego programu (podejście szerokie).

kład określić, czy oczekuje ankiet, wizyt terenowych, analiz statystycznych. W tym punkcie możliwe są dwie strategie. Albo metodologia zostaje dokładnie sprecyzowana i zespołowi ewaluacyjnemu pozostaje tylko zrobić badanie zgodnie z danymi wytycznymi, albo też określone są ogólne, ramowe i minimalne oczekiwania wobec metodologii.

Po piąte, zlecający decyduje o dokładnym budżecie i czasie przeznaczonym na badanie.

Wreszcie ostatnim, szóstym pytaniem, jest pytanie weryfikujące całą koncepcję raz jeszcze: Czy odpowiedź na dane pytania jest możliwa, biorąc pod uwagę dany budżet, czas i dostępne dane? Jeśli pojawiają się wątpliwości, warto rozważyć zawężenie pytań, zakresu badania czy uproszczenie oczekiwanej metodologii (lub oczywiście co najmniej prawdopodobne wydłużenie czasu badania i zwiększenie budżetu). Oczywistym jest, że badanie dużego programu pogłębioną metodologią wymagać będzie dużo większego budżetu i dłuższego terminu realizacji badania niż ewaluacja bieżąca słabo zaawansowanego programu. Niezachowanie proporcji miedzy wymienionymi elementami, czyli zaplanowanie nierealistycznego badania, skutkować będzie problemami na etapie jego wdrażania i oczywiście jego znikomą przydatnością[35].

Jeśli ewaluacja jest zlecana na zewnątrz, to kolejną fazą jest **przygotowanie zamówienia na ewaluację (SIWZ)**. Zlecający musi skodyfikować swoje oczekiwania (ustalone na etapie planowania i projektowania) w formie zamówienia na ewaluację. Zgodnie z polską ustawą o zamówieniach publicznych całość takiego zamówienia nosi nazwę Specyfikacji Istotnych Warunków Zamówienia (ang. *Terms of Reference*). SIWZ jest podstawą ogłoszenia konkursu ofert, będzie też podstawą egzekwowania jakości wykonanej przez ewaluatorów pracy. SIWZ przygotowuje instytucja zlecająca ewaluację.

Ostatnia faza przygotowania i zlecania ewaluacji to **wybór wykonawcy badania.** Potencjalni wykonawcy składają swoje oferty na badanie. Oferta jest interpretacją SIWZ. Potencjalny wykonawca przedstawia pomysł na badanie i wstępny zarys jego aspektów technicznych (organizacji badania), odnosząc się do wymogów SIWZ (nie tylko je wypełniając, ale i wychodząc poza nie). Trzy główne elementy gry konkurencyjnej to pomysł badawczy (podejście badawcze, w tym metodologia), powiązana z nim organizacja badania (co i kiedy będzie wykonywane, jak prezentowane) oraz cena. Przedstawiana oferta składa się zwykle z dwóch pakietów – propozycji merytorycznej (technicznej) i finansowej.

W pakiecie merytorycznym (tzw. Organizacja i Metodologia – ang. *Organisation & Methodology*) potencjalny wykonawca ma wykazać się zrozumieniem zakresu zadań (czyli celów ewaluacji, oczekiwań zlecającego) – przedstawia

[35] W praktyce badań amerykańskich instytucje zlecające, zanim podejmą decyzję o dużych badaniach skomplikowanych programów publicznych, zlecają studium pilotażowe – tzw. studium oszacowania możliwości ewaluacyjnych (ang. *evaluability assessment study*). Jak sama nazwa wskazuje, ma ono sprawdzić, czy badanie jest możliwe do zrealizowania, a co za tym idzie, zapobiec ewentualnemu zleceniu kosztownych studiów, które nie mogą (z przyczyn obiektywnych – braku danych, niemierzalności danego zjawiska, zbyt wczesnego etapu programu) dać wiarygodnej odpowiedzi na stawiane pytania.

więc kompletną propozycję realizacji badania. Podsumowanie jej zawartości przedstawia poniższa tabela.

Tabela 11. Oferta – zawartość pakietu merytorycznego

Nazwa	Opis zawartości
Cele i zakres badania	Kontekst programu – akapity mające pokazać, iż ubiegający się o kontrakt rozumie i zna kontekst programu, jaki ma poddać ewaluacji Cele, efekty badania i jego odbiorcy – w praktyce powtórzone z zakresu badań Czynniki sukcesu – jakie elementy zadecydują o dobrym wykonaniu kontraktu (w praktyce eksponowane jest tu doświadczenie wykonawcy) Założenia i potencjalne problemy – ograniczenia, które wykonawca spodziewa się napotkać w czasie badania oraz sposoby ich minimalizowania
Koncepcja badawcza	Strategia badawcza – postawione pytania i podejście zastosowane do analizy i oceny Plan badawczy, czyli opis kolejnych etapów badania, kroków badawczych, które zostaną przeprowadzone Opis zakładanych produktów badania, które będą dostarczone na każdym z etapów badania
Metodologia	Opis skrzynki narzędziowej badania – metod i technik planowanych w realizacji badania, uzasadnienie doboru próby, etc. Krótki opis uczestników projektu (np. poszczególnych instytucji, które wchodzą w skład Konsorcjum stającego do przetargu – ich doświadczenia w ewaluacji, zasobów jakimi dysponują, etc.)
Harmonogram badania	Graficzny plan organizacji i wykonania prac (odpowiadający dokładnie fazom wyróżnionym w koncepcji badawczej) z zaznaczeniem spotkań, raportów okresowych, sprawozdań i co najważniejsze momentów przełomowych (tzw. *milestones*) oraz – jeśli to konieczne – podziału pracy między ekspertami[37]

Źródło: opracowanie własne.

Pakiet drugi to oferta finansowa, czyli plan rozdysponowania oferowanego przez Zlecającego budżetu. Należy pamiętać, że w zakresie badań podany jest maksymalny możliwy budżet – można więc zaoferować cenę niższą (przy niektórych przetargach konkuruje się również cenowo). Budżet jest w oczywisty sposób powiązany z zakresem badania, jego skalą i planem pracy ekspertów.

Wybór wykonawcy badania dokonywany jest według procedury zapisanej w SIWZ (w przypadku Polski zgodnie z polską ustawą o zamówieniach publicznych).

[36] Kluczowa jest tu wizualizacja całego planu działania. Jest on przedstawiany w prostym arkuszu Excela lub też, w bardziej wyszukany w sposób, w programach zarządzania projektami (np. Microsoft Project).

Tabela 12. Typowe kryteria oceny ofert

Kryteria oceny	Elementy brane pod uwagę
Zrozumienie zakresu zadań	• Zrozumienie celów, zadań i oczekiwań zlecającego • Zadeklarowane produkty
Metodologia	• Rzetelność projektu badania • Rzetelność technik zbierania danych • Rzetelność technik analizy danych
Zarządzanie projektem	• Przejrzystość celów i momentów przełomowych (*milestones*) projektu • Przejrzystość i solidność metod pracy • Przejrzystość alokacji zasobów i rozdysponowania pracy
Kwalifikacje zespołu	• Kwalifikacje • Dotychczasowe doświadczenie • Pola specjalizacji – pokrycie dziedzin, których dotyczy ewaluacja (multidyscyplinarność)
Cena	• Wysokość budżetu • Jego racjonalność (w zestawieniu z zaproponowanym zakresem prac)

Źródło: opracowanie własne na podstawie (European Commission, 2004a).

Prowadzenie badania ewaluacyjnego

W zależności od rodzaju ewaluacji na każdym z etapów stosowane są inne metody i narzędzia badawcze. Wachlarz używanych narzędzi badań społeczno-ekonomicznych jest bardzo szeroki. Metody najczęściej stosowane w praktyce UE można podzielić na statyczne – niewymagające „wyjścia zza biurka" oraz na interaktywne – wymagające bezpośredniego kontaktu z przedmiotem badania, interesariuszami programu. Uproszczoną tabelę narzędzi badawczych przedstawia Tabela 12. Należy jednak podkreślić, że dobór narzędzi nie jest kwestią mechaniczną, lecz zależy od: tematyki programu i poziomu analizy, dostępnego budżetu i czasu, danych, a także inwencji i doświadczenia badaczy[37].

[37] Na przykład głównymi narzędziami analizy makro (Narodowe Plany Rozwoju, cała polityka) są modele ekonometryczne (np. HERMIN), przy ewaluacjach średniookresowych i bieżących poszczególnych działań przeważają wywiady i analizy raportów z monitoringu. Ewaluacje końcowe to analiza statystyczna, wywiady, wizyty studyjne, studia przypadków. Ewaluacje programów o tematyce transportowej oparte są głównie na analizach kosztów-korzyści i kosztów-efektywności. Pomoc „miękka" (szkolenia, współpraca) czy programy współpracy ponadgranicznej to wywiady, *benchmarking*, studia przypadków.

Tabela 13. Wybrane metody stosowane w ewaluacji

Metoda	Prospektywne (ewaluacja *ex ante*)				Retrospektywne (ewaluacja *ex post*)			
	strukturalizacja	obserwacja	analiza	ocena	strukturalizacja	obserwacja	analiza	ocena
Metody statyczne								
Użycie danych administracyjnych		+				+	+	
Analiza SWOT	+				+			+
Modele logiczne	+				+			
Mapy oddziaływania	+		+		+		+	
Analiza zmian udziałów			+				+	
Modele ekonometryczne			+				+	
Grupy porównawcze			+				+	
Analiza kosztów-korzyści				+		+		
Analiza kosztów-efektywności								+
Benchmarking (porównania)								+
Metody interaktywne								
Panele dyskusyjne, warsztaty	+	+		+	+	+		+
Badania ankietowe, wywiady		+				+		
Studia przypadków		+	+			+	+	
Mindmaping – mapy logiczne	+		+		+		+	
Obserwacja uczestnicząca		+				+		
Głosowanie kolorami	+			+	+			+
Metoda delficka	+			+	+			+

Źródło: Opracowanie własne na podstawie GHK Consultants (2006); Herman i współpracownicy (1987); Stufflebeam (2000a: 64).

Celami pierwszej fazy badania ewaluacyjnego – **strukturalizacji** – są: sprecyzowanie pytań badawczych, określenie zasad (kryteriów i wskaźników) oceny programu, wybór szczegółowych narzędzi obserwacji i wreszcie zaplanowanie pracy.

Fazę strukturalizacji ewaluatorzy zaczynają od zidentyfikowania i zrozumienia pól działania programu, mechanizmów instytucjonalnych, logiki programu. Właściwe zrozumienie programu – jego ograniczeń, specyfiki i kontekstu – jest kluczowe zarówno dla poprawnej analizy, jak i późniejszego proponowania racjonalnych zmian w programie. Niezbędne na tym etapie są: przegląd dokumentów programowych, wcześniejszych studiów – w tym poprzednich ewaluacji, wreszcie krótkie wywiady z instytucjami zarządzającymi programem.

Dopiero na tej podstawie ewaluatorzy precyzują pytania badawcze, podejście do programu i całego badania. Kwestia: „co badać" została rozstrzygnięta jeszcze przed wyborem wykonawcy i zapisana w zakresie zadań, a następnie w zwycięskiej ofercie. Niemniej do tej kwestii zespół ewaluatorów powraca na początku kontraktu – po to, by niektóre rzeczy uściślić, sprecyzować, czasem zaproponować dodatkowe pytania ewaluacyjne. Tutaj też ogólne kryteria ewaluacji i ogólne pytania postawione przez zlecającego są rozbijane na szereg szczegółowych pytań (opisowych – dostarczających podstaw i kontekstu do odpowiedzi na pytania szczegółowe oraz pytań normatywnych – oceniających program). Do nich dobierane są szczegółowe metody zbierania i analizy danych. Kwestią ostatnią, techniczną, jest szczegółowe zaplanowanie pracy zespołu badawczego.

Podstawowymi narzędziami, które pomagają w fazie strukturalizacji są: analiza SWOT i Pestle, modele logiczne. Bardzo pożytecznym jest też tzw. *mindmapping* – używanie map logicznych w formie drzewek do dyskusji prowadzonych w zespole.

Dobrą praktyką, a często wręcz wymaganiem kontraktu, jest zaprezentowanie wypracowanego pakietu metodologicznego zlecającemu. Zasadniczo im częściej wykonawca konsultuje się na tym etapie z operatorami programu, tym lepsza będzie metodologia. Oczywiście konsultacje nie oznaczają przyjęcia wszelkich postulatów zlecającego. Wykonawca jest niezależny, stanowi zespół badawczy. W przypadku dużych kontraktów etap strukturalizacji trwa około miesiąca.

Celem **fazy obserwacji** jest zebranie przez ewaluatorów danych niezbędnych do analizy. Wyróżniane są dwa rodzaje danych: dane wtórne – już istniejące (z monitoringu, statystyki) i dane pierwotne – zebrane przez samych ewaluatorów. Oczywiście punktem wyjścia jest użycie danych administracyjnych – dokumentacji badanego programu. Przede wszystkim są to dane pochodzące z monitoringu (a więc stopień zrealizowanych celów, monitorowane wskaźniki, etc.), dokumenty programowe, czy wyniki wcześniejszych ewaluacji i studiów. Oparcie się wyłącznie na danych wtórnych jest niewystarczające. Badania prowadzone wyłącznie zza biurka są słabe, a ich wnioski stanowią najczęściej powtórzenie faktów dobrze znanych zarządzającemu programem. Badanie ewaluacyjne (szczególnie średniookresowe i końcowe) powinno też mieć elementy uczestniczące. Paleta dostępnych narzędzi zbierania danych pierwotnych jest

szeroka. Najczęściej są to: wywiady, badania ankietowe, badania społeczne, wizyty w terenie – lustrowanie projektów czy studia przypadków.

Podczas **fazy analizy** ewaluatorzy dokonują analizy i interpretacji zebranych danych, oszacowania efektów programu, zweryfikowania postawionych hipotez oraz ostatecznej oceny programu i jego efektów. Używane są przede wszystkim metody statystyczne oraz porównawcze. Szczególny nacisk jest kładziony na oszacowanie efektów interwencji – zarówno tych zamierzonych, jak i niezamierzonych.

Wreszcie **faza oceny** jest kluczowym elementem ewaluacji i dokonywana jest w odniesieniu do uprzednio sformułowanych pytań ewaluacyjnych i wybranych kryteriów. Stwierdza się w niej faktyczną jakość danej interwencji. Jednocześnie formułowane są wstępne wnioski i rekomendacje z analizy mające służyć poprawie programu.

W przypadku oceny i formułowania rekomendacji bardzo dobrą praktyką jest włączanie w nią interesariuszy programu. Ostateczny głos należy do ewalutorów. Takie podejście daje następujące korzyści:

- wypracowana ocena, nawet bardzo krytyczna, jest łatwiejsza do zaakceptowania przez uczestników programu – stanowi krytykę konstruktywną,
- wnioski płynące z oceny są bardziej realistyczne i łatwiej przyjmowane przez uczestników programu (identyfikują się z nimi, bo brali udział w ich wypracowywaniu).

Przygotowanie, przyjęcie i rozpowszechnianie raportu końcowego

Bardzo ważnym elementem pracy nad raportem końcowym jest prezentacja i konsultacje na temat wstępnego raportu końcowego. Powinny one objąć zarówno kwestie merytoryczne (zawartość rekomendacji), jak i czysto techniczne (forma prezentacji – struktura raportu). Taki dialog prowadzony między wykonawcą a zlecającym (przy zachowaniu bezstronności i niezależności wykonawcy – badacza) znacząco przyczynia się do późniejszej skuteczności raportu – wdrażania rekomendacji, zainteresowania wynikami badań i szerokiego odzewu. Momentem kończącym formalnie badanie ewaluacyjne jest przekazanie zlecającemu przez wykonawcę ostatecznej wersji raportu.

Tabela 14. Dobre praktyki raportów ewaluacyjnych

• krótkie podsumowanie 3–5 stron (*executive summary*),
• wnioski i rekomendacje powtórzone w formie listy na wstępie dokumentu i pogrupowane tematycznie (ocena programu, system wskaźników, zarządzanie, etc.),
• podział na rekomendacje operacyjne i systemowe,
• przy obszernych raportach streszczenia rozdziałów na początku każdego z nich,
• przy dużej, zróżnicowanej grupie odbiorców mapa dla czytających (które z rozdziałów są najbardziej przydatne dla poszczególnych aktorów programu),
• krótkie podsumowania każdego z akapitów na marginesach raportu,
• sprecyzowanie, kto jest adresatem rekomendacji, jak pilne jest jej wdrożenie,
• załączane jako aneksy wzory ankiet, listy wywiadów.

Źródło: opracowanie własne.

Przygotowany raport powinien spełniać pewne standardy europejskie. Raport musi być napisany na temat, powinien odpowiadać standardom naukowym, zawierać opis przesłanki ewaluacji, cel, pytania, procedurę badawczą, przytaczać dokładne źródła informacji, musi także zawierać wyjaśnienie procesu dedukcji i interpretacji danych. Rekomendacje powinny być precyzyjne, wnioski – jasne i mające pokrycie w treści raportu.

Zdarzają się sytuacje, szczególnie w przypadku braku wcześniejszej dyskusji nad rekomendacjami, iż zlecający, bądź ktoś z kluczowych interesariuszy podważa krytyczne wnioski raportu. Takie sytuacje są szczególnie powszechne, gdy dana instytucja traktuje ewaluację jak audyt – uznaje badanie za rodzaj karnego dochodzenia i oceny jej kompetencji. Wcześniejsze konsultowanie metodologii i wstępnych wniosków może zapobiec sytuacji konfliktowej. Jeśli już jednak do niej doszło, to sposobem rozwiązania jest spisanie tzw. protokołu rozbieżności. Instytucja podważająca badanie wypisuje swoje uwagi, do których następnie odnosi się zespół ewaluatorów. Taki wewnętrzny protokół rozbieżności jest następnie podstawą do dalszej dyskusji. Ewaluatorzy do każdego krytycznego wniosku przypisują rekomendacje mające pomóc w zlikwidowaniu wad wykrytych w programie. Wspólnie ze zlecającym przygotowują realistyczny kalendarz wprowadzania tychże rekomendacji (uwzględniający ograniczenia instytucjonalne). Kalendarz jest publikowany razem z raportem. Krytyczny wydźwięk raportu może być zatem zniwelowany przez jednoczesny plan naprawczy. Stawia to instytucję odpowiedzialną za program w pozytywnym świetle, jako organizację uczącą się.

Każdy z typów ewaluacji ma już w założeniu przypisany sposób jego rozpowszechniania i wykorzystania. I tak, ewaluacja wstępna jest wykorzystywana do korekty planu interwencji. Ewaluacje średniookresowe i bieżące są używane do zmian w zarządzaniu programem, do poprawy programu w trakcie jego trwania i ewentualnie do projektowania nowego, jeśli ewaluacje są publikowane pod koniec jego trwania. Wreszcie użycie ewaluacji *ex post* ogranicza się do informowania społeczeństwa i pokazania *post factum* roli funduszy. Ewaluacje *ex post*, wbrew teorii, nie mają żadnego wpływu na kontynuowanie czy zaprzestanie danej interwencji z tej prostej przyczyny, że są kończone w połowie trwania kolejnej edycji funduszy (np. raporty z ewaluacji *ex post* dla lat 1994–1999 były dostępne w 2004 roku). Mogą jednak pomagać w poprawie innych analogicznych programów.

Ważnymi elementami przy rozpowszechnianiu wyników ewaluacji są:

- jasna strategia rozpowszechniania wyników ewaluacji – określenie grup docelowych i sposobów komunikowania wniosków (powinno być to zrobione już na etapie projektowania ewaluacji, zaś raport z badania powinien być dostosowany właśnie do tych wymagań);
- szeroki dostęp do wyników badania – publikowanie wyników na stronach internetowych instytucji zarządzającej programem, informacja dla beneficjentów, zorganizowanie dyskusji wokół wyników badania;
- zaprojektowanie systemu informacji dla kontynuacji badania (*follow up* wdrażania rekomendacji i śledzenie ich efektów).

W przewodnikach Komisji sugerowane jest dobranie odpowiednich kanałów komunikacji do odbiorców. Podsumowuje to poniższa tabela.

Tabela 15. Kanały komunikacji i odbiorcy ewaluacji

		Kanały komunikacji							
		spotkanie połączone z dyskusją	skrócony raport	nota	publikacja	synteza	prezentacja	broszury, artykuły	konferencja prasowa
ODBIORCY	Komisarz			+	+	+	+		
	Grupa sterująca programem	+	+		+		+		
	Operatorzy programu	+	+		+		+		
	Zarządzający programem	+	+		+		+		
	Instytucje europejskie, regionalne, krajowe				+	+		+	
	Obywatele, prasa				+	+		+	+

Źródło: (European Commission, 1999b: vol. 1, s. 162–163).

Podsumowując proces ewaluacji i odnosząc powyższy opis do praktyki – szczególnie polskich ewaluacji – należy zaznaczyć cztery kwestie.

Po pierwsze, planowanie ewaluacji w Polsce dla funduszy strukturalnych coraz częściej przybiera formę rocznych, a nawet kilkuletnich planów ewaluacji. Z jednej strony porządkuje to rynek i pracę samych zlecających, zwiększa przewidywalność. Z drugiej oznacza to, że instytucje antycypują swoje potrzeby informacyjne z dużym wyprzedzeniem i ograniczają własną elastyczność na tym polu. To znacząco ogranicza użyteczność ewaluacji bieżących. Gdy pojawi się nowy problem – kwestia badawcza wynikająca z bieżącej, zmieniającej się szybko sytuacji, dana instytucja będzie musiała czekać co najmniej rok na wykonanie badania. Takie nieelastyczne podejście wynikać może z rocznego cyklu budżetowania polskiej administracji i starych biurokratycznych nawyków. Kłóci się to w oczywisty sposób z wizją nowoczesnej administracji – elastycznej, szybko reagującej na zmiany, sięgającej po ewaluację w celu rozwiązania zarówno strategicznych, jak i bieżących kwestii. Drugim hipotetycznym wytłumaczeniem takiego podejścia może być fakt, iż zlecanie badań ewaluacyjnych nie jest powiązane ze strategicznym zarządzaniem, a staje się jedną z kolejnych popularnych, ale martwych procedur. Cykl zlecania ewaluacji jest więc oderwany od konkretnych potrzeb informacyjnych – nikt nie oczekuje, że ma te potrzeby zaspokajać i przynosić bazę do podejmowanych decyzji. Te toczą się swoim tempem i według własnej, politycznej logiki.

Po drugie, w praktyce strukturalizacja następuje dwukrotnie. Ta kluczowa dla formy i użyteczności badania dokonywana jest przez zlecającego na etapie

projektowania ewaluacji. Opisywanych sześć głównych decyzji determinuje ofertę wykonawcy i późniejszy przebieg badania. W polskiej praktyce błąd popełniany jest zwykle przy uczciwej odpowiedzi na zasadnicze pytanie: czy odpowiedź na zadane pytania jest możliwa i wykonalna przy określonym zakresie, dostępnych danych i przeznaczonym budżecie? Nierealne oczekiwania skutkują słabymi badaniami, a te z kolei nie mają żadnej użyteczności w podejmowaniu dalszych decyzji.

Po trzecie, należy zaznaczyć, że „idealne" cztery fazy badania ewaluacyjnego w praktyce często zlewają się ze sobą – zbieranie danych, ich interpretacja i analiza następują równocześnie, a elementy oceny (szczególnie z włączeniem w nią interesariuszy) odbywają się już w fazie zbierania danych czy przy ich wstępnej analizie.

Wreszcie ostatnią godną podkreślenia kwestią jest zaplanowanie wykorzystania wyników już na etapie projektowania ewaluacji. Nie chodzi tu tylko o wskazanie technicznych kanałów rozpowszechniania wyników, ale o strategiczną refleksję – w czym dane badanie, w jakiej decyzji, w jakim momencie ma pomóc. Jak pokazuje polska praktyka, wiąże się to z umiejętnością precyzowania przez zlecającego swoich potrzeb informacyjnych. Brak sprecyzowanych potrzeb daje potem wykonawcom możliwość różnych interpretacji (zwykle minimalistycznych) i dostarczenia badań o niskiej przydatności.

4. STUDIUM WYKORZYSTANIA EWALUACJI W PROGRAMACH INTERREG III

4.1. WSTĘP

Rozdział ten przedstawia wyniki badań empirycznych wykorzystania ewaluacji średniookresowych programów współpracy terytorialnej INTERREG III. Jego pierwsza część przybliża Inicjatywę Wspólnotową INTERREG, jej miejsce w polityce spójności UE, sposób organizacji i specyfikę ewaluacji programów terytorialnych. W części drugiej zaprezentowano rdzeń analizy – metamodel wykorzystania ewaluacji. Wreszcie w części ostatniej tego rozdziału zawarto szczegółową analizę (z zastosowaniem modelu) trzech studiów przypadków ewaluacji średniookresowych wybranych programów INTERREG.

4.2. CHARAKTERYSTYKA INICJATYWY WSPÓLNOTOWEJ INTERREG

4.2.1. Współpraca terytorialna jako element Polityki Spójności UE

Współpraca terytorialna jest częścią Polityki Spójności Unii Europejskiej. Polityka spójności została powołana w 1987 roku zapisami Jednolitego aktu europejskiego modyfikującego traktat rzymski (tytuł XVII, artykuły 158–162 traktatu rzymskiego). Jako podstawowy cel postawiono „wzmacnianie spójności ekonomicznej i społecznej Wspólnoty" (art. 158 TR). Cel ten miał być realizowany poprzez „zmniejszanie dysproporcji w rozwoju poszczególnych regionów i zacofania regionów mających najmniej korzystne warunki" (drugi akapit art. 158 TR). Nadaje on polityce spójności silny wymiar regionalny i interwencjonistyczny (Nowak, 2005: 69).

Impulsem do działania była kombinacja procesów zewnętrznych i planów politycznych. Drugi światowy kryzys gospodarczy i rozszerzenie UE (ówczesnych Wspólnot Europejskich)[1] na słabiej rozwinięte państwa (Grecja, Hiszpania, Portugalia) uwydatniły nierówności międzyregionalne. Posiłkując się teorią centrum-peryferii, Komisja oraz część z państw członkowskich skonstatowały, że Unia stanęła przed groźbą rozłamu na dynamiczne regiony centrum i słabe regiony peryferyjne (Szlachta, 1997: 27). Silna koalicja słabszych gospodarczo państw (Hiszpania, Irlandia, Portugalia, Grecja, regiony południa Włoch) wyrażała obawę, że korzyści płynące z planowanego przejścia

[1] W całej pracy zastosowano uproszczone nazewnictwo. Terminu Unia Europejska użyto zarówno do sytuacji sprzed 1993 roku – a więc do istniejących wtedy Wspólnot Europejskich, jak i do stanu prawnego po wejściu w życie Traktatu z Maastricht – formalnego początku UE.

do kolejnego, wyższego etapu integracji (Jednolitego Rynku Europejskiego)[2] pojawią się głównie w regionach wysoko rozwiniętych, kosztem regionów słabych. Państwa te groziły więc zablokowaniem projektu JRE. UE zdecydowała się więc stworzyć mechanizm zrównoważenia zysków ze wspólnego rynku i budowania spójności gospodarczej[3]. Wspólną Politykę Spójności oparto na pięciu argumentach (szerzej patrz: Amstrong, 1993: 154–155; Balme, 1997: 64–67; Martin, 1998: 76): koncentracji finansowej[4], skali – koordynacji[5], wspólnego zysku[6], solidarności[7] oraz wpływu na inne polityki[8].

Decyzji politycznej i zapisom traktatowym towarzyszyły znaczące środki finansowe, sukcesywnie zwiększane w kolejnych okresach budżetowych[9]. Polityka Spójności stała się więc jedną z najważniejszych dziedzin działania Unii Europejskiej (40% budżetu UE).

Główne cele i instrumenty realizacji polityki spójności modyfikowano w kolejnych okresach, jednak aż do 2007 roku ich logika była analogiczna: głównym celem była pomoc regionom opóźnionym (programy tzw. celu 1 przeznaczone dla regionów UE mających PKB poniżej 75% średniej UE), a kolejnym – pomoc regionom poprzemysłowym (programy celu 2) oraz słabo zaludnionym, rolniczym (programy celu 4, 5 i 6; Allen, 2000). Cele te były realizowane przy uży-

---

[2] Kolejne etapy gospodarczej integracji regionalnej to: unia celna, wspólny rynek i wreszcie etap najwyższy – unia gospodarczo-walutowa (szerzej patrz: Nowak, 2002).

[3] Wielu analityków podkreśla, że polityka spójności była częścią pakietu JRE – interwencjonistyczną równowagą dla wolnego rynku oraz ceną (tzw. *side-payment*), za którą bogate państwa „kupiły" udział biedniejszych krajów w tym projekcie (Behrens, Smyrl, 1999; Dinan, 1999: 430–432).

[4] Ang. *The financial targeting argument* zakładał, że efektywność zostanie podniesiona poprzez koncentrację środków na obszarach najbardziej uwstecznionych. Biedniejsze państwa są w stanie skierować środki pomocowe tylko do nielicznych, wybranych rejonów, podczas gdy UE może działać na szerszą skalę, obejmując pomocą większe obszary, często wręcz całe terytorium danego państwa.

[5] Uważano, że zwiększy się skuteczność koordynacji pomocy zarówno polityk krajowych, jak i wspólnotowych inicjatyw regionalnych. Brak koordynacji rozwiązywania problemów regionalnych przez państwa UE 12 zakłóciłby inne polityki (np. politykę konkurencji). Koordynacja i współpraca dadzą też większą efektywność przy wspieraniu problemowych regionów przygranicznych. Działania podlegać będą mniejszym ograniczeniom, wynikającym z granic państwowych, różnic w przepisach narodowych, etc.

[6] Argument wspólnego zysku (ang. *vested interest argument*) oznaczał, że wprowadzenie polityki spójności leży we wspólnym interesie wszystkich państw członkowskich. Ubóstwo słabszych regionów ogranicza wielkość i chłonność wspólnego rynku, co także rzutuje na dynamikę gospodarczą całej organizacji. Na regionach uwstecznionych nie zarabia nikt, za to działają one destabilizująco na sąsiadujące z nimi obszary.

[7] Zasada solidarności, nazywana również argumentem efektów integracji, dowodziła, że korzyści z jednolitego rynku rozłożą się nierównomiernie pomiędzy państwami. Należy więc zastosować jakiś mechanizm redystrybucji, tak by wszystkie państwa miały udział w korzyściach płynących z efektów integracji.

[8] Narastające i nierozwiązane różnice regionalne stwarzają poważne zagrożenie dla funkcjonowania wspólnego rynku, destabilizują i niweczą także inne polityki i działania Wspólnoty. Uzasadnienie to wiąże się z racją wspólnego zysku.

[9] Wydatki w każdym z okresów programowania planowano rosnąco. Uśrednione roczne wydatki w latach 1988–1993 wynosiły 9,6 mld ECU, dla lat 1994–1999 wzrosły do 25 mld euro, w latach 2000–2006 ograniczano je do 22 mld euro, by zwiększyć je radykalnie na lata 2007–1913 do 49 mld euro rocznie.

ciu instrumentów finansowych – Funduszy Strukturalnych. W 1993 roku do instrumentarium polityki spójności dodano Fundusz Spójności. Prócz programów głównego nurtu (tzw. *mainstream funds* – cele 1–5) funkcjonowały również Inicjatywy Wspólnotowe (ang. *Community Initiatives*), zatwierdzone w 1991 roku – pomyślane jako wparcie dla inicjatyw i programów leżących we wspólnym interesie Unii. W odróżnieniu od programów głównego nurtu (osadzonych w narodowych dokumentach strategicznych, przygotowywanych przez każdy kraj oddzielnie), Inicjatywy Wspólnotowe były programami ponadnarodowymi, których cele strategiczne wytyczano – przy kluczowym udziale Komisji Europejskiej – dla wszystkich regionów UE (Moussis, 2003: 218).

Polityka spójności UE była, i wciąż pozostaje, oparta na kilku kluczowych zasadach (Szlachta, 2000: 69): zasadzie koncentracji (kilka celów strategicznych i obszarów problemowych), programowaniu (wieloletnie planowanie w formie kompleksowych programów rozwojowych), zasadzie partnerstwa (między podmiotami publicznymi różnego szczebla zaangażowanymi w dane interwencje, jak również między aktorami sektora prywatnego, publicznego i pozarządowego) oraz zasadzie dodatkowości (środki UE nie zastępujące, a jedynie zwiększające poziom nakładów krajowych). Jej okresy były skoordynowane z pakietami budżetowymi UE. Tabela 14 podsumowuje ewolucję celów polityki spójności UE, jej instrumentów i środków w kolejnych czterech okresach programowania.

Tabela 16. Reformy Polityki Spójności Unii Europejskiej

Cele	Fundusze[10]	% udział
Okres Programowania 1988–1993	3 Fundusze Strukturalne	62 mld ECU
CEL 1 (regionalny) Promowanie rozwoju i dostosowań strukturalnych regionów zapóźnionych (PKB *per capita* poniżej 75% średniej UE)	EFRR, EFS, EFOGR-Sekcja Orientacji	57,2%
CEL 2 (regionalny) Konwersja regionów, obszarów przygranicznych i subregionów oraz miast dotkniętych upadkiem przemysłu	EFRR, EFS	10,4%
CEL 3 (horyzontalny) Zwalczanie długoterminowego bezrobocia	EFS	15,6%
CEL 4 (horyzontalny) Ułatwianie integracji zawodowej ludzi młodych	EFS	
CEL 5a (horyzontalny) Przyspieszanie adaptacji struktur rolniczych	EFOGR-SO	5,2%
CEL 5b (regionalny) Ułatwianie rozwoju i dostosowań strukturalnych regionów rolniczych (w tym struktur rybołówstwa)	EFOGR-SO, EFRR, EFS	2,6%
10 Inicjatyw Wspólnotowych: Stride, Envireg, Regis, Rechar, INTERREG, Regen, Prisma, Leader, Employment Initiatives (Euroform, Now, Horison), Telematique	Wszystkie FS – zależne od tematu	9,0% (3,8 mld)

[10] Fundusze Strukturalne (FS) to: Europejski Fundusz Rozwoju Regionalnego (EFRR), Europejski Fundusz Społeczny (EFS), Europejski Fundusz Orientacji i Gwarancji Rolnej – sekcja orientacji (EFOGR-SO), Finansowy Instrument Wspierania Rybołówstwa (FIWR). Funduszem nie należącym do Funduszy Strukturalnych jest natomiast Fundusz Spójności (FSp).

Okres Programowania 1994–1999	4 Fund. Strukturalne + Fundusz Spójności	152 mld euro
CEL 1 (regionalny) – Bez zmian	EFRR, FIWR, EFS, EFOGR-SO	61,8%
CEL 2 (regionalny) – Bez zmian	Bez zmian	10,1%
CEL 3 (horyzontalny) Zwalczanie długoterminowego bezrobocia, ułatwianie integracji zawodowej ludzi młodych zagrożonych wypadnięciem z rynku oraz ułatwianie powrotu na rynek osobom zagrożonym stałym bezrobociem	Bez zmian	8,5%
CEL 4 (horyzontalny) Ułatwianie adaptacji zatrudnionych obojga płci do zmian zachodzących w przemyśle i do przemian w systemach produkcji	Bez zmian	1,5%
CEL 5a (horyzontalny) – bez zmian	EFOGR-SO, FIWR	4,0%
CEL 5b (regionalny) – bez zmian	Bez zmian	4,5%
CEL 6 (regionalny) – od 1995 roku Rozwój regionów o niskiej gęstości zaludnienia (poniżej 8 osoby na km^2)	EFRR, FIWR, EFS, EFOGR-SO	0,5%
13 Inicjatyw Wspólnotowych: INTERREG II, Leader II, Regis II, Employment Intitaitves (Now & Youth, Horizon), Adapt, Rechar II, Resider II, Retex, Konver, Kobi, Urban, Pesca, Peace	Wszystkie FS – zależne od tematu	9,2% (13,9 mld)

Cele	Obsługujące fundusze[10]	% udział
Okres Programowania 2000–2006	4 Fund. Strukturalne + Fundusz Spójności	129 mld euro
CEL 1 (regionalny) Promowanie rozwoju i dostosowań strukturalnych regionów zapóźnionych, w tym regionów o bardzo niskiej gęstości zaludnienia	Bez zmian	70%
CEL 2 (regionalny) Rekonwersja gospodarcza i społeczna regionów dotkniętych upadkiem przemysłu lub zmianami w sektorze usług, zacofanych obszarów wiejskich, regionów kryzysowych uzależnionych od sektora rybołówstwa, problemowych obszarów miejskich	Bez zmian	13%
CEL 3 (horyzontalny) Pomoc w adaptacji i modernizacji systemów oświaty, szkolenia zawodowego i zatrudnienia	Bez zmian	12%
4 Inicjatywy Wspólnotowe: INTERREG III, Urban II, Equal, Leader +	EFRR, EFS, EFOGR-SO	5% (5,3 mld)
Okres Programowania 2007–2013	3 Fundusze	347 mld euro
CEL 1 (regionalny) Konwergencja – wsparcie rozwoju regionów zapóźnionych, których poziom PKB *per capita* jest niższy od 75% średniej UE oraz regionów, gdzie ten poziom nieznacznie przewyższa powyższy próg (wygasająca pomoc przejściowa w związku z efektem statystycznym rozszerzenia).	EFRR, EFS, FK	81,54%

CEL 2 (regionalny) Regionalna konkurencyjność i zatrudnienie – umocnienie konkurencyjności i atrakcyjności regionów (spoza celu konwergencji), jak również zwiększenie zatrudnienia	EFRR, EFS	15,95%
CEL 3 (regionalny) Europejska współpraca terytorialna	EFRR	2,52% (8,7 mld)

Źródło: opracowanie własne na podstawie dokumentów UE.

Z zestawienia płynie kilka wniosków:

- głównym celem polityki spójności jest niewątpliwie pomoc regionom zapóźnionym, definiowanym jako obszary o poziomie PKB *per capita* poniżej 75% średniej UE. Rdzeniem działań polityki spójności jest, jak sama nazwa wskazuje, konwergencja – oparta na logice pomocy słabszym układom;
- w kolejnych latach następowała koncentracja środków na wybranych celach strategicznych. Należy jednak zwrócić uwagę, że często była to koncentracja pozorna – redukując liczbę celów, rozszerzano ich zapisy, tak by poszczególne obszary pomocowe wciąż otrzymywały wsparcie (np. likwidacja celu 6 i włączenie go w cel 1);
- w kolejnych okresach programowania sukcesywnie zwiększano środki (z wyjątkiem okresu budżetowego 2000–2006), zwiększając poziom interwencji pięciokrotnie;
- zmiany w instrumentach polityki spójności powiązane były z przechodzeniem do kolejnych etapów integracji UE (pogłębianie), ale przede wszystkim rozszerzaniem jej terytorialnego zasięgu o nowe państwa (poszerzanie). Jak zauważa Szlachta (1997: 32), wymiar polityczny jest więc główną przesłanką do interwencjonizmu w polu polityki regionalnej;
- w kolejnych latach upraszczano organizację instrumentów finansowych obsługujących poszczególne cele, redukując ich liczbę, upraszczając relację cel-fundusz i ostatecznie doprowadzając do zintegrowania Funduszy Strukturalnych i Funduszu Spójności;
- Inicjatywy Wspólnotowe stanowiły nieznaczny procent w budżecie całej polityki spójności. W kolejnych okresach redukowano ich liczbę (zwiększając jednocześnie środki) i tym samym koncentrując te innowacyjne, ogólnounijne inicjatywy na kilku kluczowych polach;
- głównym polem w ramach Inicjatyw Wspólnotowych (o największym budżecie), które znalazło swoją kontynuację również w nowym okresie programowania, okazała się współpraca terytorialna – Inicjatywa Wspólnotowa INTERREG.

Należy jednak podkreślić, że filozofia działania polityki spójności zmienia się Gorzelak (2007). „Konwergencja" jest coraz częściej rozbijana na dwa zjawiska: konwergencję w skali krajów i dywergencję zachodzącą w skali regionów. Gorzelak (2007: 16) stawia tezę, iż ten drugi element (zróżnicowanie) jest jednym z czynników dynamizujących tempo wzrostu (czyli skracanie dystansu nowych krajów UE do starych państw członkowskich). Dowodzi również, odwołując się do szeregu najnowszych badań, że ciągłe publiczne wsparcie dla stałych układów przestrzennych ma skutki raczej socjalne niż rozwojowe, ugruntowuje ich słabość, demobilizuje, nie przyczynia się do trwałego rozwo-

ju, ale wręcz przeciwnie – może go nawet spowalniać. Takie ujęcie prowadzi do akceptacji zróżnicowań (jako nieodłącznego a nawet pożądanego elementu rozwoju) i nakierowania interwencji publicznych na dalsze wzmacnianie już dynamicznych układów – zdolnych do sprostania globalnej konkurencji. Tę filozofię widać w strukturze celu polityki spójności na lata 2007–2013 (Cel 2 – Wsparcie konkurencyjności).

4.2.2. INTERREG jako narzędzie współpracy terytorialnej

Jak już zaznaczono w poprzednim punkcie, współpraca terytorialna była tylko jednym z wielu, i to nieznacznym (w zestawieniu z narodowo zorientowanymi programami głównego nurtu funduszy) elementem polityki spójności UE. Była jednak największą z Inicjatyw Wspólnotowych, a po likwidacji tychże, dla lat 2007–2013, stała się jednym z trzech głównych celów polityki spójności.

Inicjatywa Wspólnotowa INTERREG zasługuje na uwagę z racji na jej wyjątkowy charakter. Po pierwsze, celem INTERREG jest współpraca terytorialna, a więc ponad granicami, ponadeuropejska. Ponadto INTERREG, będąc inicjowany przez Komisję Europejską i nakierowany na ogólnoeuropejskie wyzwania, jest inicjatywą *sensu stricte* europejską. Po drugie (co zostanie szerzej omówione w dalszej części), w odróżnieniu od głównego nurtu funduszy strukturalnych programy INTERREG były wdrażanie nie przez pojedyncze administracje krajowe, a przez instytucje z różnych krajów, w ramach współpracy. To tworzyło więc prawdziwie ponadnarodowe, międzykulturowe struktury i procesy zarządzania. Po trzecie, INTERREG z racji ograniczonych środków w mniejszym stopniu koncentrował się na wyrównywaniu poziomów i infrastrukturze, a w większym na tworzeniu połączeń funkcjonalnych między krajami i instytucjami. Interakcje i współpraca między aktorami z różnych krajów była więc rdzeniem tych programów. Po czwarte, mniejsze pieniądze paradoksalnie dawały większą elastyczność i możliwości innowacji tak Komisji, jak i administracjom regionalnym, gdyż centralne administracje krajowe koncentrowały się na środkach z głównych celów polityki spójności. Po piąte wreszcie, inicjatywa INTERREG okazała się wyjątkowo nośna i była kontynuowana przez wszystkie kolejne lata, zaś w nowym okresie programowania jej ranga wzrosła. Powyższe argumenty sprawiają, że INTERREG jest wyjątkowo ciekawym przedmiotem studiów związanych z procesami zarządzania i uczenia się, gdyż osadza je w środowisku ponadnarodowym i międzykulturowym.

Współpraca terytorialna ma w Unii Europejskiej kilka aspektów, wynikających z jej zróżnicowanych celów. Jako główne cele programów współpracy terytorialnej określono (Council of the European Union, 1999: art. 10.1a; 2006: art. 7):

- wspieranie współpracy transgranicznej (ang. *cross-border cooperation*) pomiędzy sąsiadującymi ze sobą regionami UE położonymi przy granicach wewnętrznych i zewnętrznych (lądowych i morskich);

- wzmacnianie współpracy transnarodowej (ang. *transnational cooperation*) w ramach ponadnarodowych stref współpracy obejmujących duże, europejskie obszary funkcjonalne (np. region Morza Bałtyckiego, region Alp);
- wspieranie współpracy międzyregionalnej (ang. *interregional cooperation*) – czyli wymianę doświadczeń w ramach sieci współpracy zawiązywanych pomiędzy dowolnymi regionami Unii Europejskiej.

W założeniu współpraca terytorialna ma:

- pomóc regionom peryferyjnym w przełamywaniu ich izolacji,
- przezwyciężać negatywne efekty granicy jako bariery rozwoju społeczno-gospodarczego,
- promować sieciowe, wspólne, skoordynowane rozwiązania problemów i wyzwań stojących przed regionami europejskimi.

Tym samym wzmacnia harmonijny, zrównoważony i trwały rozwój całego obszaru Wspólnoty, a co za tym idzie, przyczynia się do realizacji ogólnego celu polityki spójności UE – spójności gospodarczej i społecznej.

Z perspektywy regionalnej, istnieją dwa główne czynniki mobilizujące do nawiązywania współpracy terytorialnej: przełamywanie wspólnych problemów, które dotykają dane regiony (np. peryferyjność, kwestie ekologiczne) i powiększanie wspólnych korzyści (np. podnoszenie przewagi konkurencyjnej regionu przygranicznego jako funkcjonalnej całości) (Olejniczak, 2006).

Inicjatywę Wspólnotową INTERREG (ang. *Community Initative INTERREG*) stworzono w 1990 roku, a więc w dwa lata po uruchomieniu polityki spójności. INTERREG, będąc częścią polityki spójności, był również zorganizowany według zasady wieloletnich pakietów planowania i budżetowania. Dotychczas zrealizowano więc III edycje tego programu: INTERREG I (1990–1993), INTERREG II (1994–1999) i INTERREG III (2000–2006). Jak już wspomniano, w nowej perspektywie budżetowej współpraca terytorialna stała się jednym z trzech głównych celów polityki regionalnej UE[11]. Formalnie Inicjatywy Wspólnotowe przestały więc istnieć od 2007 roku. W praktyce aktorzy europejscy zaangażowani w dotychczasowe programy INTERREG posługują się terminem „INTERREG IV" przy opisywaniu i zarządzaniu nowymi programami współpracy terytorialnej.

W kolejnych edycjach rozłożenie akcentów i orientacja poszczególnych części Interregu nieco się zmieniała, a każda z tych inicjatyw finansowała szereg programów bilateralnych lub multilateralnych. Tabela 15 podsumowuje tę ewolucję dla lat 1990–2006.

[11] To niewątpliwie podkreśla wzrost znaczenia współpracy terytorialnej jako elementu rozwoju regionalnego UE. Z drugiej strony należy zaznaczyć, że środki przeznaczone na Cel 3 są stosunkowo nieduże w porównaniu z finansowaniem Celu 1 i 2. Na Cel 3 – Europejską Współpracę Terytorialną – alokowano 2,5% z 308 mld euro (w cenach z 2004 roku) przeznaczonych na politykę spójności UE.

Tabela 17. Ewolucja Inicjatywy INTERREG

Inicjatywa Wspólnotowa	INTERREG I 1990–1993	INTERREG II 1994–1999	INTERREG III 2000–2006
Budżet ogółem	1,082 mld euro	3,600 mld euro	4,875 mld euro
Liczba programów	31 programów	79 programów	72 (83) programy[12]
Współpraca transgraniczna (ang. *Cross-border cooperation*) Współpraca między sąsiadującymi regionami przygranicznymi	INTERREG I 31 programów (w tym 4 morskie)	INTERREG II A 59 programów (w tym 16 morskich) z czego: – 31 na wewnętrznych granicach UE – 28 na zewnętrznych granicach UE	INTERREG III A 53 programy (w tym 15 morskich) z czego: – 24 na granicach wew. UE15 – 29 granicach zew. UE15 (z czego 12 na granicach z *NMS)* – 11 nowych programów (po 2004)
Sieci energetyczne (ang. *Energy networks*)	–	INTERREG II B 3 prog. – kontynuacja REGEN CI	–
Współpraca transnarodowa (ang. *Transnational cooperation*) Współpraca w ramach przestrzennych stref geograficznych	–	INTERREG IIC i Działania Pilotażowe art. 10 13 programów nt. Planowania przestrzennego (w kontekście ESPD)[13] 7 programów współpracy ponadnarodowej 2 programy przeciwpowodziowe 4 programy walki z suszą 4 Dział. Pilotażowe art. 10	INTERREG III B 13 prog. nakierowanych na współpracę regionalną I planowanie przestrzenne w ramach wyznaczonych stref – tzw. mezoregionów (np. Region Morza Bałtyckiego, Region Adriatycki)
Współpraca międzyregionalna (ang. *Interregional cooperation*) Współpraca między sieciami regionów	–	–	INTERREG III C Paneuropejskie sieci współpracy podzielone na 4 strefy zarządzania
Programy – parasole (ang. *Umbrella programmes*)	–	–	ESPON – program badań rozwoju przestrzennego INTERACT zorientowany na transfer wiedzy i doświadczeń między komponentami A,B,C

Źródło: (Bachtler et al., 2005).

Główne wnioski z powyższej tabeli są następujące:

- Obecny podział na komponenty A, B, C odpowiada trzem celom współpracy terytorialnej – współpracy transgranicznej, transnarodowej i międzyregionalnej;

[12] Po rozszerzeniu UE w 2004 roku o dziesięć nowych krajów liczba programów wzrosła do osiemdziesięciu trzech. Stworzono jedenaście dodatkowych programów współpracy transgranicznej na granicach nowych krajów członkowskich (NMS), głównie z krajami spoza UE.

[13] *European Spatial Development Perspective.*

- Istnieje ciągłość między kolejnymi edycjami – najdłuższa (piętnastoletnia) w przypadku programów współpracy transgranicznej (komponent A);
- Liczba programów w ramach komponentów zmieniała się – większość z nich była łączona; rozszerzenie UE sprawiło, że pojawiły się nowe granice – pola współpracy;
- Od 2000 roku funkcjonują programy spajające trzy komponenty, to tzw. programy parasolowe (ang. *umbrella programmes*).

Kilka słów komentarza wymagają programy – „parasole". Ich celem jest zbieranie doświadczeń, uchwycenie szerszych zależności i wymiana wiedzy powstającej w ramach różnych inicjatyw współpracy terytorialnej. Program ESPON współfinansuje szereg projektów badawczych związanych z przestrzennymi aspektami polityki regionalnej (szeroko pojętym planowaniem i zagospodarowaniem przestrzennym w skali Europy). Natomiast Program INTERACT został pomyślany jako platforma wymiany doświadczeń, *know-how* między zarządzającymi i koordynatorami programów INTERREG. INTERACT współfinansuje również studia badawcze dotyczące praktyk, głównie z dziedziny zarządzania w programach INTERREG[14].

INTERREG, jak już zaznaczono, mógł finansować wyłącznie projekty realizowane w regionach Unii Europejskiej. Na granicach zewnętrznych UE następowała więc silna asymetria, a działania były jednostronne. Dlatego też w połowie lat 90. współpraca terytorialna została rozszerzona na przygraniczne regiony krajów spoza UE, sąsiadujące z regionami unijnymi. Stworzono dla nich fundusze współpracy przygranicznej w ramach programu Phare, a z czasem również Tacis, Cards, Meda.

W teorii programy CBC miały przystawać do programów INTERREG A. W praktyce koordynacja inicjatyw między INTERREG a CBC była bardzo ograniczona. Przyczynami były asymetryczne środki i odmienna organizacja[15].

Dla nowego okresu realizacji polityki regionalnej (2007–2013) Unia Europejska stara się zintegrować działania na granicach zewnętrznych UE. Stworzono jeden instrument finansowy dla regionów położonych na zewnętrznych granicach UE (zarówno regionów z państw członkowskich UE, jak i państw sąsiadujących) – Europejski Instrument Sąsiedztwa i Partnerstwa (ang. *The European Neighbourhood and Partnership Instrument* – ENPI) (European Commission, 2004c).

4.2.3. Organizacja i zarządzanie INTERREG III

Dobiegający końca INTERREG III[16] od samego początku był podzielony na trzy komponenty (odpowiadające trzem aspektom współpracy terytorial-

[14] Szerzej patrz strony internetowe tychże programów: http://www.espon.eu/ oraz www.interact-eu.net.

[15] Programy CBC były finansowane i wdrażane w cyklu rocznym, INTERREG w perspektywie wieloletniej. Programy były koordynowane przez różne dyrekcje Komisji Europejskiej, środki finansowe CBC były znacznie mniejsze, a partnerzy po dwóch stronach granicy mieli zupełnie odmienne potrzeby (regiony spoza UE koncentrowały się na rozbudowie brakującej infrastruktury, regiony UE na miękkiej współpracy).

[16] Okres programowania INTERREG III to lata 2000–2006, ale zgodnie z zasadą n+2 środki z INTERREG III będzie można wydatkować do 2008 roku.

nej), z szeregiem programów w ramach każdego z nich i specjalnymi zasadami zarządzania (European Commission, 2000a). Poniżej scharakteryzowano pokrótce każdy z nich.

Komponent A – współpraca transgraniczna

Ten komponent obejmował pięćdziesiąt trzy programy współpracy transgranicznej realizowane na wydzielonych, sąsiadujących ze sobą obszarach przygranicznych Unii Europejskiej (poziom NTS3 – podregiony). Ich ogólna orientacja to wspieranie centrów rozwoju społeczno-gospodarczego w obszarach przygranicznych (lista programów i mapa zasięgu terytorialnego – patrz Aneks).

Współpracujący aktorzy regionalni i krajowi, działając w ramach instytucji programu, razem wypracowywali jego strategię – główne priorytety, podział środków, szczegółowe zasady promocji i procedury.

Na tej podstawie instytucje zarządzające ogłaszały konkursy na projekty. Projekty zgłaszane przez instytucje, firmy, stowarzyszenia z obszaru objętego programem musiały wpisywać się w założone cele danego programu. Każdy z pięćdziesięciu trzech programów finansował kilkaset projektów. Projekty realizowane w ramach programów IIIA obejmowały zarówno inwestycje infrastrukturalne, jak i „miękkie" projekty z zakresu: stymulowania przedsiębiorczości i rozwoju małych firm, ochrony środowiska, modernizacji i rozbudowy istniejącej sieci powiązań transportowych, wsparcia inicjatyw społeczności lokalnych (mikroprojekty wspierające działania *People-to-People*), rozwoju kapitału ludzkiego i instytucjonalnych form współpracy transgranicznej, innowacji, rozwoju technologicznego, edukacji, współpracy kulturalnej, etc.

Komponent B – współpraca transnarodowa

INTERREG IIIB objął trzynaście programów współpracy ponadnarodowej między władzami krajowymi, regionalnymi i lokalnymi zlokalizowanymi w ramach wydzielonych paneuropejskich regionów przestrzennych. Ich ogólna orientacja to wspieranie (głównie koncepcyjne) zrównoważonego i skoordynowanego rozwoju przestrzennego oraz integracji przestrzennej (lista programów i mapa zasięgu terytorialnego – patrz Aneks).

Podobnie jak w przypadku komponentu A współpracujący aktorzy regionalni, krajowi i lokalni, działając w ramach instytucji programów, razem przygotowywali strategię – główne priorytety, podział środków, szczegółowe zasady i procedury funkcjonowania danego programu.

Następnie instytucje zarządzające ogłaszały konkursy na projekty. Projekty zgłaszane przez instytucje z obszaru objętego programem musiały wpisywać się w założone cele danego programu. Każdy z trzynastu programów INTERREG IIIB finansował kilkadziesiąt projektów, studia koncepcji rozwoju infrastruktury transeuropejskiej, strategie rozwoju przestrzennego, ochrony środowiska, etc. – a więc inicjatywy „miękkie", przygotowujące grunt pod inwestycje infrastrukturalne.

Komponent C – współpraca międzyregionalna

INTERREG III operował nie na poziomie programów, lecz ogólnoeuropejskich projektów wymiany doświadczeń z zakresu polityki regionalnej.

Dla potrzeb zarządzania wydzielono cztery strefy: Północ, Południe, Wschód i Zachód. Każda z nich posiadała swoje instytucje do prowadzenia naboru projektów, ich finansowania i kontroli. Podział ten nie przekładał się na geografię projektów – łączyły one instytucje z różnych części Europy.

O projekty ubiegać się mogły wyłącznie instytucje publiczne. W sumie w ramach komponentu IIIC realizowanych było ponad trzysta projektów. Ich celem była poprawa efektywności polityk i instrumentów rozwoju regionalnego poprzez rozwój sieci współpracy, wymianę doświadczeń, najlepszych praktyk, technik i podejść do zarządzania rozwojem regionalnym.

Typy projektów

Jak już opisano powyżej, tematyka projektów pomiędzy komponentami różniła się. Projekty z programów komponentu A dotyczyły zarówno inwestycji infrastrukturalnych, jak i miękkich inicjatyw, programy IIIB wspierały „miękkie" projekty – strategie i plany przestrzenne, zaś IIIC projekty sieciowej

Tabela 18. Typy projektów INTERREG

Typ projektu / Rodzaj Programu	Granice zewnętrzne i wewnętrzne granice NMS	Komponent A – wewnętrzne granice UE15 (+. LI, CH)	Komponent B – transnarodowy	Komponent C – międzyregionalny
Projekty przygraniczne – (ang. *Close to border*) – realizowanie własnych potrzeb i pomysłów przy uwzględnieniu opinii partnera po drugiej stronie granicy	X			
Projekty przystające (ang. *Parallel projects*) – niezależne wdrażanie dwóch projektów po dwóch stronach granicy, ale projektów które przystają do siebie, uzupełniajac się	X	X		
Projekty ponadgraniczne (ang. *Cross-border projects*) – wykonywane wspólnie po dwóch stronach granicy – jako jedna interwencja w oparciu o wspólnie wypracowaną wizję	X	X		
Wąskie projekty transnarodowe i międzyregionalne – projekty kilku partnerów z kilku krajów/regionów o bardzo szczegółowym, wąskim zakresie tematycznym	X	X	X	X
Szerokie projekty transnarodowe i międzyregionalne – wielu partnerów, szersza orientacja tematyczna – mogą być płytkie (konferencje, wymiana *know-how*) lub trwałe (wspólne planowanie, współpraca po zamknięciu projektu na innych forach).		X	X	X

Źródło: Bachtler et al. (2005).

współpracy. Niemniej jednak w praktyce między tymi projektami z trzech komponentów pojawiły się pewne zbieżności. Analizując sposób współpracy (jej zakres, charakter, intensywność relacji) pomiędzy stronami uczestniczącymi w projektach, można wyróżnić kilka typów projektów współpracy terytorialnej i ich relację z rodzajem programu.

Interesariusze programów INTERREG III

Do programów INTERREG stosowały się te same zasady organizacji wdrażania, co do programów głównego nurtu funduszy strukturalnych. Podstawowe instytucje określało Rozporządzenie na temat Funduszy Strukturalnych (Council of the European Union, 1999). Dodatkowo, dokumenty Komisji precyzowały rozwiązania instytucjonalne specyficzne dla INTERREG (European Commission, 2000a). W rezultacie każdy program INTERREG, niezależnie od komponentu do którego należał, musiał posiadać następujące podstawowe instytucje zarządzające:

Instytucja Zarządzająca – IZ (ang. *Managing Authority – MA*) to instytucja (publiczna lub prywatna) wyznaczona przez państwa członkowskie biorące udział w programie, mająca zapewnić skuteczne i poprawne zarządzanie programem, to znaczy przygotowanie dokumentu programowego – strategii programu (ang. *CIP – Community Initiative Programme*) i jego uzupełnienia (ang. *Programme Complement*), prowadzenie promocji programu i naboru wniosków projektowych, stworzenie systemu zbierania informacji finansowych i statystycznych na temat postępów we wdrażaniu programu, dostarczanie informacji oraz wsparcia dla prac i decyzji podejmowanych przez Komitety Monitorujący i Sterujący, zlecanie ewaluacji (Council of the European Union, 1999: art. 9n, art. 34; European Commission, 2000a: pkt. 30).

Instytucja Płatnicza – IP (ang. *Paying Authority – PA*) to instytucja wyznaczona przez Państwa Członkowskie biorące udział w programie do zarządzania przepływami finansowymi – sporządzania i składania wniosków o płatności oraz otrzymywania płatności z Komisji i przekazywania ich do beneficjentów (wykonawców projektów) (Council of the European Union, 1999: art. 9o).

Wspólny Sekretariat Techniczny – WST (ang. *Joint Technical Secretariat – JTS*) jest instytucją pełniącą funkcje sekretariatu danego programu. Ponadto ma wspomagać prace Organu Zarządzającego przede wszystkim w zakresie promocji programu, naboru wniosków, wsparcia dla procedury wyboru projektów przez Komitet Sterujący, bieżących kontaktów z beneficjentami (European Commission, 2000a: pkt. 30).

Komitet Monitorujący (ang. *Monitoring Committee – MC*) – składa się z przedstawicieli organów regionalnych i lokalnych, organów państwowych (jeżeli wyrażą takie życzenie), przedstawiciela Komisji Europejskiej (w roli obserwatora), partnerów gospodarczych i społecznych, przedstawicieli NGOs. Komitet zatwierdza uzupełnienie programu i w trakcie jego funkcjonowania wprowadza ewentualne poprawki w dokumentach programowych, ustala kryteria wyboru projektów, monitoruje i ocenia program jako całość (Council of the European Union, 1999: art. 35; European Commission, 2000a: pkt. 28).

Komitet Sterujący (ang. *Steering Committee – SC*) – skład Komitetu Sterującego jest analogiczny do Komitetu Monitorującego (dopuszczalne jest, by obowiązki SC pełnił MC działając w jego roli). Komitet Sterujący dokonuje ostatecznego wyboru projektów, które mają być finansowane z danego programu, a także prowadzi skoordynowane monitorowanie wszystkich operacji programu.

Ponadto zarówno rozporządzenie, jak i przepisy wykonawcze dopuszczały tworzenie instytucji pomocniczych (punktów kontaktowych, koordynatorów regionalnych, etc).

W rzeczywistości powyższa prosta lista instytucji owocowała w różnych programach różnymi konfiguracjami. Zróżnicowanie przebiegało w odniesieniu do dwóch głównych kwestii:

- różnego dzielenia się krajów i regionów odpowiedzialnością w ramach podstawowych instytucji (i różnic w delegacji kompetencji pomiędzy instytucjami np. z IA do JTS),
- tworzenia różnych dodatkowych, wspomagających instytucji.

Co do kwestii pierwszej (różnicy w podziale kompetencji), występowało kilkanaście kombinacji, przykładowo: IZ i IP w dwóch różnych krajach, w tym samym kraju ale w różnych instytucjach, w tym samym kraju i tej samej instytucji, jeden WST o słabej i wspomagającej pozycji wobec IZ, WST o silnej pozycji (główna instytucja prowadząca bieżące zarządzanie programem), WST z delegacjami w regionach objętych programem, KM i KS łączone, KM i KS jako dwie niezależne instytucje z różnym składem personalnym, etc. (Bachtler et al., 2005: 91–96). Na podstawie podziału kompetencji można było określić poziom zintegrowanego zarządzania programem. Programy o jednym zestawie instytucji (bez struktur lustrzanych tworzonych dla każdego uczestniczącego regionu) były najbardziej zintegrowane.

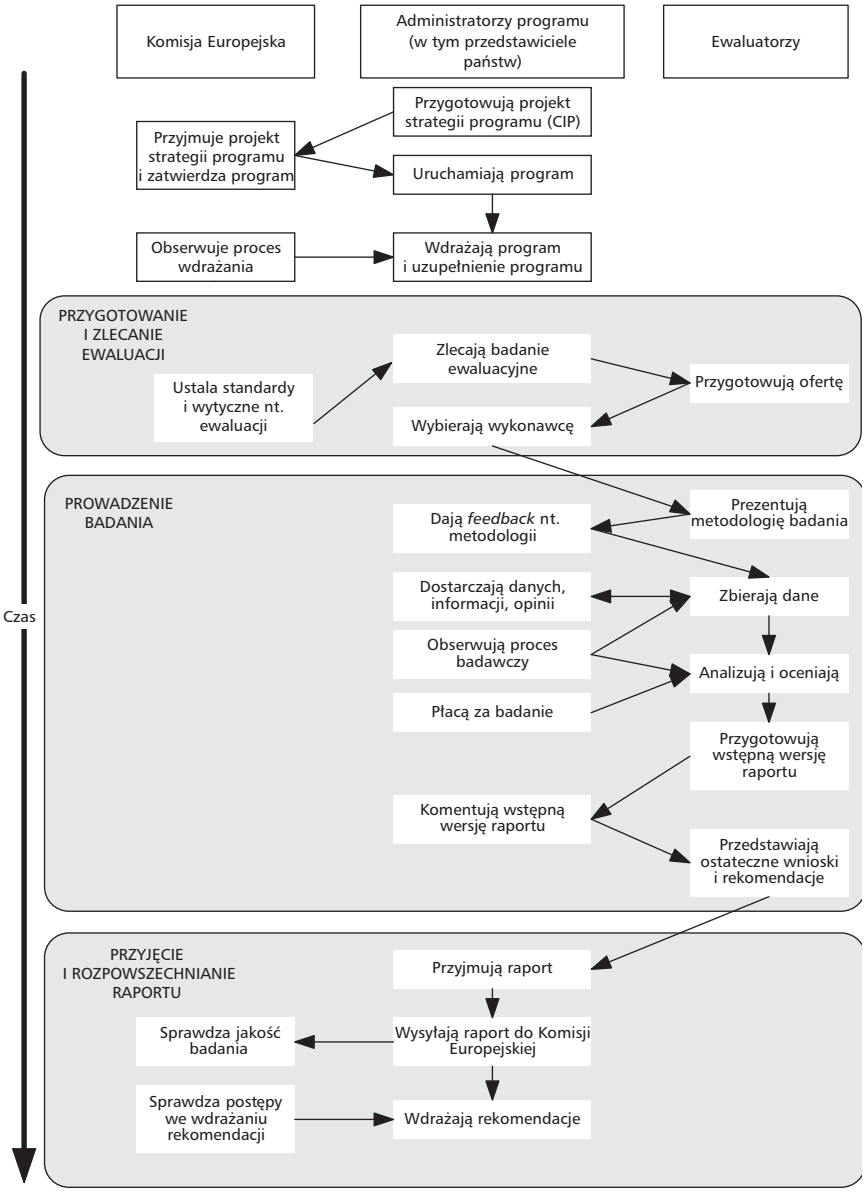
W przypadku kwestii drugiej, dodatkowe instytucje tworzone *ad hoc* można połączyć według ich funkcji w kilka grup (Bachtler et al., 2005: 96–97):

- Instytucje promujące i animujące program, typowe dla dużych ponadnarodowych programów obejmujących wiele państw, miały za zadanie zbliżać program do lokalnych aktorów, dawać pierwsze, bieżące wsparcie beneficjentom programu);
- Punkty koordynacji – będące głównymi punktami wymiany informacji i koordynacji pomiędzy lokalnymi animatorami programu;
- Instytucje wspierające – dające wsparcie w zarządzaniu programem dla głównych instytucji (sekretariatu, instytucji zarządzającej, komitetów) szczególnie na poziomie kontaktów z administracjami krajowymi czy też w trakcie fazy naboru wniosków, późniejszego monitoringu;
- Platformy współpracy, których głównym zadaniem było zagwarantowanie reprezentacji krajowych interesów na poziomie programu poprzez włączanie różnych polityków i oficjalnych przedstawicieli w strategiczne dyskusje nad programem. Ich funkcja była z reguły doradcza.

Podsumowując, na potrzeby niniejszej pracy można sformułować następującą typologię interesariuszy ewaluacji programów INTERREG:

- Kluczowe grono administratorów – osoby kluczowe z Instytucji Zarządzającej (IZ) i Sekretariatu Programu (WST), na co dzień zaangażowane w sprawy programu;

Schemat 6. Relacje interesariuszy INTERREG w ewaluacji średniookresowej



Opracowanie: Karol Olejniczak na podstawie Eser, Nussmeller (2006).

- Szersze grono zarządzających – członkowie Komitetu Sterującego (KS) i Komitetu Monitorującego (KM);
- Uczestnicy programu – personel instytucji wspomagających (np. punktów kontaktowych), beneficjenci, aplikujący o fundusze;
- Decydenci wyższego szczebla – politycy, przedstawiciele Komisji Europejskiej;

- Aktorzy zewnętrzni – poziom europejski: Program INTERACT, personel zarządzający analogicznymi programami, eksperci współpracy terytorialnej, organizacje zajmujące się współpracą terytorialną;
- Opinia publiczna – mass-media, dziennikarze prasowi i telewizyjni, internetowe serwisy informacyjne) i szersza opinia publiczna.

Należy zaznaczyć, że w procesie ewaluacji kluczowe role odgrywali zwykle przedstawiciele Komisji Europejskiej, personel programu (IZ, WST), ewaluatorzy oraz – gdy taka się pojawiała – Grupa Sterująca Ewaluacją. Standardowy schemat relacji i ich sekwencje dla ewaluacji średniookresowej podsumowano schemacie 6.

4.2.4. Specyfika ewaluacji programów współpracy terytorialnej

INTERREG III podlegał takim samym wymogom ewaluacji, jak programy z głównego nurtu funduszy strukturalnych. Wymagane były:

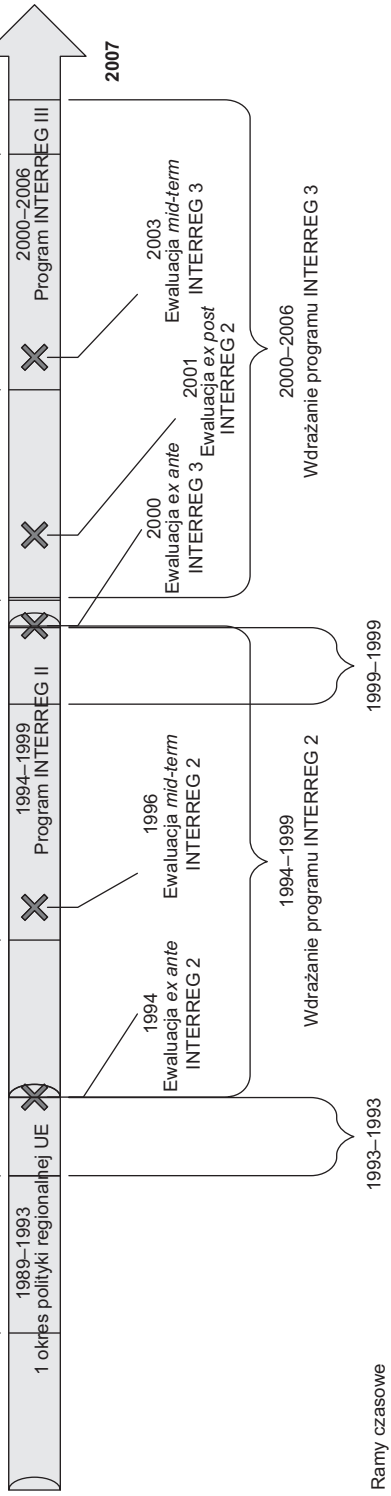
- **Ewaluacja *ex ante*** – zlecana i przyjmowana przez instytucję zarządzającą oraz komitet monitorujący programu, a obejmująca zakres pytań, problemów badawczych i ogólnych narzędzi wyznaczony przez Komisję Europejską;
- **Ewaluacja** średniookresowa – zlecana i przyjmowana przez IZ programu, jak powyżej obejmująca zakres pytań i problemów badawczych określony dokumentem Komisji Europejskiej;
- **Uaktualnienie ewaluacji** średniookresowej – zlecane na zewnątrz przez IZ programu, w swej części merytorycznej będące w zasadzie powtórzeniem ewaluacji średniookresowej;

Tabela 19. Wytyczne KE dotyczące zakresu ewaluacji średniookresowej INTERREG

Wytyczne Komisji Europejskiej w zakresie ewaluacji średniookresowej programów INTERREG
Głównym celem wyznaczonym ewaluacjom średniookresowym była ocena struktury programu i jego wstępnych rezultatów oraz przedstawienie rekomendacji mających poprawić realizację celów programu. Każde badanie ewaluacyjne musiało objąć pięć następujących kwestii: • Sprawdzenie aktualności i trafności analizy wykonanej w ramach ewaluacji *ex ante* (w tym sprawdzenie aktualności analizy SWOT – głównie zmian w kontekście programu) oraz zaproponowanie rekomendacji, które powiążą program ściślej ze zmieniającym się środowiskiem zewnętrznym; • Powtórne zweryfikowanie adekwatności (*relevance*) i spójności (*consistency*) strategii programu – sprawdzenie, czy obrane priorytety i cele wciąż trafnie odpowiadają potrzebom, zbadanie, na ile realizowane działania wywołują oczekiwane efekty, zweryfikowanie po dwóch latach realizacji programu, na ile założone cele są rzeczywiście możliwe do osiągnięcia, wreszcie poprawienie wewnętrznej spójności – między poszczególnymi priorytetami i działaniami; • Sprawdzenie poprawności kwantyfikacji celów – wskaźników produktu, rezultatu i oddziaływania, sprawdzenie, czy zastosowane wskaźniki odzwierciedlają charakter realizowanych działań, czy mają wartość informacyjną (pomagają w monitoringu programu); • Zbadanie efektywności programu i oczekiwanego oddziaływania społecznego-gospodarczego, w tym ocena Wspólnotowej Wartości Dodanej programu i jego wpływu na kwestie horyzontalne, a w przypadku zidentyfikowania problemów zaproponowanie przesunięć środków; • Sprawdzenie jakości funkcjonowania wspólnych struktur zarządzania i monitoringu (zbierania danych, wdrażania, kryteriów selekcji projektów etc.). Komisja wyznaczyła również kalendarz badań. Badania miały być zlecone najpóźniej do marca 2003 roku, wykonane między marcem a lipcem, zaś raporty z badań zatwierdzone do grudnia 2003 roku.

Źródło: European Commission (2002b).

Schemat 7. Kalendarz planowania i ewaluacji Inicjatywy Wspólnotowej INTERREG



Ramy czasowe
okresy polityki regionalnej:
- początek roku budżetowego
- koniec roku budżetowego

Planowanie inicjatywy INTERREG:
- początek – data publikacji raportu kohezyjnego lub początek konsultacji (*Commission Green paper*)
- koniec – zatwierdzenie komunikatu Komisji na temat INTERREG (*CEC Communication to MSs laying down guidelines for a Community Initiative INTERREG*)

Planowanie programu:
- początek – w drugiej połowie wdrażania programu edycji poprzedniej
- koniec – przesłanie do Komisji Europejskiej propozycji programu operacyjnego (*Community Indicative Programme – CIP*)

Wdrażanie Programu:
- początek – data zatwierdzenia uzupełniania programu (*Programme Complement*)
- koniec – koniec okresu programowania

Ewaluacje – zaznaczone są okresy prowadzenia badań (badanie + publikacja wyniku, nigdy nie przekraczały 12 miesięcy)

Źródło: opracowanie własne.

- **Ewaluacja *ex post*** – będzie zlecana przez Komisję Europejską dla wszystkich programów INTERREG III (według planów ma być wykonana po zamknięciu inicjatywy – około 2008 roku).

Dopuszczane były również tzw. **ewaluacje *on-going*/*ad hoc*** (bieżące). Decyzję o ich podjęciu podejmował Komitet Monitorujący programu we współpracy z Instytucją Zarządzającą i Sekretariatem Technicznym.

W latach 1994–1999 (INTERREG II) praktyka ewaluacji była nieregularna i zróżnicowana (Eser, Nussmueller, 2006: 251)[17]. Dopiero dla INTERREG III regulacje Komisji wymusiły przeprowadzenie w miarę jednoczesnego i zestandaryzowanego ćwiczenia ewaluacyjnego dla wszystkich programów, szczególnie w odniesieniu do ewaluacji średniookresowej.

Dotychczas Komisja Europejska przygotowała dwa specjalne przewodniki dostosowane do specyfiki INTERREG, a określające szczegółowy zakres tematyczny badania i główne pytania ewaluacyjne. Dotyczyły one ewaluacji *ex ante* i średniookresowej (European Commission, 2000f, 2002b). W zasadzie pokrywały się z wytycznymi dla ewaluacji głównych programów Funduszy Strukturalnych. Podsumowanie szczegółowych wytycznych dla ewaluacji średniookresowej, będącej przedmiotem tej pracy, zawiera tabela 19.

Analizując powyższe wytyczne warto zauważyć za Eser i Nussmueller (2006), że przed ewaluacjami średniookresowymi postawiono dwoiste cele. Z jednej strony miały one rozliczać, podsumowywać dotychczasowe osiągnięcia (a więc pełnić funkcję podsumowującą – przede wszystkim funkcję wzmacniania odpowiedzialności), a z drugiej służyć wewnętrznemu procesowi uczenia się.

Kilka słów należy poświęcić planowanemu kalendarzowi badań. Ewaluacja średniookresowa jest jednym z elementów cyklu planowania i wdrażania polityki – w tym wypadku współpracy terytorialnej. Warto więc pokazać kalendarz założony przez Komisję Europejską w szerszej perspektywie planowania i oceny wcześniejszych i późniejszych edycji programów. Przedstawiono to na schemacie 7.

Pozornie programy INTERREG dotyczyły podobnej problematyki, choć o dużo mniejszej skali niż główny nurt funduszy strukturalnych (inwestycje infrastrukturalne, ale również projekty „miękkie"). Istnieje jednak szereg znaczących różnic, wymagających specjalnego podejścia do badania. Specyfikę ewaluacji programów INTERREG można podsumować w siedmiu punktach:

- Złożony kontekst – programy pokrywały odmienne jednostki przestrzenne o odmiennych kulturach, powiązaniach funkcjonalnych, różnym poziomie rozwoju. Ewaluacja w takim wypadku musiała obejmować obydwie strony granicy, traktować obszar przygraniczny jako jednolitą przestrzeń funkcjonalną. To wymagało zintegrowanego podejścia;
- Dwoista realizowały zarówno cele rozwojowe (wyrównywanie potencjałów, wzmacnianie rozwoju regionalnego), jak i cel integracji (przełamy-

[17] W części programów zdecydowano się na ewaluacje wewnętrzne, w innych na zewnętrzne. Większość raportów miała charakter dokumentów wewnętrznych, tylko niektóre zostały upublicznione. Na przykład dla ewaluacji *ex post* INTERREG II ponad połowa studiów jest niedostępna.

wanie granicy jako bariery, łączenie obszarów i partnerów po obu stronach granicy). Cele rozwojowe dawały efekty bardziej namacalne, zaś cele integracyjne przynosiły „miękkie" rezultaty. W ewaluacji wymagało to użycia zróżnicowanych metod analizy oddziaływania;

- Multisektorowa orientacja – programy funduszy strukturalnych głównego nurtu były zwykle zorientowane sektorowo. Tymczasem programy INTERREG pokrywały różne sektory (biznes, transport, ochrona środowiska, innowacje), a interwencje były prowadzone na różnych poziomach (regionalnym, subregionalnym, lokalnym). To wymagało multisektorowego podejścia w badaniach, zróżnicowanych technik oraz mieszanki metod ilościowych i jakościowych;
- Efekty obserwowane lokalnie – ograniczona skala interwencji sprawiała, że efekty projektów były obserwowalne głównie w skali lokalnej (gminy, kilku gmin). Trudno było je więc badać za pomocą analiz statystycznych na poziomie regionu. Bardziej wskazane wydawało się użycie metod jakościowych, studiów przypadku, etc.;
- Niematerialne efekty (ang. *intangible effects*) – ponieważ budżety INTERREG były niewielkie, to promowane były tzw. miękkie projekty (budowanie sieci współpracy, wspólne wypracowywanie strategii), a w mniejszym stopniu projekty czysto infrastrukturalne. W rezultacie prosty model: nakłady-proces-produkty-efekty był trudniejszy do zastosowania (trudniej było zmierzyć efekt i wypreparować relację przyczynowo-skutkową). Od ewaluatorów wymagało to podejścia jakościowego;
- Proces równie ważny jak efekty – kluczowym elementem projektów współpracy był sam proces – robienie rzeczy wspólnie. Prosty pomiar nakład-produkt miał ograniczone zastosowanie, gdyż nie mówił nic o jakości relacji. Ewaluacja musiała więc uwzględniać zarówno tradycyjne efekty, jak i jakość współpracy, która do nich doprowadziła (intensywność interakcji, ich trwałość, wspólne uczenie się). Tu również zastosowanie znajdowały metody jakościowe, narzędzia analizy procesów – zbliżone do ewaluacji innowacji;
- Perspektywy różnych interesariuszy – zaangażowanie aktorów z różnych kultur administracyjnych, różnych poziomów i sektorów sprawiało, że proces wdrażania był bardziej skomplikowany. Gdy dochodziło do momentu oceny programu, te zróżnicowane punkty widzenia winny zostać uwzględnione. Co więcej, ewaluacja mogła się również stać narzędziem integrowania uczestników, włączania ich w program i budowania ich poczucia współodpowiedzialności. Wskazane wydawały się więc podejścia interaktywne i negocjacyjne.

Podsumowując opis Inicjatywy Wspólnotowej INTERREG raz jeszcze należy podkreślić jej wartość jako przedmiotu studiów związanych z procesami zarządzania i uczenia się w środowisku ponadnarodowym i międzykulturowym. INTERREG był swoistym laboratorium współpracy *sensu stricte* europejskiej i ponadnarodowej, a także, z racji swej specyfiki, wyzwaniem dla ewaluacji.

4.3. MODEL ANALIZY WYKORZYSTANIA EWALUACJI

4.3.1. Prezentacja modelu

Wypracowany metamodel został przedstawiony na schemacie 5 w Aneksie 6.2. Oparto go na wnioskach z przeglądu typologii, modeli teoretycznych i wyników badań empirycznych prezentowanych w literaturze poświęconej wykorzystaniu ewaluacji programów publicznych[18]. Źródła te były omawiane i analizowane we wcześniejszych rozdziałach tej pracy (przede wszystkim w rozdziale 2) – na ich podstawie budowano cząstkowe kategorie i elementy modelu. Dodatkowym źródłem przy jego modelu były konsultacje z wybranymi teoretykami i praktykami ewaluacji (dotyczące listy źródeł, planu badawczego i wstępnej wersji modelu)[19].

Ogólna charakterystyka modelu

Wypracowany model ma dwie charakterystyczne cechy warte podkreślenia, a różniące go od wcześniejszych prac z tego zakresu.

Po pierwsze, zastosowano w nim **podejście systemowe**. Model jest próbą przedstawienia w sposób całościowy, w jednym ciągu logicznym, relacji między zmiennymi niezależnymi, wyjaśniającymi (czynnikami wykorzystania ewaluacji), a zmiennymi zależnymi, wyjaśnianymi – rodzajami użycia ewaluacji i szerszymi funkcjami, które spełnia w sektorze publicznym. Źródłami inspiracji były między innymi prace Cousinsa i Leithwooda (1986) oraz Burke'a Johnsona (1998).

Po drugie, zastosowano **podejście procesowe**. Proces ewaluacji był rozumiany jako proces tworzenia i wykorzystania wiedzy. Przedstawiono go analogicznie do ciągu logicznego stosowanego w planowaniu, wdrażaniu i analizie projektów oraz programów publicznych (European Commission, 2000e, 2004b). To ujęcie zostało zaczerpnięte z prac Cousinsa (2003) oraz Marka i Henry'ego (2004) i zaadaptowane do specyfiki Unii Europejskiej[20]. Etapy procesu ewaluacyjnego przedstawiono więc jako „proces produkcyjny". Jednak w tym wypadku produktami i efektami projektu nie są infrastruktura, szkolenia czy wsparcie przedsiębiorczości i ich użyteczność (jak w przypadku projektów funduszy strukturalnych), a raport ewaluacyjny, powstała wiedza i jej użycie.

---

[18] Na tym etapie badania wykorzystano standardowe metody i techniki przeglądu literatury (Fink, 2005; Hart, 1998).

[19] Konsultowanymi badaczami byli: prof. Marvin Alkin – University of California Los Angeles (UCLA), Eliot Stern – Tavistock Institute, prof. David Reingold – School of Public & Environmental Affairs, Indiana University (SPEA), prof. John Bachtler – European Policies Research Center – Strathclyde University.

[20] Dalszymi inspiracjami były prace Johna Owena (2007), Daniela Stufflemeama (2000a, 2005) czy Malcolma Provusa (1971).

Główne segmenty modelu

Model przedstawiający ciąg wykorzystania ewaluacji, składa się z pięciu segmentów – następujących po sobie etapów tegoż procesu (określona jest więc sekwencja czasowa i logiczna). Są to:

- wkład do ewaluacji (ang. *evaluation input*), czyli czynniki kontekstowe oraz wkład wnoszony na początku procesu przez uczestników ewaluacji (tak wykonawców, jak i zlecających);
- proces ewaluacji (ang. *evaluation process*) – standardowa sekwencja działań podejmowana podczas unijnego badania ewaluacyjnego (jego szczegółowe zaplanowanie i realizacja);
- produkty ewaluacji (ang. *evaluation outputs*) – opisują cechy bezpośrednich produktów powstałych z badania – wniosków i przygotowanego materiału, a więc raportu, stworzonej wiedzy i procesu rozpowszechnienia informacji;
- wykorzystanie ewaluacji (ang. *evaluation utilization*) – ten segment oraz segment kolejny zawierają zmienne zależne – różne sposoby użycia ewaluacji (włączając w to brak użycia, jak i manipulację wynikami) w różnej perspektywie czasowej;
- oddziaływanie – funkcje ewaluacji (ang. *evaluation influence – functions of evaluation*) opisują jej szersze oddziaływanie – jej przełożenie na zarządzanie w sektorze publicznym.

W dalszej części tego podrozdziału opisano szczegółowo zawartość kolejnych segmentów modelu, podając źródła literatury, bądź też odniesienia do fragmentów przeglądu literatury na podstawie których były wprowadzane. W nawiasach podawano użyte bądź też wypracowane (na podstawie własnej analizy) terminy angielskie, po to by ułatwić czytelnikowi połączenie niniejszego tekstu z bogatą anglosaską literaturą przedmiotu.

Segment 1 – WKŁAD (ang. *evaluation input*)

W ramach tego segmentu umieszczono dwie grupy czynników.

Pierwszą grupą są **warunki kontekstowe** ewaluacji (ang. *context*) – czyli to, co zewnętrzne dla podejmowanego badania ewaluacyjnego. W ramach tej grupy wyróżniono:

- kulturę ewaluacyjną danej administracji/kraju (ang. *evaluation culture of the country*), przede wszystkim tradycję w wykorzystywaniu badań ewaluacyjnych;
- regulacje i wytyczne UE dotyczące ewaluacji (ang. *EU regulations*) – chodzi tu przede wszystkim o omawianą wcześniej serię Dokumenty Robocze KE, które narzucają minimalny zestaw pytań ewaluacyjnych i metody dla każdego z typów ewaluacji;
- kontekst polityczno-decyzyjny (ang. *window of opportunity*), czyli w jaki moment cyklu decyzyjnego, debaty politycznej wokół programu i polityki trafia dana ewaluacja (nazywany też „momentem" badania);
- formalny moment podjęcia badania – typ czasowy ewaluacji (ang. *research timing*) – czyli to, w którym momencie trwania programu podejmowane jest

badanie, czy jest to ewaluacja *ex ante*, *mid-term* (średniookresowa), *on-going* (bieżąca) czy też *ex post*;

- charakter i stan ewaluandy – badanego programu (ang. *characteristic & stage of the evaluand*) – chodzi o wielkość programu, jego tematykę, stan zaawansowania, organizacyjną strukturę wdrażania, ale również o to, czy istnieje konsensus, czy też konflikt wokół programu oraz czy jest, czy też nie ma presji do zmian w programie.

Drugą grupą jest **wkład uczestników** (ang. *input from participants*). Są to rzeczy, zachowania, warunki wnoszone przez głównych uczestników ewaluacji – zarówno wykonawców badania, jak i zlecających ewaluację, czyli odbiorców badania.

Chodzi konkretnie o:

- oczekiwania zlecających (ang. *expectations of the contracting authority*) – założone cele ewaluacji tak formalne, jak i ukryte, nieformalne (ang. *hidden agenda*) oraz wiążące się z tym potrzeby informacyjne, oczekiwania wiązane z ewaluacją;
- precyzyjność zapisów Zakresu Zadań (ang. *quality of Terms of Reference*), czyli na ile jasno i konkretnie zlecający sformułował i zapisał w Zakresie Zadań swoje wymagania dotyczące merytorycznej strony kontraktu ewaluacyjnego;
- przeznaczone zasoby (ang. *assigned resources*), czyli budżet przypisany do realizacji badania; środki finansowe w sposób oczywisty warunkują zakres ewaluacji, dobór metod, skalę badania;
- przeznaczony czas (ang. *assigned time*), czyli długość kontraktu ewaluacyjnego, czy jest to krótkie badanie, czy też długie, kilkunastomiesięczne, umożliwiające szersze spojrzenie na temat;
- charakterystyka uczestników (ang. *users characteristics*), ich struktura organizacyjna, zdolność uczenia się, obecność liderów – jednostek zainteresowanych promowaniem badania;
- charakterystyka ewaluatorów (ang. *evaluators characteristic*) – doświadczenie zespołu ewaluacyjnego, „background" zawodowy i posiadane umiejętności, kultura organizacyjna (w przypadku dużych firm).

Segment 2 – PROCES EWALUACJI (ang. *evaluation process*)

Proces badawczy podzielono na kilka standardowych etapów: przygotowanie kontraktu, projektowanie badania – wybór podejścia i metod (w tym przegląd poprzednich studiów z danej tematyki), zbieranie danych, analizę, ocenę, pisanie wniosków i redagowanie raportu (European Commission, 1999b: vol. 1; Herman et al., 1987: 26; Pawson, 2006: 103). W niniejszym modelu zastosowano podział bazujący na podejściu unijnym – MEANS i Evalsed (porównaj rozdział 3.3.4), wyróżniając trzy fazy: strukturalizacji, obserwacji i analizy[21], oceny. W każdej z faz wyróżniono szereg czynników warunkujących wykorzystanie ewaluacji.

[21] Połączono te dwie fazy z uwagi na występujący w nich analogiczny zakres czynników wpływu na ewaluację oraz ze względu na płynność granic między zbieraniem a analizą danych.

W **fazie strukturalizacji** (ang. *structuring phase*) są to:

- cechy pytań priorytetowych (ang. *priority questions*) – chodzi po pierwsze o zakres pytań – na ile główne pytania i hipotezy badawcze są spójne, a na ile rozstrzelone i szeroko zakreślone (w formie „listy życzeń"). Po drugie zaś chodzi o to, czy podnoszone kwestie są kontrowersyjne, postrzegane przez różnych interesariuszy jako newralgiczne;
- wybrane podejście badawcze (ang. *evaluation design*) – chodzi o wybór sposobu podejścia do badania. Bazując na typologiach Alkina i jego współpracowników (2006), a także Dahler-Larsena (2005) wyróżnionych zostało pięć opcji – ewaluacja jako: badanie naukowe, ewaluacja oparta na teorii, zorientowana na cele, na zarządzanie, ewaluacja responsywna i partycypatywna (porównaj rozdział 2.2.3);
- paleta narzędzi i technik (ang. *chosen toolbox*) planowanego zbierania danych, analizy, oceny. W przypadku INTERREG szczególnie ważnym jest dobranie ich do specyfiki programu. Wybory te są w dużej mierze warunkowane obranym podejściem badawczym;
- role – relacje – interakcje (ang. *role, relations, interactions*) pomiędzy zlecającymi, ewaluatorami i interesariuszami na etapie precyzowania metodologii i planu badania. Możliwe są opcje: podejście sztywne, statyczne, interaktywne (patrz rozdział: 2.3.3).

Czynnikami wyróżnionymi w **fazie obserwacji i analizy** (ang. *observation & analysis*) są:

- jakość danych zastanych, tzw. wtórnych (ang. *quality of monitoring data*) – pochodzących przede wszystkim z monitoringu programu. Jakość oznacza dostępność danych, kompleksowość, przystawalność do działań programu;
- zróżnicowanie metod i źródeł (ang. *mixture of methods and sources*) ma kilka aspektów. Po pierwsze, chodzi o mieszankę metod jakościowych – miękkich i ilościowych – *quasi*-eksperymentalnych[22]. Po drugie, mieszankę podejść interaktywnych i statycznych[23]. Po trzecie, chodzi o zróżnicowanie danych, opieranie się na kilku źródłach jednocześnie – na tzw. triangulacji (ang. *use of triangulation*) – zarówno na etapie zbierania danych, jak i ich analizy (spojrzenia na to samo zagadnienie z kilku perspektyw, przy użyciu kilku źródeł);
- poprawność zbierania danych pierwotnych przez ewaluatorów (ang. *quality of data collection*) – chodzi o trzymanie się standardów naukowych w prowadzeniu badań ilościowych i jakościowych;
- poprawność analizy (ang. *correctness of data analysis*) – tak jak powyżej chodzi o przestrzeganie standardów metod i warsztatu badawczego nauk społecznych;

---

[22] Metody „miękkie" to wywiady, ankiety, obserwacja uczestnicząca – w tym wizyty studyjne, warsztaty z beneficjentami. Metody twarde to eksperymenty, *quasi*-eksperymenty, modele ekonometryczne, modele teoretyczne.

[23] Metody statyczne to np. użycie danych administracyjnych i przegląd dokumentów (*desk-research*), analiza SWOT, modele logiczne, mapy oddziaływania, analiza zmian udziałów, modele ekonometryczne, analiza kosztów-korzyści/efektywności, zaś metody interaktywne to np. panele dyskusyjne, warsztaty, badania ankietowe, wywiady (szerzej patrz tabela 12; opisy metod: GHK Consultants 2006).

- rola ewaluatorów (ang. *evaluators' role*) – zgodnie z teoriami ewaluacji kluczową w procesie badania jest rola ewaluatora. Można wyróżnić trzy role ewaluatora, w dużej mierze zależne od przyjętego podejścia badawczego: rola naukowca, konsultanta, animatora-edukatora (porównaj: 2.3.3)[24]. Natomiast w momencie pojawienia się silnych dychotomii konsensusu lub konfliktu wokół programu, w momencie wystąpienia presji zmian lub jej braku, role dodatkowe to rola mediatora (konflikt i brak presji zmian), arbitra (konflikt i silna presja zmian), eksperta budzącego świadomość (konsensus i brak potrzeby zmian), reformatora (konsensus i potrzeba zmian);
- relacje – interakcje (ang. *relations, interactions*) pomiędzy zlecającymi, ewaluatorami i interesariuszami, a nawet szerszą wspólnotą. Relacje wynikające w dużej mierze z wybranego podejścia badawczego oraz wybranej przez ewaluatora roli mogą mieć cztery formy: izolacji, ograniczonego kontaktu (relacje jednostronne), interakcji – procesu dwukierunkowego, interakcji szerokich (kontakt również z instytucjami i sieciami tematycznymi spoza programu) (porównaj: rozdział 2.3.3).

Wreszcie w **fazie oceny** (ang. *assessment phase*) jako wpływające na wykorzystanie zidentyfikowano:

- zróżnicowanie technik oceny programu (ang. *differentation of assessment techniques*) – od strony wykorzystanych źródeł (ankiety, opinie z wywiadów, dane z monitoringu, obserwacje własne), włączonych osób i uwzględnionych perspektyw (ocena ewaluatorów – ekspertów, samoocena beneficjentów, ocena zarządzających);
- poprawność procesu oceny (ang. *correctness of the assessment process*) – czy od strony metodycznej i technicznej poprawnie użyto danych narzędzi, obiektywnie pokierowano procesem oceny;
- role – rolę ewaluatora na tym etapie można opisać analogicznie jak w fazie obserwacji i analizy (naukowiec, konsultant, animator-edukator), natomiast w odniesieniu do aspektu konfliktów-presji zmian w programie role: mediatora (konflikt i brak presji zmian), arbitra (konflikt i silna presja zmian), eksperta budzącego świadomość (konsensus i brak potrzeby zmian) i reformatora (konsensus i potrzeba zmian);
- relacje – interakcje (ang. *relations, interactions*) pomiędzy zlecającymi, ewaluatorami i interesariuszami. Chodzi o stopień zaangażowania interesariuszy w ocenę programu na podstawie zebranych danych – a więc o stopień współoceny. Te relacje mogą mieć trzy formy: izolacji, ograniczonego kontaktu (relacje jednostronne), interakcji – procesu dwukierunkowego.

Segment 3 – PRODUKTY EWALUACJI (ang. *evaluation outputs*)

Produkty ewaluacji wpływające na jej wykorzystanie przedstawione są w czterech grupach. Pierwszą stanowią kontakty i relacje będące wynikiem badania, drugą są cechy raportu ewaluacyjnego – jego jakość i forma, trzecią

[24] Ten podział jest zmodyfikowaną wersją kategoryzacji zaproponowanej przez Morabito (2002).

grupą wiedza, jaką przyniósł raport i badanie, zaś ostatnią grupą produktów proces udostępniania i rozpowszechniania wyników.

Stworzone kontakty (ang. *established contacts & networks*) – uznano, że ich rozległość i intensywność (zbudowany klimat zaufania) może mieć kilka aspektów:

- kontakty ze zlecającym (ang. *relations with contracting authority*) – formalne, słabe i nietrwałe lub też dobre, gwarantujące, że zlecający staną się promotorami („mecenasami") badania;
- kontakty z interesariuszami (ang. *stakeholders network*) – brak sieci lub zbudowana sieć interakcji i komunikacji między interesariuszami (np. zarządzającymi, beneficjentami, aplikującymi, grupami tematycznymi), albo nawet zacieśnione interakcje oraz koalicja dająca mobilizację i wsparcie wyników;
- szersza sieć (ang. *wider networking*) – brak nawiązanych kontaktów zewnętrznych – poza programem – lub też wyjście poza program i stworzenie kanałów komunikacji z badaczami analogicznych interwencji, innymi programami, europejskimi sieciami tematycznymi.

W przypadku **raportu ewaluacyjnego** jako cechy wpływające na wykorzystanie ewaluacji wyróżniono jego poprawność metodyczną i jakość formy (na ile skutecznie przechodzi „test prawdy" odbiorców), a konkretnie:

- poprawność zakresu badania – czy pokrywa zlecony zakres zadań, do jakiego stopnia traktuje program całościowo;
- poprawność planu badawczego – trafność wewnętrzna, trafność teoretyczna, uwzględnienie specyfiki programu;
- poprawność doboru i użycia danych – rzetelność badania, triangulacja;
- przejrzystość i poprawność analizy – jasność procesu dedukcji/indukcji, trafność konkluzji statystycznych;
- wiarygodność wyników – ich oparcie w zebranych danych i dokonanych obserwacjach;
- jakość rekomendacji – ich bezstronność, konkretność, operacyjność;
- przystępność i przejrzystość raportu – tzw. przyjazność dla użytkownika (ang. *user friendliness*), komunikatywność formy prezentacji – odesłania, podsumowania, streszczenia, wyróżnienia, „mapy" dla różnych typów odbiorców.

Kolejną grupę czynników – **rodzaj wiedzy** będącej produktem badania ewaluacyjnego (wnioski i rekomendacje) ujęto w pięciu kategoriach (bazując na: Lundvall, Johnson, 1994; Nutley et al., 2003). Należy podkreślić, że kluczowym czynnikiem jest w nich to, na ile oferowana wiedza jest nowa dla jej odbiorców. I tak odbiorcy z badania i raportu mogą:

- dowiedzieć się o… (ang. *know-about*) – raport dostarcza nowej wiedzy o problemach, wyzwaniach, konfliktach, potencjalnych zagrożeniach, ale i szansach, które pojawiają się w szerszym kontekście, otoczeniu programu;
- dowiedzieć się co… (ang. *know-what*) – raport zawiera informacje, co działa a co nie, które działania i rozwiązania przynoszą oczekiwane efekty lub nieoczekiwane sukcesy, a które są opóźnione, sprawiają problem, nie działają;
- dowiedzieć się dlaczego… (ang. *know-why*) – raport opisuje i wyjaśnia źródła oraz przyczyny tak problemów, jak i sukcesów, interpretuje dane procesy,

odnosi program do szerszej wiedzy z danej tematyki, pokazuje go na tle teorii, norm, założeń polityk, trendów międzynarodowych;

- dowiedzieć się jak i kiedy… (*know-how & when*) – raport zawiera konkretną wiedzę praktyczną, w jaki sposób i kiedy poprawić lub zmienić interwencję, jak i w jakiej perspektywie czasowej efektywnie wdrażać usprawnienia, może również zawierać alternatywne rozwiązania lub przykłady *know-how* z innych programów;
- dowiedzieć się kto… (ang. *know-who*) – raport wskazuje, kto (jakie instytucje, aktorzy) powinien być odpowiedzialny za podjęcie danych działań naprawczych lub zaradczych, a kogo należy lub warto zaangażować w dane procesy.

Ostatnią grupą produktów są czynniki związane z **rozpowszechnianiem wyników** (ang. *evaluation dissemination*) – raportu i wniosków z badania. Wyróżnić należy dwa elementy – rolę ewaluatora i charakterystykę komunikacji.

Rola ewaluatora (ang. *evaluators role*) w procesie rozpowszechniania wyników może być aktywna – ewaluator proponuje kanały i sposoby prezentacji wyników, promuje badanie lub też pasywna – prezentuje wyniki tylko na wyraźne życzenie i według planu zlecającego.

Drugą kwestią jest charakterystyka komunikacji. Ważnymi czynnikami są:

- strategia rozpowszechniania wyników (ang. *dissemination strategy*) – istnienie lub brak sformalizowanej strategii (krótkiego dokumentu strategicznego, notatki, planu);
- wybrane grupy odbiorców (ang. *target audience*) – wąska grupa operatorów programu, głównie decydenci polityki (w tym KE), szeroka grupa interesariuszy (beneficjenci, aplikujący), szersza publika, inne ministerstwa i instytucje – odbiorcy europejscy;
- kanały i formy komunikacji (ang. *communication channels*) – są uzależnione od wybranych grup docelowych, np. rozesłanie raportu wewnętrznymi kanałami wymiany informacji (w ramach organizacji, programu – dokument wewnętrzny), prezentacja na wąskim bądź szerokim forum programu (np. na komitecie monitorującym lub konferencji i warsztatach z interesariuszami), krótka nota dla głównych decydentów, publikacja raportu na stronie internetowej, podsumowanie wniosków dla prasy;
- czas publikacji (ang. *publication timing*);
- nieformalny proces komunikacji (ang. *informal communication*) – czy wyniki są rozpowszechniane również w sposób nieformalny.

Segment 4 – UŻYCIE EWALUACJI (ang. *evaluation use & utilization*)

W tym segmencie zintegrowano kilka perspektyw spojrzenia na użycie ewaluacji. Po pierwsze, rozróżniono – głównie za Cousinsem (2004) oraz Christie i Alkin (1999) – między użyciem a jego brakiem, użyciem właściwym i niewłaściwym. Po drugie, wyróżniono perspektywę czasową – natychmiastowe użycie oraz perspektywę średnio i długookresową – wykorzystanie ewalu-

acji (za Kirkhart, 2000)[25]. Po trzecie, zaznaczono zakres merytoryczny użycia (rekomendacje, koncepcje, nawiązane kontakty) i rodzaj użytkowników (interesariusze programu vs. aktorzy zewnętrzni – osoby i instytucje spoza programu). Wreszcie uwzględniono podstawowe rodzaje użycia (instrumentalne vs. koncepcyjne). W rezultacie otrzymano cztery główne kategorie: pozytywne użycie ewaluacji w perspektywie krótkoterminowej, pozytywne wykorzystanie w perspektywie długoterminowej, neutralne użycie i wreszcie złe użycie. Są one omówione poniżej.

Pierwszym rodzajem jest **pozytywne użycie ewaluacji – perspektywa krótkoterminowa (ang. *evaluation use*)**. Jest to użycie produktów ewaluacji wyróżnionych w poprzednim segmencie modelu w odniesieniu do danego programu i przez aktorów bezpośrednio zaangażowanych w tenże program. W tej opcji istnieją dwie formy użycia: koncepcyjne i instrumentalne. Ważna jest ich kolejność – najpierw następuje użycie koncepcyjne, potem instrumentalne (Owen, 2007: 106). Najpierw użytkownicy ewaluacji przyswajają (internalizują) wiedzę zdobytą w procesie badania lub zawartą w raporcie, a następnie na tej podstawie działają.

Użycie koncepcyjne oznacza więc:

- zmiany w myśleniu odbiorców ewaluacji (ang. *internalizing findings*) – przyswojenie nowej wiedzy (głównie technicznej) zawartej w raporcie lub zdobytej podczas badania – wiedzy o programie, jego problemach, interakcjach (*know-what, how, when & who*);
- zdobycie świeżego, szerszego spojrzenia (ang. *gaining new perspective*) na program i jego kontekst – wiedzy o procesach zachodzących w otoczeniu *(know-about, know-why*);
- wiedza i doświadczenie z tematyki danego programu nabyte przez ewaluatorów, którzy wykonywali dane badanie.

Dowodami użycia instrumentalnego jest zaś wdrożenie rekomendacji lub praktyczna reakcja na wnioski wysuwane w raporcie. Konkretnie chodzi o:

- wprowadzenie zmian w logice i strategii programu – przesunięcia środków między celami, modyfikacja struktury celów, zmodyfikowanie analizy SWOT, włączenie nowych grup beneficjentów, etc.;
- zmiany w organizacji, procedurach, zasadach i procesach – przesunięcia kompetencji między instytucjami, zmiany w procedurach programu i sposobie zarządzania (np. rozszerzenie promocji programu, uproszczenie wniosków);
- zatrzymanie wdrażania programu – w praktyce jest to jednak mało prawdopodobne;
- wykorzystanie wniosków i rekomendacji raportu jako punktu odniesienia do następnego badania ewaluacyjnego tego samego programu (np. punktu porównań dokonanych postępów);
- utrzymanie kontaktów i komunikacji między interesariuszami programu, które zostały nawiązane w procesie ewaluacji;

[25] Tę odmienną perspektywę czasową dodatkowo uwypuklono zastosowaniem różnego nazewnictwa. Dla perspektywy krótkoterminowej przyjęto termin „użycia ewaluacji" (ang. *evaluation use*), zaś dla długoterminowej – „wykorzystania ewaluacji" (ang. *evaluation utilization*).

- podtrzymanie sieci kontaktów zewnętrznych (z instytucjami spoza programu, innymi programami), które zostały nawiązane podczas ewaluacji.

Drugim rodzajem jest **pozytywne wykorzystanie ewaluacji – perspektywa średnio i długoterminowa (ang. *evaluation utilization*)**. Jest to wykorzystanie produktów danej ewaluacji do innych interwencji przez aktorów nie zaangażowanych bezpośrednio w realizację zbadanego programu. W tej opcji również istnieją dwie formy użycia: instrumentalne i koncepcyjne.

Przejawami wykorzystania instrumentalnego są natomiast:

- wdrożenie danych rekomendacji z opóźnieniem, w kolejnej edycji programu (w sytuacji, gdy program jest kontynuowany bez większych zmian – to zdarza się jednak w praktyce programów UE rzadko);
- wykorzystanie raportu jako źródła zewnętrznych (dla zbadanego programu) analiz, badań, porównań;
- prezentowanie raportu lub programu na podstawie wniosków z raportu, jako dobrej praktyki (w ramach europejskich sieci tematycznych, forów współpracy);
- nowe kontrakty, o zbliżonej tematyce, zdobyte przez ewaluatorów na podstawie doświadczenia uzyskanego w poprzedniej ewaluacji.

Przejawami wykorzystania koncepcyjnego są natomiast:

- przytaczanie wniosków w debatach, dyskusjach na temat polityki, analogicznych interwencji;
- wykorzystanie wniosków z badania przy planowaniu logiki i strategii kolejnej generacji programów (w tym wypadku programów na lata 2007–2013);
- używanie raportu lub programu na podstawie opisów w raporcie, jako wzoru i punktu odniesienia (*benchmark*) dla innych interwencji/badań europejskich;
- zmiany w percepcji polityki, teorii analogicznych interwencji – w tym wypadku teorii współpracy terytorialnej.

Trzecim rodzajem jest **neutralne użycie ewaluacji (ang. *neutral use*)**. Za takie należy uznać:

- użycie symboliczne (ang. *symbolic use*) – wykonanie badania w celu otrzymania dodatkowego potwierdzenia podjętych już decyzji, swoistej merytorycznej podkładki;
- uzasadniony brak użycia (ang. *justified non-use*) – odbiorca nie używa raportu z powodu jego słabej jakości (brak poparcia w danych, wnioski bez mocnych dowodów, stojące w sprzeczności z powszechną wiedzą o programie i wynikami innych badań), braku użytecznych rekomendacji; uzasadniony brak użycia ma miejsce również wtedy, gdy dobry raport nie zostaje użyty, ponieważ kontekst programu zmienił się w sposób radykalny (wtedy rekomendacje, choć trafne, przestały być aktualne).

Ostatnim, czwartym rodzajem wykorzystania ewaluacji jest je **złe użycie (ang. *misuse of evaluation*)**. Głównym czynnikiem rozróżniającym możliwe opcje są intencje użytkownika – czy dopuszcza się złego użycia ewaluacji w sposób świadomy, czy nie:

- pozorne użycie (ang. *sham use*) zachodzi, gdy badanie (raport lub cały jego proces) jest potraktowane wyłącznie jako spełnienie formalnego wymogu

administracyjnego (w przypadku programów UE – spełnienie wymogów Komisji Europejskiej). Raport z badania zostaje całkowicie zignorowany, nie jest używany ani rozpowszechniany. Warto tu dodać, że cały proces ewaluacji może być od samego początku papierową rzeczywistością – zarówno zlecający, jak i ewaluator tworzą jedynie pozory badania i produkują raport wyłącznie na potrzeby Komisji;

- błędne użycie (ang. *mistaken use*) następuje, gdy użytkownik mając słabe rozeznanie w standardach nauk społecznych, przyjmuje i bezkrytycznie używa (w dobrej wierze) słabego raportu ewaluacyjnego;
- nadużycie – manipulacja (ang. *abuse, mischievous use*), czyli manipulowanie raportem i wynikami badania przez jego odbiorców, prezentowanie ich w sposób wybiórczy (wyjmowanie wniosków z kontekstu), przerysowywanie wniosków pozytywnych, a zatajanie krytycznych (ang. *spin*), wszystko po to, by przedstawić program i osoby zarządzające nim w sposób pozytywny;
- zatajenie (ang. *suppression*) – nieupublicznianie krytycznego raportu, zatajenie krytycznych wniosków i rekomendacji.

Segment 5 – ODDZIAŁYWANIE EWALUACJI (ang. *influence of evaluation*)

Logikę tego segmentu oparto na publikacjach Kirkharta (2000) oraz Marka i G. Henry'ego (2004) postulujących, by użyteczność wykorzystania ewaluacji rozpatrywać w szerszym kontekście jej oddziaływania na dobrobyt społeczny. Operacjonalizując tak szeroki koncept, osadzono go w kontekście zarządzania sektorem publicznym w UE i rozbito na konkretne cele – funkcje ewaluacji.

Opierając się na analizie literatury, w modelu uwzględniono pięć funkcji ewaluacji.

Funkcja 1. Poprawa planowania (ang. *Improving planning*):

- lepsze uzasadnienie interwencji publicznych (ang. *improving justification/legitimization*) – dostosowanie ich do potrzeb, pojawiających się problemów, sytuacji, oczekiwań adresatów pomocy;
- bardziej racjonalna i spójna alokacja zasobów publicznych odnosi się do spójnych strategii, logicznych celów programów (ang. *improving coherence & rationality of allocation*).

Funkcja 2. Poprawa wdrażania i kontrola jakości (ang. *Improving performance & quality*):

- lepsze wdrażanie interwencji publicznych (ang. *improving programme management*) obejmuje poprawę struktur, procedur, procesów wdrażania interwencji publicznych;
- minimalizowanie negatywnych efektów zarządzania opartego na wynikach (ang. *minimizing side-effects of performance management*) dotyczy zapobiegania pojawianiu się w praktyce sektora publicznego takich efektów ubocznych, jak zabijanie innowacji, koncentrowanie się tylko na technicznych celach lub efektach ilościowych, a ignorowanie efektów jakościowych (szerzej patrz: Tabela 4);

- lepsze efekty wdrażania (ang. *improving programme delivery*) odnosi się do zdobywania wiedzy i do weryfikacji postępów w realizacji planu, podnoszenia na bieżąco jakości dostarczanych produktów i usług.

Funkcja 3. Wzmocnienie odpowiedzialności (ang. *Improving accountability*):

- pokazanie szerszej opinii publicznej skuteczności interwencji publicznych – informowanie, na ile zakładane cele polityk są osiągane, jakie pojawiają się efekty prowadzonych interwencji publicznych;
- pokazanie szerszej opinii publicznej wydajności interwencji publicznych – przejrzyste informowanie o kosztach prowadzonych interwencji i o tym, na ile zasoby publiczne są racjonalnie wydatkowane.

Funkcja 4. Wsparcie procesów uczenia się

- wyjaśnienie czynników i procesów warunkujących działanie interwencji, jej sukces lub porażkę, w tym wyjaśnianie zaistniałych czynników ubocznych;
- wkład do teorii leżących u podstaw danego typu interwencji (ang. *testing assumptions*) – weryfikowanie teorii i założeń leżących u podstaw danych polityk;
- wiedza na temat dobrych praktyk i wzorców (ang. *building up know-how*) – identyfikowanie dobrych praktyk, promowanie wzorów i innowacyjnych, skutecznych podejść do interwencji publicznych;
- budowanie zdolności ewaluacyjnych (ang. *building capacity*) – testowanie i upowszechnianie technik i skutecznych podejść do badań, oceny i uczenia się z programów publicznych.

Funkcja 5. Wzmocnienie partnerstwa i poczucia współwłasności (ang. *improving partnership & ownership*):

- silniejsze, nowe powiązania, relacje, interakcje – rozwijanie współpracy między interesariuszami interwencji publicznych, pomoc w nawiązywaniu kontaktów, łączenie w sieci współpracy (ang. *networking & strengthening cooperation*);
- zaktywizowanie i uprawnianie interesariuszy – ułatwianie włączenia nowych głosów, nowych aktorów w dyskusję i proces decyzyjny polityk, interwencji (ang. *activation and empowerment*).

4.3.2. Zastosowanie modelu w studiach przypadków

Stworzony metamodel został zastosowany w studiach przypadku raportów z ewaluacji średniookresowych programów INTERREG III. Aby ułatwić analizę danych ze studiów przypadków, elementy modelu zostały przełożone na pytania weryfikujące daną kwestię. W opisie studiów przypadków pytania te były adresowane przy użyciu zróżnicowanych źródeł danych (wywiadów, przeglądu dokumentów, analizy tekstu, etc. – szerzej patrz Aneks na temat metodologii badania).

Model dostosowano również od strony zawartości merytorycznej do specyfiki ewaluacji programów współpracy terytorialnej. Główne dostosowania w odniesieniu do schematu 5 w Aneksie 6.2 omówiono poniżej.

Do studiów przypadków zostały wybrane raporty o wysokiej jakości (patrz Aneks na temat metodologii badania). Dokonując takiego wyboru wyłączono z analizy jakość raportu – zmienną najczęściej podawaną jako kluczową dla użycia ewaluacji (patrz: Balthasar, 2006; Cousins, Leithwood, 1986; EPEC, 2005; European Commission, 2004a; Morriset al., 1987; Owen, 2007; Rossi et al., 1999; Weiss, Bucuvalas, 1980a). Z analizy wyłączony był więc cały segment P2 (Raport ewaluacyjny).

Wymogi dotyczące celu i zakresu ewaluacji średniookresowych programów INTERREG były takie same dla wszystkich badań i zostały szczegółowo określone w Dokumencie Roboczym Komisji Europejskiej (porównaj rozdział 4.2.4 tabela 19). Dlatego też nie były one analizowane w modelu (w segmencie W1 wyłączono punkty 2 i 4, w W2 część punktu 1, a w D1 punkt 1). Ponadto temat wszystkich analizowanych raportów był ten sam – współpraca terytorialna (stąd wyłączony podpunkt w W1.5).

Definicja dobrego raportu obejmuje poprawne zebranie i analizę danych, poprawność oceny, a także użyteczność prezentowanej wiedzy. Ponieważ analizowane były tylko dobre raporty, więc nie było konieczne sprawdzanie w modelu raz jeszcze tych kwestii (wyłączenie w D2 triangulacji i punktów 3 i 4, w D3 punktu 2, w P3 punktów 4 i 5).

W segmentach dotyczących użycia ewaluacji wyłączono opcję zatrzymania programu (niemożliwa w praktyce programów funduszy strukturalnych).

Wreszcie w przypadku neutralnego lub złego użycia (segmenty U3 i U4), z racji tego, że analizowane były tylko dobre raporty, wyłączono używanie lub nieużywanie słabego badania. Nie uwzględniono też opcji zatajenia raportu, gdyż wszystkie raporty zostały upublicznione (miał do nich wolny dostęp między innymi autor niniejszej pracy).

4.4. STUDIA PRZYPADKÓW

Niniejszy rozdział zawiera opis wyboru, a następnie szczegółową analizę trzech studiów przypadków w oparciu o model analityczny skonstruowany w poprzednim podrozdziale. Każde studium przypadku zostało przeprowadzone i zaprezentowane według tej samej procedury – pięciu kroków analitycznych, odpowiadających pięciu segmentom testowanego modelu (w tekście zaznaczono kody danych segmentów modelu). Były to:

- Analiza wkładu w ewaluację (W1-warunki kontekstowe, W2 – wkład uczestników);
- Analiza procesu ewaluacji (D1-strukturalizacja, D2-obserwacja i analiza, D3-ocena);
- Analiza produktów ewaluacji (P1-stworzone kontakty, P2-wiedza zawarta w raporcie, P3-rozpowszechnienie wyników);
- Analiza użycia ewaluacji (U1-użycie ewaluacji U2-wykorzystanie ewaluacji, U3-neutralne lub U4-złe użycie ewaluacji);
- Wnioski ze studium przypadku odnoszące się do hipotez badawczych (hipotezy pomocnicze dotyczące wypełnienia funkcji ewaluacji oraz hipotezy dotyczące czynników wpływających na wykorzystanie badania).

4.4.1. Wybór studiów przypadków

Liczba programów INTERREG III poddanych w 2003 roku ewaluacji *mid-term* (średniookresowej) wyniosła siedemdziesiąt dwa (53 programy INTERREG IIIA, 13 INTERREG IIIB, 4 INTERREG IIIC, Program ESPON i INTERACT). Liczba raportów z tych badańwynosiła sześćdziesiąt dziewięć. Różnica wynikała z tego, że cztery programy INTERREG IIIC były przebadane w ramach jednego, połączonego badania ewaluacyjnego (jego produktem był jeden raport).

Do analizy z zastosowaniem zbudowanego metamodelu zdecydowano się wybrać sześć raportów (trzy badania) o wysokiej jakości. Skoncentrowano się na raportach na wysokim poziomie, gdyż to jakość raportu jest warunkiem wyjściowym jego pozytywnego użycia w zarządzaniu sektorem publicznym. Termin „raportu o wysokiej jakości" zdefiniowano w oparciu o kryteria przyjęte w odniesieniu do oceny ewaluacji programów UE. Raport o wysokiej jakości to taki, który spełnił następujące kryteria:

- odpowiadał zakresowi złożonego przez zleceniodawcę zamówienia;
- miał poprawny plan badawczy (trafność wewnętrzna, trafność teoretyczna, uwzględnienie specyfiki programu);
- opierał się na poprawnym doborze i użyciu danych (rzetelność badania i doboru próby, triangulacja);
- zawierał przejrzystą i poprawną analizę (jasny proces dedukcji/indukcji, trafność konkluzji statystycznych);
- przedstawiał wiarygodne wyniki (takie, które mają poparcie w zebranych danych i dokonanych obserwacjach);
- przedstawiał jasne, konkretne i bezstronne rekomendacje;
- był napisany w sposób przystępny i przejrzysty (tzw. przyjazność dla użytkownika, komunikatywność formy prezentacji).

Wybór konkretnych raportów – studiów przypadku według powyższych kryteriów – oparto na wynikach wcześniejszego badania – metaewaluacji badań średniookresowych programów INTERREG III (Bachtler et al., 2005)[26]. Ogólnym celem tego studium porównawczego był bilans doświadczeń i postępów we wdrażaniu programów współpracy terytorialnej. Jako szczegółowe pola badawcze zlecający – Sekretariat INTERACT podał: analizę orientacji tematycznych programów, ocenę efektywności programów, analizę struktur zarządzania i procesów wdrażania, ocenę systemów monitoringu oraz identyfikację potencjalnych pól transferu doświadczeń między aktorami INTERREG. Dodatkowo wykonawca – European Policies Reseach Centre (University of Strathclyde) – zaproponował porównawczą analizę jakości raportów ewaluacyjnych (według powyższych siedmiu kryteriów) oraz metodologii użytych do ewaluacji programów INTERREG.

---

[26] Autor niniejszej pracy był członkiem zespołu badawczego European Policies Research Centre, University of Strathclyde. Otrzymał również stosowane oficjalne pozwolenie od Sekretariatu INTERACT (instytucji zamawiającej badanie) i EPRC na wykorzystanie danych źródłowych i wyników cząstkowych projektu, w tym bazy danych raportów ewaluacji średniookresowych.

Studium oparte było wyłącznie na syntezie raportów – a więc na danych zawartych w ewaluacjach średniookresowych. Należy zwrócić uwagę, że przy takim podejściu jakość raportów determinowała wizerunek programów. Główną metodą była analiza treści raportów ewaluacyjnych, analiza literatury na temat współpracy terytorialnej w Europie i wcześniejszych ewaluacji programów INTERREG (ewaluacji *ex post* INTERREG II), wywiady z pracownikami Komisji Europejskiej zajmującymi się tą Inicjatywą Wspólnotową. Raporty z ewaluacji miały analogiczną strukturę rozdziałów i wniosków, ale były obszernymi dokumentami (przeciętnie liczyły 100 stron) i były publikowane w różnych językach i formatach. Konieczne więc było ich zestawienie. Narzędziem technicznym, umożliwiającym syntezę, była baza danych, w której zespół recenzentów wpisywał podsumowania raportów według ujednoliconej struktury pytań. Baza składała się z czterech komponentów tematycznych: informacji ogólnych o programie, metodach użytych w ewaluacji, jakości ewaluacji oraz najbardziej obszernego komponentu – *meritum* raportu ewaluacyjnego (z segmentami na temat orientacji tematycznej i aktualności strategii, efektywności programu, struktury zarządzania, monitoringu, omówienia projektów strategicznych).

Główne wnioski z badania pogrupowane według wymienionych już pól analizy można streścić następująco (Bachtler et al., 2005).

Analiza orientacji tematycznych programów przyniosła spostrzeżenie na temat podobieństw między niektórymi programami ścieżki A, a programami IIIB. Programy IIIA o długiej tradycji, operujące w ramach „starych" granic UE, koncentrowały się na wspieraniu wspólnych projektów „miękkich" i budowaniu wspólnych strategii (podobnie jak programy IIIB). Programy IIIA funkcjonujące w ramach nowych granic preferowały projekty przygraniczne, dotyczące rozwoju infrastruktury, zwykle prowadzone przez partnerów po dwóch stronach granicy niezależnie, albo przy organicznej koordynacji (szerzej typologia projektów w ramach różnych programów była omawiana w rozdziale poświęconym INTERREG).

Ocena efektywności programów udowodniła, że tylko w 10% z nich wydawanie pieniędzy idzie zgodnie z planem. Tak więc *gros* programów miała poważne opóźnienie w zagospodarowywaniu środków UE, a co za tym idzie była narażona na regułę N + 2 i utratę pomocy UE. Wynikało to z ich bardzo późnego startu, ale również z problemów w koordynacji i we wspólnym wypracowaniu ponadnarodowych procedur.

Analiza struktur zarządzania pokazała natomiast zaskakującą sytuację. Mimo standardowego modelu narzuconego dla wszystkich programów (Sekretariat Techniczny, Organ Zarządzający, Płatniczy, Komitet Monitorujący, Sterujący, instytucje pomocnicze), wymuszającego uniformizację wśród programów, istniało wyjątkowe zróżnicowanie i mnogość układów instytucjonalnych. Istniały rozwiązania wysoce ponadgraniczne i partnerskie (np. w przypadku programu Hiszpania-Francja – instytucja płatnicza we Francji, zarządzająca w Hiszpanii, w programie Euroregio Moza-Ren jeden sekretariat programu), ale i rozwiązania mało zintegrowane (głównie programy na zewnętrznych granicach UE, czy programy południa UE z sekretariatami – lustrami po obu stronach granicy). Szczególnie szerokie było spektrum instytucji pomocniczych

(oficerowie kontaktowi, punkty pierwszego kontaktu, platformy współpracy i instytucje promocyjne). Analiza ujawniała, że stopień integracji struktur zarządzania zależy w dużej mierze od historii współpracy na danej granicy (istnienia dynamicznego euroregionu).

W przypadku procesów wdrażania słaby stan zaawansowania programów ograniczył analizy ewaluacji średniookresowych (a co za tym idzie i ich syntezy) do kwestii promocji, wyboru projektów i tworzenia baz monitoringu. Analiza pokazała, że sukces odnosiły te programy, które w ramach promocji wyodrębniły specjalną strategię wspierania pomysłów na projekty transgraniczne (bazy pomysłów, oficerowie kontaktowi w terenie, pomoc w wypełnianiu wniosków od strony technicznej, listy rezerwowe projektów). Przy wyborze projektów wyraźna była przede wszystkim grupa programów skandynawskich, o jasnych – merytorycznych i technokratycznych – kryteriach wyboru projektów. W wielu programach proces wyboru projektów okazał się jednak nieprzejrzysty i mocno polityczny. W polu monitoringu nierozwiązanym problemem były wskaźniki i uchwycenie w nich trasngraniczoności projektów, miękkiego charakteru ich efektów. Temat ten był rozwijany w kolejnym studium zleconym przez INTERACT (Fiorini et al., 2006).

Wreszcie w polu transferu wiedzy i doświadczeń między aktorami INTERREG, raport EPRC zidentyfikował szeroką listę tematów (od doświadczeń w efektywnych strukturach zarządzania, poprzez rozwiązania z zakresu promocji programów i budowania pomysłów na projekty, po kwestie finansowe i rozwiązania dotyczące baz danych z monitoringu rzeczowego programów). Ciekawym elementem studium było zidentyfikowanie przesłanek dla wymiany doświadczeń i zaproponowanie strategii uczenia się dostosowanej do poziomu programów. Stwierdzono, że operatorzy programów mogą się kierować trzema przesłankami: chęcią rozwiązania bieżących problemów, uniknięciem potencjalnych problemów oraz szukaniem szerszych, długofalowych inspiracji i wizji. W oparciu o to stwierdzono, że istnieją trzy opcje uczenia się:

- wymiana między „rówieśnikami" (ang. *Peer-to-peer exchange*) – dzielenie się wiedzą i doświadczeniem między programami o podobnej charakterystyce i na podobnym poziomie rozwoju, tzn. o podobnym poziomie zaawansowania wydatków, położeniu w tej samej strefie kulturowej, podobnym poziomie integracji obszaru przygranicznego (np. programy Polska–Niemcy i Czechy–Austria);
- konsultacje z poprzednikami (ang. *Consultation with predecessors*) – czyli wymiana doświadczeń z programami bardziej zaawansowanymi, które przełamały już podobne problemy w przeszłości, ale dzielą z danym programem analogiczną ścieżkę rozwoju (np. Programy CBC Polska-Niemcy i program sąsiedzki Polska–Ukraina)
- inspiracje od liderów (ang. *Inspiration from the leaders*) – obserwacja rozwiązań najbardziej zintegrowanych i rozwiniętych programów, czerpanie od nich inspiracji na temat kierunków rozwoju (np. program Niemcy–Holandia–Francja i program Niemcy–Polska).

Użyteczność tych metod różni się w zależności od potrzeb, którymi kierują się operatorzy programu, co można podsumować w formie tabeli.

Tabela 20. Przesłanki i rodzaje transferu wiedzy między instytucjami programów UE

		Przesłanki do uczenia się		
		rozwiązywanie problemów	unikanie błędów	szukanie inspiracji
Rodzaje transferu wiedzy	Wymiana między rówieśnikami	**	**	*
	Konsultacje z poprzednikami	***	***	**
	Inspiracje od liderów	*	*	***

* mało przydatne, ** użyteczne, *** wysoce użyteczne.
Źródło: (Bachtler et al., 2005).

To podejście pomogło Sekretariatowi programu INTERACT lepiej ukierunkować wymianę doświadczeń i łączyć aktorów INTERREG stosownie do ich potrzeb informacyjnych.

Kluczowe z perspektywy niniejszej pracy były wyniki analizy EPRC w polu jakości i metodologii raportów ewaluacyjnych. Dwa główne wnioski były następujące. Po pierwsze, ewaluacje w niewielkim stopniu starały się stworzyć specjalne techniki i metody dopasowane do specyfiki programów współpracy terytorialnej (ich nastawienie zarówno na miękkie, jak i na twarde efekty, nakierowanie na efekty rozwojowe i efekty łączenia, interakcje między aktorami i rolę różnic międzykulturowych). Ewaluacje opierały się w dużej mierze na standardowym warsztacie głównego nurtu funduszy strukturalnych. Drugim spostrzeżeniem było, iż programy słabo zaawansowane, o poważnych, jasno widocznych problemach miały słabe raporty, wypełnione ogólnymi wnioskami i mało konstruktywnymi rekomendacjami. Paradoksalnie więc, tam gdzie była pilna potrzebna naprawy stanu rzeczy, badanie ewaluacyjne dawało bezużyteczny produkt. Był to przypadek przede wszystkim programów z południa UE. Wyraźnie wyłaniała się też grupa dobrych raportów – o jasnej formie i zwięzłych rekomendacjach. Badanie EPRC nie pozwalało jednak prześledzić, jaka może być rzeczywista rola tych potencjalnie użytecznych raportów w naprawianiu problemowych sytuacji (praktycznie rzecz biorąc wszystkie programy miały problemy). Prezentowana w niniejszej książce analiza była więc swoistą kontynuacją badania EPRC. Punktem wyjścia do tej analizy były informacje z dwóch części bazy EPRC, dotyczące jakości raportów ewaluacji średniookresowych:

- „Metody użyte w ewaluacji" (ang. *Methods used*);
- „Jakość ewaluacji" (ang. *Quality of evaluation*).

W przypadku pierwszego z tych segmentów zawartość bazy miała charakter informacyjny – opisowy, często statystyczny. Recenzenci zaznaczali, które z możliwych i dostępnych metod były wykorzystane podczas danego badania ewaluacyjnego, w jakiej formie i skali. Drugi komponent – jakość ewaluacji – miał charakter jakościowy i oceniający. Recenzenci oceniali w formie krótkich opisów jakość danego raportu. Szablon oceny został skonstruowany na podstawie szablonów Komisji Europejskiej i zawierał siedem omawianych już kryteriów oceny jakości raportu.

Po analizie zawartości obydwu komponentów bazy, wniosków z wywiadów z urzędnikami Komisji Europejskiej, jak również uwzględnieniu opinii członków zespołu badawczego EPRC, zidentyfikowano dziesięć raportów o najwyższej jakości. Wybrane przypadki zweryfikowano następnie pod względem dostępności dokumentów (możliwość prowadzenia wywiadów w języku angielskim i dostępność językowa dokumentów programowych) oraz możliwości nawiązania kontaktu z kluczowymi aktorami procesu ewaluacji. Ostatecznie odrzucono cztery raporty, a do studiów przypadku wybrano sześć (patrz poniższa tabela).

W praktyce przeprowadzono trzy studia przypadków, ponieważ badania ewaluacyjne dla programów Austria–Słowenia, Austria–Słowacja, Austria–Czechy i Austria–Węgry okazały się realizowane i przyjmowane przez ten sam zespół, a źródła danych były wspólne.

Tabela 21. Raporty ewaluacyjne wybrane do badania

Tytuł raportu	Komentarz
INTERREG IIIA – Austria – Slovenia	Raport wybrany do studium przypadku: przypadek 1
INTERREG IIIA – Austria – Czech Republic	Raport wybrany do studium przypadku: przypadek 1
INTERREG IIIA – Austria – Slovakia	Raport wybrany do studium przypadku: przypadek 1
INTERREG IIIA – Austria – Hungary	Raport wybrany do studium przypadku: przypadek 1
INTERREG IIIA – Ireland – Northern Ireland	Raport wybrany do studium przypadku: przypadek 2
INTERREG IIIB – North-West Europe	Raport wybrany do studium przypadku: przypadek 3
INTERREG IIIA – Ireland – Wales	Raport wyłączony z badania po wstępnej analizie i wywiadach z częścią interesariuszy. Struktura zarządzania programem umożliwiła dotarcie tylko do irlandzkiej strony – tak wykonawców, jak i operatorów programu. Tymczasem główne ośrodki decyzyjne i koordynacja konsorcjum badawczego okazały się leżeć po stronie walijskiej. Tam też nie było ciągłości personalnej. Materiał zebrany po stronie irlandzkiej miał więc ograniczoną wartość analityczną.
INTERREG IIIB – Baltic Sea Region	Raport wyłączony z badania z powodu słabej dostępności geograficznej oraz problemów w nawiązaniu kontaktów z głównymi aktorami (brak ciągłości personalnej).
INTERREG IIIA – FI/SE/NO/RU – Nord	Raport wyłączony z badania z powodu ograniczeń językowych (dokumentacja programu w języku szwedzkim i fińskim) oraz dużej, zróżnicowanej językowo grupy kluczowych aktorów
INTERREG IIIA – ES/FR – Pyrénées	Raport wyłączony z badania z powodu kwestii językowych (wszystkie dokumenty programu tylko po hiszpańsku i francusku)

Źródło: opracowanie własne.

4.4.2. Przypadek 1: IIIA–Austria – Czechy, Słowacja, Słowenia, Węgry

Warunki kontekstowe [segment W1 modelu]

Austria uczestniczyła aż w sześciu programach współpracy ponadgranicznej (patrz Aneks – Mapa 1). Program INTERREG był więc dla Austrii ważny, gdyż stanowił narzędzie budowania stosunków dobrosąsiedzkich. W latach 2000–2003 sytuacja była o tyle wyjątkowa, że cztery z tych programów funkcjonowały na granicach zewnętrznych – z państwami kandydującymi do UE. W 2004 roku regiony sąsiedzkie z Czech, Słowacji, Słowenii, Węgier (dotychczas finansowane z programu PHARE-CBC) miały zostać włączone w trwający już (od 2000 roku) program INTERREG. Kwestia płynnego włączenia NMS – nowych krajów członkowskich (ang. *New Member States*) – w działający już system była dla Austrii na tyle wysokim priorytetem, że koordynację nad całym procesem rozszerzania INTERREG-ów przejął rząd federalny.

Analizując warunki kontekstowe, należy zwrócić uwagę na fakt, że we wszystkich czterech omawianych tu programach (Austria–Czechy, Austria–Słowacja, Austria–Słowenia, Austria–Węgry) następowało zderzenie kultury administracyjnej Europy Zachodniej (nowoczesnego zarządzania opartego na wynikach) z administracjami postkomunistycznymi – a więc kierującymi się inną logiką. Mimo długiego procesu dostosowań do wymogów UE, wciąż zachowały one wiele z klasycznej administracji publicznej (sztywne procedury, hierarchiczność procesu decyzyjnego, niska elastyczność i innowacyjność). Brak kultury ewaluacyjnej w tych państwach był jedynie jednym z pomniejszych elementów różnic między kulturami administracji. Dodatkowym utrudnieniem – związanym również z kulturą administracji – był fakt, że w przypadku zarządzania programami europejskimi biurokracje krajów kandydujących były przyzwyczajane przez ostatnie 15 lat do procedur zarządzania DG-External Aid (charakteryzującej się, w dużym uproszczeniu, ścisłymi regułami i dużym stopniem centralizacji – na poziomie Komisji), podczas gdy Austria i wdrażany w niej INTERREG był nadzorowany przez Dyrekcję Generalną-Regio (dającą dużo większą niezależność i samodzielność administracjom krajowym). Sytuację pogarszała bardzo ograniczona koordynacja między tymi dwoma dyrekcjami.

Wszystkie analizowane programy austriackie były średniozaawansowane[27]. Paradoksalnie okazało się to problemem, gdyż w ten zaawansowany system i w ponad połowie przypisane pieniądze (a więc w trwające projekty) należało włączyć nowych partnerów.

Struktura organizacyjna była standardowa dla wszystkich programów i zgodna z rozporządzeniami UE. Instytucją Zarządzającą (IZ) był Federalny Urząd Kanclerski (*Federal Chancellery*) w Austrii. Działały też Wspólne Sekretariaty Techniczne (WST) wspomagające na co dzień działania IZ (obsługiwane przez austriacką instytucję ÖIR – Österreichisches Institut für

[27] Dla AT-CZ rozdysponowano 50% środków, dla AT-SI 57%, dla AT-SK 41%, a dla AT-HU 34,6%. Rzeczywiste wydatki wynosiły odpowiednio: 13,4%, 8,4%, 6,8%, 7,9% (źródło: dane z programów i baza EPRC).

Raumplanung), Wspólne Komitety Monitorujące (WKM) włączające wszystkie instytucje zaangażowane w programy i Wspólne Komitety Sterujące (WKS). Bardzo silną pozycję miały w tych instytucjach poszczególne austriackie władze regionalne – na co dzień zajmujące się wdrażaniem poszczególnych programów (instytucje pośredniczące). Wszystkie te instytucje były ulokowane w Austrii i odnosiły się tylko do strony austriackiej. Kraje sąsiedzkie, będąc poza UE, nie działały w ramach INTERREG-u, tylko programu PHARE–CBC, który miał oddzielne struktury. Austriacy starali się integrować te różne systemy instytucjonalne poprzez włączanie w prace Komitetów i Sekretariatu przedstawicieli instytucji zarządzających PHARE–CBC w Czechach, Słowacji, Słowenii i Węgier. Wyzwaniem dla planowanego rozszerzenia programu INTERREG było przejście od lustrzanych struktur narodowych na rzecz jednego systemu, wspólnego dla dwóch stron granicy.

Struktura celów analizowanych programów była zbliżona[28]. Należy jednak zwrócić uwagę, że w momencie podejmowania ewaluacji średniookresowych były to cele tylko dla regionów austriackich. Partnerzy z państw kandydackich (Czech, Węgier, Słowacji, Słowenii) mieli swoje strategie i własne, dużo skromniejsze źródło finansowania (PHARE–CBC). Strategie PHARE–CBC były zwykle nakierowane na inwestycje infrastrukturalne, gdyż słaba infrastruktura stanowiła główną bolączkę regionów przygranicznych państw kandydackich. Oczywiście koordynowano wizje i strategie (INTERREG vs. PHARE–CBC) prowadząc wspólne rozmowy i konsultując kształt strategii. Niemniej wciąż były to różne systemy.

Podsumowując kwestie kontekstowe badanych ewaluacji, można stwierdzić, że nie istniał konflikt wokół żadnego z programów. Natomiast presja reform była niewątpliwa – stanowiła po prostu wymóg formalny, bo po rozszerzeniu UE zasady i pieniądze INTERREG automatycznie zaczynały obejmować nowe kraje członkowskie.

Wkład uczestników [segment W2 modelu]

Oczekiwania zlecającego – Federalnego Urzędu Kanclerskiego – były bardzo jasne. Po pierwsze, ewaluacja miała pomóc operatorom programu w podniesieniu ich efektywności („use if for programmer, to improve their efficiency"). Po drugie, oczekiwano, że pomoże w stosunkowo płynnym procesie integrowania nowych krajów w działające już programy INTERREG. Za tym szły konkretne oczekiwania co do formy i podejścia badawczego. Urząd Kanclerski dążył do tego, by był to proces interaktywny, by obserwacje ewaluatorów były na bieżąco używane w programie, by mieć ciągłość śledzenia rekomendacjami, a wreszcie by rozbudzić tym badaniem zainteresowanie partnerów z NMS, pomóc im włączyć się w bieżącą dyskusję.

Zlecający zdecydował się na ewaluację *on-going* – w tym wypadku najtrafniej tłumaczoną jako „ewaluacja ciągła", a nie ewaluacja bieżąca, gdyż nie

[28] Pierwszym priorytetem we wszystkich czterech programach był rozwój gospodarczy, drugim pod względem wielkości finansowania – rozwijanie infrastruktury, a trzecim zasoby ludzkie (Öir-Managementdisenste Gmbh, 2001a, b, c, d, 2002, 2004a, b).

była badaniem punktowym, lecz objęła cały kilkuletni pakiet badań. Kontrakt na badanie ewaluacyjne zawierał więc wykonanie ewaluacji średniookresowej, bieżącą asystę instytucji programowych w kolejnych trzy latach oraz uaktualnienie badania średniookresowego (w sumie trzy i pół roku badań i działań). W ToR umieszczono zapis, by w skład zespołu ewaluacyjnego wchodzili eksperci z nowych krajów członkowskich.

Taka strategia zlecającego wynikała z kilku kwestii. Po pierwsze, miał on negatywne doświadczenia z poprzedniego okresu programowania odnośnie tradycyjnej, punktowej ewaluacji średniookresowej – ówczesne badania okazały się, z perspektywy operatorów programu, bezużyteczne. Były tylko dodatkowym narzędziem kontroli dla Komisji. Z drugiej strony zlecający miał już pozytywne doświadczenia płynące z zastosowania interaktywnej, partnerskiej ewaluacji ciągłej[29]. Wreszcie, zlecający był świadomy, że z powodu opóźnień programu ewaluacja średniookresowa będzie powtórzona za dwa lata. Doszedł więc do wniosku, że skoro i tak będzie musiał zaangażować środki w spełnienie tego wymogu, to warto użyć go do własnych celów.

W instytucji zlecającej (FC) był wyraźny lider i promotor pomysłu – jeden z urzędników wysokiej rangi, który przez lata był koordynatorem wszystkich programów INTERREG, a w badanym okresie silnie zaangażował się (na poziomie europejskim i samej Austrii) w prace nad nowym kształtem polityki współpracy terytorialnej po 2007 roku. Drugą zaangażowaną instytucją był JTS. Na potrzeby ewaluacji stworzono dwie grupy. Pierwszą była Grupa Sterująca Ewaluacją – składały się na nią trzy lub cztery osoby z kluczowych instytucji. Drugą, szerszą grupą była Międzyprogramowa Grupa Sterująca (*Cross-Programme Steering Group – CPSG*). W jej skład wchodzili przedstawiciele instytucji zarządzających i głównych interesariuszy ze wszystkich czterech programów (w tym przedstawiciele Nowych Krajów Członkowskich). Grupa ta miała być platformą wymiany poglądów, poszukiwania wspólnych cech i rozwiązań dla wszystkich czterech programów.

Wykonawcą kontraktu został austriacki instytut studiów regionalnych. Zespół ewaluatorów składał się z ośmiu do dziesięciu osób (po 2–3 osoby do każdego programu). Prowadzącym dla każdego z badań był ewaluator austriacki, ale współpracował z nim ewaluator z danego nowego państwa członkowskiego. Kierownikiem zespołu i głównym koordynatorem był ewaluator austriacki z wieloletnim doświadczeniem.

Należy podkreślić tu zupełnie odmienną filozofię kontraktu i badania wynikającą z różnicy między ewaluacją średniookresową a bieżącą. Pierwszą jest kwestia różnych funkcji – kontroli vs. uczenia się. Opisując to słowami ewaluatora, „w ewaluacji średniookresowej zarządzający mówią nam: znajdźcie i oceńcie, jak sprawy się mają. Wiedzą, że to rodzaj kontroli, że całość czytać będzie Komisja i starają się wszystko przedstawić w dobrych barwach. Przy ewaluacji bieżącej my mówimy – nie marnujmy czasu, powiedźcie, jakie są wasze problemy, czego chcecie się dowiedzieć i od razu będziemy nad tym

[29] Podejście „ciągłe" zastosowano z sukcesem przy ewaluacjach celu 2 w latach 1995–1999, zaś podejście interaktywne w ewaluacji *ex ante* INTERREG-ów austriackich.

pracować. Pracujemy z nimi, ale Komisja nie ma dostępu do informacji roboczych – zobaczy tylko rezultat końcowy. To musi pozostać wewnętrzne, gdyż inaczej nikt nie odważy się szczerze rozmawiać". Drugą różnicą jest strategia badania. Zamiast jednego, punktowego badania dotyczącego dużych kwestii, bada się wiele drobnych pytań ważnych dla zlecających i przez nich wysuwanych. Założeniem jest, że użytkownikom łatwiej jest na bieżąco przyswajać partie informacji, niż cały raport jednocześnie. Wreszcie trzecia różnica to czas. Badanie średniookresowe prowadzone było według sztywnego kalendarza (równo w połowie programu). Tymczasem półmetek programu nie zawsze był kluczowym momentem. Ewaluacja bieżąca w tym wypadku była ciągła, łatwiej więc było dostosować badanie do tempa programu (tu najważniejszym momentem było rozszerzenie UE).

W używanym do analizy modelu trudno jest umieścić dwie kwestie zidentyfikowane podczas badania. Pierwszą jest różnica w hierarchii priorytetów Austrii i krajów kandydackich, jak również wynikające z tego ich niskie zaangażowanie w proces ewaluacji (wbrew oczekiwaniom Austrii). Dla nowych krajów program INTERREG był w tamtym okresie kwestią drugoplanową, gdyż przygotowując się do członkostwa koncentrowały się na dużo bardziej znaczących dla nich sprawach – głównych funduszy strukturalnych czy polityki rolnej.

Drugą kwestią był swoisty lobbing prowadzony przez głównego ewaluatora. Od początku lat 90. rozwijał on zarówno naukowo, jak i w praktyce (w kontraktach dla rządu austriackiego) podejście partycypatywne. Jego głównymi inspiracjami było myślenie systemowe – w tym systemowe podejście do procesów uczenia się i poprawy zarządzania programami (Hummelbrunner, 2000) oraz podejście skoncentrowane na użyciu badania (*utilization-focused approach* M. Pattona). Do swoich pomysłów przekonał osoby z Federalnego Urzędu Kanclerskiego. Pomysł został przetestowany, a po tym jak sprawdził się w ewaluacjach programów celu 2 oraz w badaniu Inicjatywy URBAN, zyskał zwolenników w administracji austriackiej. Tę sytuację można określić jako lobbing bądź też uprzedni proces edukacyjny administracji (przygotowanie gruntu pod badanie).

Analiza procesu ewaluacji [segmenty D1, D2, D3]

Z racji obranej przez zlecającego formy kontraktu (ewaluacja ciągła) trudno jest omawiać wyłącznie ewaluację średniookresową i jej efekty. Była ona jedynie jednym z elementów (trzymiesięcznym badaniem) całego, trzyipółrocznego procesu złożonego z szeregu badań i konsultacji oraz raportów roboczych. Kalendarium tej ewaluacji w kontekście całego planowania i zarządzania programami INTERREG dla Austrii przedstawiono na schemacie 8. Trudno jest wydzielić z tego procesu tylko raport średniookresowy, zresztą byłoby to podejście zawężające i zniekształcające obraz. W związku z tym w dalszych krokach pokazana została ewaluacja ciągła, zaś element średniookresowy został wyeksponowany tam, gdzie jest to uzasadnione.

Z uwagi na długi czas badania i jego interaktywność trudno było w sposób jednoznaczny wydzielić fazy strukturalizacji, obserwacji i analizy, oceny. Można było jednak zidentyfikować kilka etapów całego procesu na dwóch płaszczyznach. Pierwszą płaszczyzną była orientacja merytoryczna prac. I w tym wypadku można zauważyć jak prace przesuwały się od identyfikowania problemów i wspólnych kwestii, poprzez wykonanie badania średniookresowego na potrzeby Komisji (analiza procesów na poziomie programu), etap integrowania aktorów z nowych krajów (przygotowanie do rozszerzenia), łączenie procedur, analizę problemów na poziomie projektów, i wreszcie w trakcie oraz po uaktualnieniu badania – wsparcie prac nad nowymi dokumentami programowymi. Natomiast na płaszczyźnie organizacji procesu badawczego wyróżniały się trzy etapy (Hummelbrunner, 2004). Pierwszym było wypracowanie pytań, wspólnych definicji i kontaktów. W drugim etapie prowadzono analizy dla każdego z czterech programów. Trzecim etapem były studia przypadku projektów. Dalsza analiza będzie przebiegać właśnie zgodnie z tymi trzema fazami.

Celem **etapu pierwszego** prac ewaluatorów było zidentyfikowanie obszarów zainteresowań do dalszych badań i wypracowanie wśród głównych interesariuszy czterech różnych programów wspólnego rozumienia i wspólnych definicji. Etap ten miał więc wyraźne charakter fazy strukturalizacji.

Głównym narzędziem były spotkania ewaluatorów z interesariuszami na forum Grupy Międzyprogramowej (GM). Podczas dyskusji animowanych przez ewaluatorów wypracowano listę kluczowych kwestii, które miały być następnie badane. Była to wspólna lista dla wszystkich czterech programów, co dawało możliwość przyszłych porównań. Ewaluatorzy nie tylko animowali dyskusję, ale także sugerowali nowe pytania (wychodzące poza perspektywę programu, na której koncentrowali się uczestnicy spotkań).

Jednocześnie na tym etapie przygotowywany był raport z ewaluacji średniookresowej. Z uwagi na krótki czas i inaczej postawione priorytety, tak ewaluatorom, jak i zlecającym chodziło o jak najszybsze wykonanie raportu (głównie pod kątem Komisji Europejskiej). Ewaluatorzy wykorzystali ewaluację średniookresową do wsparcia procesu identyfikowania problemów i wyzwań. Zastosowanym narzędziem były warsztaty bilateralne z udziałem IZ, KM i instytucji pośredniczących (w praktyce wdrażających) każdego z czterech programów. Zrobiono też jeden wspólny warsztat przedstawicieli czterech programów, podczas którego ewaluatorzy zbierali pomysły od interesariuszy (w tym pomysły wstępnych rekomendacji). Pomysły rekomendacji zweryfikowano w wywiadach z IZ. Wnioski z badania zasiliły dyskusję na forum Grupy Międzyprogramowej na temat listy kluczowych kwestii badawczych.

Od samego początku wybrano podejście partycypatywne, zorientowane na użytkowników – operatorów programu, a co z tego wynika zorientowane na zarządzanie (poprawę procedur i procesów, wyjaśnienie problemów, efekty projektów). Stosowane techniki były wysoce interaktywne (warsztaty, dyskusje; same warsztaty dla badania miały jednak charakter dość formalny). Za interaktywne należy też uznać relacje w tej fazie między ewaluatorem a zlecającym. Warto podkreślić, że w proces włączona została szersza rzesza intere-

sariuszy, choć zainteresowanie i udział niektórych regionów był ograniczony[30]. Specyfiką tej ewaluacji było również to, że rozmowy we wszystkich czterech programach były jednoczesne i wspólne. Dawało to możliwość dodatkowych interakcji, ale też odkrywania synergii między czterema programami.

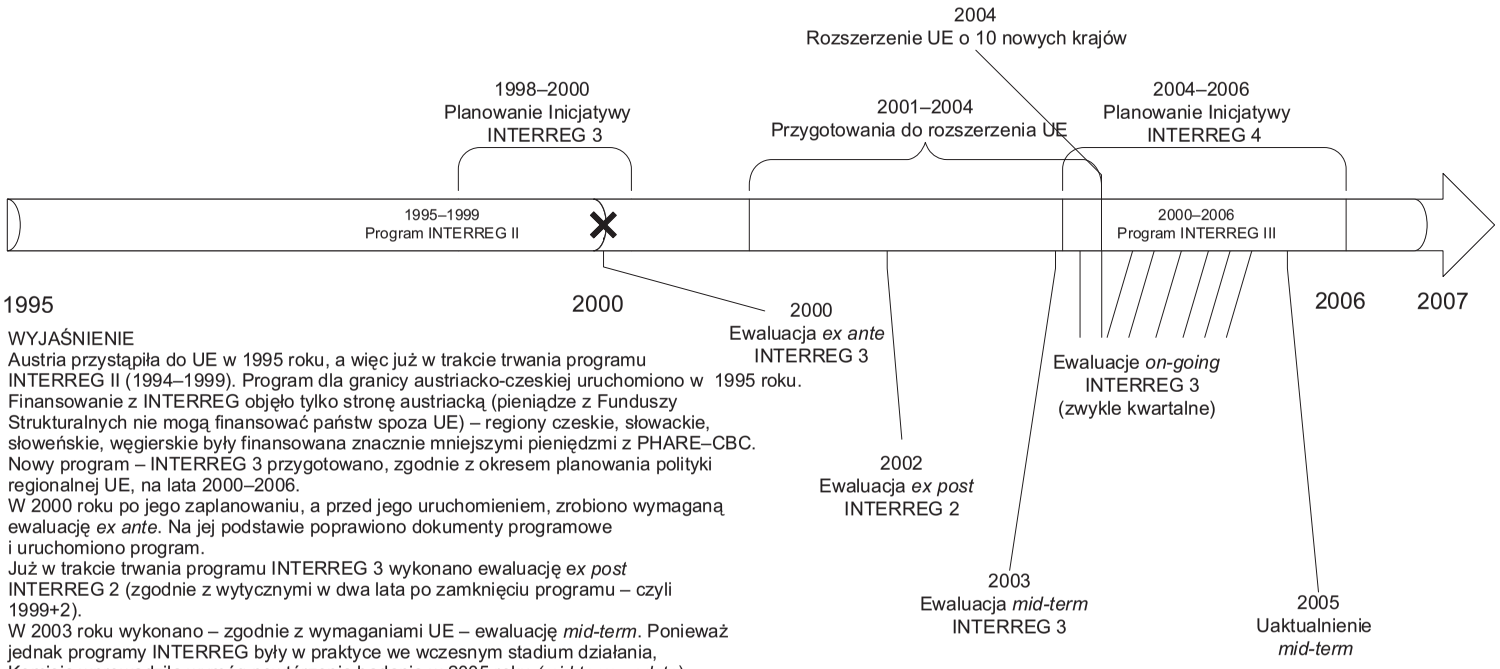
W etapie drugim pytania z etapu pierwszego zostały szczegółowo zbadane dla każdego z programów. Ewaluatorzy odgrywali głównie rolę animatorów dyskusji – organizowali i prowadzili spotkania ułatwiające dialog między partnerami programu. Szczególny akcent był kładziony na włączanie uczestników z nowych krajów członkowskich. Prowadzono na przykład kwestie wskaźników i monitoringu, analizę przystawalności projektów, analizy szybkości procedur (od ogłoszenia konkursu na projekty, poprzez aplikacje, wybór, pierwsze wypłaty).

Prezentowanie wyników analiz pod koniec prac było ciekawą, wspólną (ewaluatorów i operatorów programów) formą oceny. Ponieważ analiza dla wszystkich czterech programów była prowadzona w ten sam sposób, według tych samych pytań, dawało to możliwość porównań. Wyniki badań danych programów pokazywano na tle innych, co automatycznie prowadziło do postawienia pytania o efektywność procedur, przyczyny zróżnicowań. Tu ewaluatorzy weszli w rolę edukatorów (pokazując szerszą perspektywę, zadając pytania o procesy), ale i doradców. Relacje nawet w tym momencie były interaktywne – operatorzy programów przedstawiali swoje wyjaśnienia sytuacji, a ewaluatorzy na bieżąco, wspierając to dalszymi danymi, komentowali. Fakt, że dyskusja na forum Grupy Międzyprogramowej przebiegała w wewnętrznym gronie, mimo poruszenia wielu kwestii krytycznych, umożliwiła dialog, a sytuacje konfliktowe (nawet przy omawianiu istotnych problemów) były rzadkie.

Etapem trzecim były studia przypadków. Jedynym kluczem wyboru była decyzja operatorów programów. Ewaluatorzy przyglądali się w terenie tym projektom, które zostały wyznaczone przez instytucje zarządzające i wdrażające. Z reguły były to projekty problemowe, bo chciano się dowiedzieć o rzeczywistej sytuacji na poziomie beneficjentów. Dla każdego programu wykonano od piętnastu do osiemnastu studiów (po obu stronach granicy). Wyniki były prezentowane Instytucjom Zarządzającym każdego z programów, ale projekty łączono też w grupy tematyczne i dyskutowano je na forum międzyprogramowym. W tym trzecim etapie ewaluacji pojawiła się dodatkowa inicjatywa zlecającego. Z uwagi na obłożenie kwestiami rozszerzenia Unii Europejskiej Grupa Międzyprogramowa przestała funkcjonować, wyczerpały się też tematy do wielostronnej dyskusji. W związku z tym uruchomiono bilateralne platformy nauki (gdzie interesariusze z obu stron granicy spotykali się i dyskutowali przy asyście Urzędu Kanclerskiego i pomocy ewaluatorów).

[30] Głównymi aktorami procesu były Federalny Urząd kanclerski i MA poszczególnych programów. Udział partnerów z nowych krajów członkowskich był zróżnicowany. Najsilniejszy głos pochodził ze Słowacji, najsłabszy był udział Węgier.

Schemat 8. Kalendarz planowania i ewaluacji programów INTERREG dla Austrii



WYJAŚNIENIE
Austria przystąpiła do UE w 1995 roku, a więc już w trakcie trwania programu INTERREG II (1994–1999). Program dla granicy austriacko-czeskiej uruchomiono w 1995 roku. Finansowanie z INTERREG objęło tylko stronę austriacką (pieniądze z Funduszy Strukturalnych nie mogą finansować państw spoza UE) – regiony czeskie, słowackie, słoweńskie, węgierskie były finansowana znacznie mniejszymi pieniędzmi z PHARE–CBC.
Nowy program – INTERREG 3 przygotowano, zgodnie z okresem planowania polityki regionalnej UE, na lata 2000–2006.
W 2000 roku po jego zaplanowaniu, a przed jego uruchomieniem, zrobiono wymaganą ewaluację *ex ante*. Na jej podstawie poprawiono dokumenty programowe i uruchomiono program.
Już w trakcie trwania programu INTERREG 3 wykonano ewaluację e*x post* INTERREG 2 (zgodnie z wytycznymi w dwa lata po zamknięciu programu – czyli 1999+2).
W 2003 roku wykonano – zgodnie z wymaganiami UE – ewaluację *mid-term*. Ponieważ jednak programy INTERREG były w praktyce we wczesnym stadium działania, Komisja wprowadziła wymóg powtórzenia badania w 2005 roku *(mid-term update*).
Od 2000 roku Austriacy przygotowywali się do włączenia krajów sąsiedzkich do programu INTERREG (po rozszerzeniu UE, pieniądze przydzielone w 2000 roku na program miały współfinansować również projekty czeskie, słowackie, słoweńskie, węgierskie). Aby to włączanie sąsiadów odbywało się płynnie, strona austriacka zlecała od 2003 do 2005 roku szereg szczegółowych badań (np. kwestii poprawy promocji programu w Czechach, kwestii blokad finansowania, poprawienia spójności procedur).

Źródło: opracowanie własne.

Warto zwrócić uwagę, że działania z zakresu obserwacji, analizy i oceny były procesem ciągłym i opierały się na sprzężeniu zwrotnym (analiza – prezentacja wyników – reakcja interesariuszy – wnioski do dalszego badania – pogłębienie kwestii, które pojawiły się w reakcji na raport). Ewaluator cały czas odgrywał rolę animatora. Był też nie tyle konsultantem, co doradcą (gdyż wychodził daleko poza minimum swoich działań). Stopniowo, w momencie uzyskiwania szerszej perspektywy (informacji, zestawień i porównań z czterech programów) wchodził w rolę edukatora.

Analiza produktów ewaluacji

Stworzone kontakty [segment P1 modelu]

Niewątpliwie ewaluator stworzył silne, stałe kontakty ze zlecającym Federalnym Urzędem Kanclerskim oraz Sekretariatem koordynującym zlecenie. Urząd Federalny był też promotorem wyników (to on wyszedł z propozycją stworzenia platform wiedzy).

Omawiając zbudowane kontakty z interesariuszami, trzeba podkreślić różnice między instytucjami austriackimi a partnerami z nowych krajów członkowskich. W przypadku tych pierwszych były to kontakty stałe i rozległe – objęły wszystkich aktorów. W przypadku nowych krajów członkowskich problemem był brak ciągłości instytucjonalnej i silna rotacja kard. Uczestnicy procesu i negocjacji często się zmieniali, zmieniały się również instytucje odpowiedzialne za rozmowy i współpracę w programie. To wywoływało pewną frustrację tak u zlecających, jak i ewaluatorów, gdyż musieli odbudowywać kontakty po każdej zmianie. Zlecający podkreślał, że poczucie współwłasności, współuczestnictwo i kontakt z ewaluatorem byłby dużo silniejszy, gdyby wszyscy aktorzy procesu wnieśli wkład finansowy w kontrakt (był on finansowany wyłącznie przez stronę austriacką, centralną instytucję austriacką). Warto zwrócić uwagę, że w budowaniu kontaktów z szerszą grupą interesariuszy pomagała ewaluatorom instytucja zlecająca kontrakt. W tworzeniu relacji, jak też w całym badaniu ważną rolę odegrał fakt, że zespoły ewaluacyjne były mieszane (Austriak i osoba z kraju sąsiedzkiego). Zasadniczo pomagało to w rozeznaniu kontekstu i w samym badaniu na terenie nowych krajów członkowskich (choć w przypadku Węgier okazało się to problematyczne – konflikty między krajowym ewaluatorem a instytucją węgierką zostały przeniesione z innych projektów i zablokowały na jakiś czas działania w analizowanej ewaluacji).

Ocena rozległości kontaktów nawiązanych poza programem zależy od tego, jak zdefiniujemy sam program. Jeśli uznamy, że program to INTERREG, natomiast PHARE–CBC był czymś zupełnie zewnętrznym, wtedy niewątpliwie kontakty zewnętrzne były bardzo szerokie i głębokie. Jeśli jednak będziemy poszukiwać kontaktów nawiązanych na poziomie europejskim, poza PHARE–CBC i INTERREG austriackim, to należy stwierdzić, że nie miały one miejsca.

Wiedza zawarta w raporcie [segment P3 modelu]

W analizowanym przypadku ten punkt modelu odnosi się nie tylko do wiedzy zawartej w raporcie średniookresowym, ale także do wiedzy z całego procesu ewaluacji ciągłej.

Sam raport z badania nie przyniósł żadnych zaskakujących wniosków, a zawarta w nim wiedza dotyczyła głównie procedur, procesów i problemów z nimi związanych (*know-what*). Poszukiwanie przyczyn tego stanu rzeczy (*know-why*) oraz sposobów rozwiązywania problemów (*know-how & when*) były krystalizowane dopiero w późniejszych badaniach i rozmowach. Te rodzaje wiedzy zostały silnie rozbudowane w interaktywnym procesie badawczym, na przestrzeni kolejnych miesięcy. Dotyczyły zarówno procedur, jak i instytucji, procesów. Ciekawym jest jednak, że kwestie oceny funkcjonowania instytucji centralnych – głównego zlecającego – zostały zupełnie pominięte.

Badanie ciągłe (*on-going*), mimo iż dotyczyło kwestii technicznych, z zakresu zarządzania programami, dało uczestnikom nową wiedzę również o szerszym kontekście procesów czy o źródłach inspiracji. Wiedza ta płynęła z porównań i zestawień wyników czterech programów.

Bardzo ważnym w perspektywie koordynacji i wspólnego zarządzania czterema programami było zidentyfikowanie różnic w postrzeganiu między aktorami, a następnie wypracowanie wspólnego rozumienia terminów, pojęć, procedur (np. kwestia rozumienia „projektu ponadgranicznego" i związane z tym różnice w wyborze projektów). Ten rodzaj wiedzy nie mieści się w żadnej z kategorii wiedzy uwzględnionych w modelu.

Silny był również moduł wiedzy o aktorach (*know-who*) – kogo włączać, ale również w jaki sposób, jakie metody aktywizują poszczególnych uczestników – szczególnie tych z nowych krajów, jak z nimi dyskutować.

Rozpowszechnianie wyników [segment P4 modelu]

Raport ewaluacji *mid-term* był rozpowszechniany w wąskim gronie głównych interesariuszy (zlecający, Komitety Monitorujące, Instytucje Zarządzające i pośredniczace, sekretariaty techniczne), a następnie umieszczony na stronie internetowej. Był on raczej przyczynkiem do dalszych prac.

Zlecający miał jasną strategię rozpowszechniania wyników. Wychodził z założenia, że fragmenty badań prezentowane na bieżąco i jako wkład w określone dyskusje są dużo lepiej przyswajalne przez uczestników programu. Dlatego też wyniki ewaluacji były rozpowszechniane na bieżąco – w ramach prac i spotkań instytucji programu, głównie Komitetów Monitorujących i Sterujących (zwykle w formie krótkich not, prezentacji, podsumowań – były one potem zbierane w raporty, ale tylko w celu kontraktowym, jako podstawa wypłat). Zawsze towarzyszyła im dyskusja (animowana przez Sekretatriat Techniczny i ewaluatorów). Od 2006 roku kluczowym dla rozpowszechniania wyników były platformy wiedzy (*learning platforms*), na których spotykali się operatorzy programów z dwóch stron granicy, wspomagani przez ewaluatorów i WST. Rola ewaluatorów w tym procesie była aktywna.

Analiza użycia ewaluacji

Użycie ewaluacji [segment U1 modelu]

Niewątpliwie odbiorcy przyswoili sobie nową wiedzę. Nie zawsze postrzegali, w sposób uświadomiony, że jest to wiedza nowa, gdyż będąc włączonymi od samego początku w jej tworzenie, szybko ją interalizowali. W przypadku nowych krajów członkowskich duża część wiedzy została stracona z powodu niestabilności instytucjonalnej, braku ciągłości osobowej. Największy przyrost wiedzy nastąpił wśród aktorów słowackich (według oceny Federalnego Urzędu Kanclerskiego).

Odbiorcy ewaluacji dowiedzieli się nowych rzeczy przy okazji zderzenia wniosków z czterech programów. Pokazanie jednego z programów na tle pozostałych stworzyło nową perspektywę. Ewaluatorzy austriaccy pogłębili swoją wiedzę na temat skutecznego stosowania ewaluacji partycypatywnej. Niestety ewaluatorzy z krajów ościennych z powodu braku ciągłości personalnej skorzystali z tego w mniejszym stopniu.

Ciągłość procesu ewaluacji i jej wewnętrzny, nieformalny charakter uniemożliwiły ilościowe podejście do rekomendacji (ile zatwierdzono, ile wdrożono). Wdrożone rekomendacje czy raczej sugestie (bo wnioski z badań nie zawsze były przedstawiane w formie rekomendacji – te tworzyli dopiero uczestnicy dyskusji nad wynikami) dotyczyły zarówno strategii programu (dostosowania strategii do potrzeb nowych krajów), jak i procesów (procedur na poziomie programów i projektów). Rekomendacje z ewaluacji średniookresowej zostały też użyte jako punkt odniesienia dla jej uaktualnienia (Bauer-Wolf, 2005; Hummelbrunner, Omersel-Weeks, 2005; Kojan, Navratilova, 2005; Reiner, Karasz, 2005).

Badanie w znaczący sposób poszerzyło platformę współpracy i kontakty interesariuszy austriackich z nowymi krajami członkowskimi, jak również operatorów austriackich pomiędzy programami. Rozwijane kontakty zostały podtrzymane i są obecnie wykorzystywane.

Warto zaznaczyć, że w wyniku ewaluacji pojawiły się dwa specyficzne rodzaje wiedzy (nieuwzględnione w modelu). Pierwszym była wiedza zlecających na temat użyteczności podejścia interaktywnego (była ona budowana przez ewaluatora jeszcze przed kontraktem). Drugim rodzajem wiedzy było zidentyfikowanie różnic między aktorami w postrzeganiu tych samych terminów, a następnie wypracowanie wspólnego rozumienia kwestii kluczowych z perspektywy programów. Ten drugi rodzaj wiedzy można uznać za wiedzę koncepcyjną.

Wykorzystanie ewaluacji [segment U2 modelu]

Wiedza budowana w procesie ewaluacji ciągłej została wykorzystana w planowaniu kolejnych inicjatyw. Używano jej na bieżąco w pracach Grupy Zadaniowej ds. Rozszerzenia (przy przekształcaniu programów z lat 2000–2004), a także odwoływano się do niej (zarówno Urząd Kanclerski, jak i nowe kraje członkowskie) w dyskusjach nad programami na lata 2007–2013.

Wykorzystanie instrumentalne raportu można też prześledzić w publikacjach zewnętrznych. Samego raportu użyto jako źródła informacji w studiach zewnętrznych prowadzonych przez INTERACT (Bachtler et al., 2005; INTERACT, 2005a, b, c, d).

Ewaluatorzy austriaccy, zdobywając to kolejne wieloletnie doświadczenie, wzmocnili swoją i tak już znaczącą pozycję na rynku austriackim. Ewaluatorzy z krajów członkowskich również posiedli doświadczenie, które następnie wykorzystali w uaktualnieniu ewaluacji średniookresowej (dla nich będącej oddzielnym kontraktem). Nie wiadomo jednak, na ile skorzystali z niego w kolejnych studiach, poza omawianym badaniem ciągłym.

Trudno było w sposób jednoznaczny oszacować koncepcyjne wykorzystanie raportu w debatach dotyczących nowej polityki współpracy terytorialnej na lata 2007–2013. Jednak pewne ślady wniosków z ewaluacji średniookresowej można było dostrzec w dokumentach Federalnego Urzędu Kanclerskiego z 2004 roku, będących głosem w ówczesnej ogólnounijnej debacie nad przyszłym kształtem programu INTERREG IV. Ponadto spostrzeżenia na temat nowoczesnego systemu monitoringu dla programów współpracy terytorialnej rozwinięto w formie dalszych prac i studiów już na poziomie europejskim (Hummelbrunner, 2006). Wpłynęły one na sposób myślenia i praktyki monitorowania programu INTERREG IV. Ten efekt w dużej mierze wynika z aktywności naukowej samego ewaluatora.

Neutralne lub złe użycie [segmenty U3 i U4 modelu]

Raport nie został użyty w sposób symboliczny. Decyzje zmian nie były z góry zaplanowane, *gros* rozwiązań wypracowano w procesie ewaluacji. Moment dostarczania wniosków i rekomendacji był całkowicie zsynchronizowany z cyklami poszczególnych programów.

Raport – także ten dotyczący badania średniookresowego – nie był wyłącznie spełnieniem unijnego obowiązku. Stanowił punkt początkowy całego procesu – zarysował horyzont badań, pozwolił wstępnie zidentyfikować tematy, następnie pogłębiane. Nie stwierdzono też manipulowania wnioskami. Raport w całości został udostępniony w Internecie. Wnioski z ewaluacji ciągłej (*on-going*) były prezentowane i stale dyskutowane na forach programu.

Wnioski ze studium przypadku

Poniższe wnioski odnoszą się nie tylko do ewaluacji średniookresowej, ale do całego procesu, czyli ewaluacji ciągłej (*on-going*). Wyeksponowano to, pisząc nie o jednym raporcie, a o całym badaniu.

Przeprowadzona analiza dowiodła, iż **badanie poprawiło jakość planowania** w poszczególnych programach (głównie przystosowując programy do potrzeb nowych krajów członkowskich). Co więcej, dało też znaczący merytoryczny wkład do planowania kolejnych interwencji w nowym okresie programowym.

Badanie **wyraźnie spełniło też funkcję poprawy wdrażania projektów** we wszystkich czterech analizowanych programach. Obserwacje z ewaluacji i idące za tym wnioski praktyczne umożliwiły tzw. *streamlining* projektów (usprawnienie ich wdrażania, skrócenie czasów poszczególnych faz), poprawę szeregu procedur na poziomie programów, a przede wszystkim usprawnienie przepływów informacji między instytucjami zaangażowanymi we wdrażanie (po stronie austriackiej, jak też nowych państw członkowskich).

Natomiast **realizację funkcji *wzmocnienia odpowiedzialności w sektorze publicznym* (ang. *accountability*) należy uznać za bardzo ograniczoną**. Pokazanie skuteczności programów adresowano głównie do Komisji Europejskiej, a i tak było ono ograniczone (stopień analizy osiągnięcia celów został praktycznie pominięty). Pokazaniu stanu programu służył właśnie raport średniookresowy (przygotowywany, gdy programy były wciąż w swoim początkowym stadium – nie dotykał więc nawet kwestii efektów projektów). Wynikało to po części ze słabego stopnia zaawansowania programu (brak efektów, które można by odnieść do celów), a po części z przyjętej optyki ewaluacji (skoncentrowanie się na procedurach i potrzebach informacyjnych operatorów programu). Kwestia efektów i przystawalności projektów po obu stronach granicy była badana w dalszej części ewaluacji, ale wnioski z badań nie były prezentowane w formie oddzielnego raportu skierowanego do szerszej opinii publicznej czy Komisji.

Wreszcie **funkcja wsparcia procesów uczenia się w sektorze publicznym była bardzo silna**. Uczenie się miało kilka aspektów. Sam raport średniookresowy pomógł zidentyfikować mapę problemów – kolejne działania przynosiły zaś na bieżąco wyjaśnienia czynników i procesów warunkujących działanie interwencji. Ważnym elementem wiedzy nabytej przez uczestników było wypracowanie wspólnego rozumienia procesów, definiowanie ich i myślenie o projektach współpracy terytorialnej. Badanie spełniło tę funkcję ewaluacyjną głównie w zakresie budowania zdolności ewaluacyjnych – wśród aktorów austriackich było to oswojenie ich i przekonanie do ewaluacji partnerskiej, interaktywnej. Uczestnicy z nowych krajów członkowskich nauczyli się też współpracy i dyskusji. Funkcja uczenia się wyszła też poza badane programy i może mieć wpływ na całą politykę. Obserwacje z badań posłużyły do wypracowywania nowego podejścia do monitoringu pomocy terytorialnej (ang. *impact assessment*). Badanie dawało wiedzę na temat dobrych praktyk i ciekawych rozwiązań, czerpało je jednak przede wszystkim z czworokąta badanych programów, a więc w znikomym stopniu sięgając poza badaną rzeczywistość. Wreszcie badanie, koncentrując się na bieżących potrzebach informacyjnych uczestników, nie dało znaczącego wkładu w rozumienie szerszych teorii, podstaw polityki.

Ostatnia **funkcja – wzmocnienia partnerstwa i poczucia współwłasności uczestników – również została spełniona**. Zgodnie z założeniem zlecających, program pomógł nawiązać, a następnie wzmocnić relacje z instytucjami z nowych krajów członkowskich. Interaktywność procesu i jego sprzężenie dla czterech programów umożliwiło zbudowanie sieci współpracy i kontaktów także miedzy programami. Badanie aktywizowało interesariuszy i wią-

czyło w dyskusję nad programami instytucje i osoby dotychczas słabo obecne (NMS, niektóre regiony). Najsilniej jest to widoczne w zwiększeniu aktywności aktorów słowackich.

Przechodząc natomiast do podsumowania **czynników, które wpłynęły na wykorzystanie ewaluacji**, można wyciągnąć następujące wnioski.

Kluczowym czynnikiem okazał się **kontekst geopolityczny** – perspektywa rozszerzenia Unii Europejskiej i związana z tym konieczność zreformowania programów pod kątem udziału nowych krajów. Z tym łączył z kolei niezmiernie ważny czynnik wykorzystania synergii między czterema programami – jednoczesne ich badanie i współanalizowanie. Było to silnie eksploatowane tak przez zlecających, jak i przez ewaluatora.

Ważnym okazała się też **postawa głównych zlecających** (w tym czynnik osobowościowy) – ich przekonanie o wartości nowego podejścia do ewaluacji, zaangażowanie oraz silna promocja samego pomysłu i pojawiających się efektów. Łączyło się to i wzajemnie wzmacniało z zaangażowaniem głównego ewaluatora, jego determinacją w użyciu partycypatywno-interaktywnego podejścia.

Interaktywny, partycypatywny i zorientowany na odbiorców sposób podejścia do badania, a także aktywna rola ewaluatora jako animatora, doradcy i edukatora, w sposób jednoznacznie pozytywny wpłynęły na użycie badania. Wysoka aktywność zlecającego i stworzenie instytucjonalnych platform wymiany informacji (Międzyprogramowa Grupa Sterująca, platformy wiedzy) były dodatkowym, znaczącym czynnikiem.

Również **sposób przekazywania wiedzy i forma relacji** okazały się ważne. Użycie ewaluacji zostało zintensyfikowane poprzez przyjętą formę raportowania – a właściwie formę przekazywania wiedzy – wprowadzanej na bieżąco, w sposób ciągły, jako wkład w toczącą się dyskusję. Była ona udostępniana nie w formie jednolitego raportu, a w małych pakietach – prezentacji, podsumowań dostosowanych do bieżącej grupy odbiorców. Bliski kontakt oparty na zaufaniu między zlecającymi, zaangażowanymi interesariuszami a ewaluatorem umożliwiał swobodniejszą dyskusję dotyczącą trudnych tematów, problemów programu i projektów. Należy jednak podkreślić, że zaufanie udało się zbudować dzięki ciągłości procesu (horyzont czasowy badania to trzy i pół roku) oraz dzięki temu, że był to proces wewnętrzny – dyskusja odbywała się w szerokim, ale stałym gronie operatorów programu.

Nie bez znaczenia wreszcie był **rodzaj wiedzy**. Była ona bardzo praktyczna, oparta w zasadzie na sumie doświadczeń uczestników (a więc dużo łatwiejsza do internalizacji), dostosowana do ich potrzeb informacyjnych.

4.4.2. Przypadek 2: IIIA–Irlandia-Irlandia Północna

Analiza wkładu w ewaluację

Warunki kontekstowe [segment W1 modelu]

Program INTERREG IIIA Irlandia – Irlandia Północna obejmował swoim działaniem sześć podregionów – pięć po stronie Wielkiej Brytanii i jeden re-

gion Republiki Irlandii (z hrabstwami Cavan, Donegal, Lentrim, Louth, Monaghan i Sligo, nazywanymi regionem BMW – *Border Midlnands & Western Region*, najbardziej peryferyjnym podregionem Irlandii). Kultura administracyjna współpracujących regionów była teoretycznie bardzo zbliżona (anglosaska tradycja zarządzania opartego na wynikach z silnie rozwiniętą kulturą ewaluacji), jednak w praktyce wyraźny był ostry podział wynikający z trwającego od 500 lat konfliktu politycznego o Irlandię Północną. W 2000 roku konflikt ten wciąż się tlił, skutkując nieufnością między mającymi współpracować instytucjami i wspólnotami. Tak więc głównym politycznym celem programu INTERREG (trzeciej już edycji realizowanej na tym obszarze) było stworzenie wspólnoty interesów i impulsu do współpracy między różnymi wspólnotami lokalnymi (protestanckimi, katolickimi, lojalistami, republikanami) po obu stronach granicy. Jednocześnie polityczne i strategiczne znaczenie programu było stosunkowo małe. Analogiczny cel, ale na dużo większą skalę, realizował inny program UE – *PEACE & Reconciliation* (Pokój i Pojednanie), dysponujący czterokrotnie większym budżetem (na PEACE przeznaczono 750 mln euro, rozszerzając je następnie o dodatkowe 200 mln, podczas gdy cały INTERREG III miał do dyspozycji 179 mln euro). Uwaga polityków, administracji publicznej, potencjalnych beneficjentów, opinii publicznej oraz mediów koncentrowała się w sposób oczywisty na programie PEACE.

Idea partnerstwa i zaangażowania szerokiej gamy interesariuszy ze wszystkich regionów (strategia łączenia poprzez wiązanie instytucji) zaowocowała bardzo skomplikowaną i rozbudowaną strukturą administracyjną programu. Instytucją, która pełniła rolę Instytucji Zarządzającej i Wspólnego Sekretariatu Technicznego był SEUPB (ang. *Special European Union Programmes Body*) – biuro zlokalizowane w Republice Irlandii, w przygranicznym mieście Monaghan[31]. Do jego kompetencji należało zarządzanie programem, bieżący monitoring i koordynacja zaangażowanych w program instytucji. Komitet Monitorujący (KM) złożony z dwudziestu ośmiu osób (przedstawiciele ministerstw, agencji rządowych i regionalnych, partnerów lokalnych i pozarządowych) nadzorował proces całego wdrażania, podejmował strategiczne decyzje. Oprócz niego funkcjonował Komitet Sterujący (KS) składający się z osiemnastu osób (po 50% z północy i południa wyspy), odpowiedzialny za wybór projektów. W odróżnieniu od innych programów INTERREG III skład personalny Komitetu Sterującego był zupełnie inny niż Monitorującego. Co więcej, spotkania tego pierwszego odbywały się w różnych miejscowościach naprzemiennie po obu stronach granicy (tak więc członkowie KS ciągle podróżowali, a miejsca ich spotkań rzadko się powtarzały). Wdrażaniem poszczególnych priorytetów tematycznych programu i płatnościami na rzecz beneficjentów zajmowały się Agencje Wdrażające (ang. *Implementing Agents*). Z reguły były to *quangosy* irlandzkie i północnoirlandzkie, lub departamenty ministerstw zajmujące się da-

[31] SEUPB był jednym ze wspólnych, ponadgranicznych agencji powołanych na mocy pokojowego Porozumienia Wielkopiątkowego do współpracy i zarządzania programami w regionie. SEUPB zajmował się również wdrażaniem programu PEACE. Programem INTERREG zajmowały się tylko cztery z kilkunastoosobowej załogi. Pozostałe dwa biura znajdowały się w Belfaście i w Omagh.

nym tematem. Stworzono również specjalne ponadgraniczne partnerstwa złożone z instytucji – głównie zrzeszeń aktorów lokalnych z Republiki Irlandii i Irlandii Północnej – i działające nie tylko na potrzeby INTERREG, ale i programu PEACE. Zasadą było, że Instytucja Wdrażająca składała się z partnerów po obu stronach granicy. Listę zaangażowanych instytucji zwiększał szeroki zakres programu. Co prawda program miał tylko cztery priorytety, ale te były podzielone na jedenaście działań. Każde z nich, jako że skoncentrowane na specyficznym temacie, było wdrażane przez oddzielną Agencję Wdrażającą. Agencje dodatkowo były nadzorowane przez irlandzkie i północnoirlandzkie ministerstwa. W rezultacie tylko od strony zarządzania (a więc wyłączając beneficjentów) w program było zaangażowanych około dwadzieścia instytucji (Pricewaterhousecoppers, 2003: 37), a szacunkowo ponad pięćdziesiąt osób! To wyjątkowo dużo, biorąc pod uwagę skalę programu.

Jak już wspomniano, program miał cztery priorytety i jedenaście działań. Priorytet 1 (69 mln) – „Zintegrowane Strategie Rozwoju Lokalnego" obejmował: działania z biznesu, wsparcie gospodarki opartej na wiedzy, wsparcie zasobów ludzkich oraz rozwój inicjatyw wiejskich. Priorytet 2 (81 mln) – „Rozwój infrastruktury i ochrona środowiska" obejmował ponadregionalną infrastrukturę gospodarczą, zarządzanie ochroną środowiska oraz rozwój odnawialnych zasobów energii. Priorytet 3 (19 mln) – „Społeczeństwo obywatelskie i współpraca" prowadził działania na rzecz infrastruktury społecznej i wspólnot lokalnych oraz ochrony zdrowia. Zaś 4 Priorytet (10 mln) objął pomoc techniczną (Pricewaterhousecoppers, 2003: 71; Special EU Programmes Body, 2001).

Program miał wyjątkowo późny start. Negocjacje polityczne w regionie, a potem negocjacje z Komisją Europejską trwały do końca 2002 roku. W momencie podejmowania badania ewaluacyjnego (maj 2003) wybrano dopiero pierwszych czterdzieści dziewięć projektów, przeznaczając na nie 15% środków programu (przy czym były to pieniądze zakontraktowane, w praktyce zaś nie wydano nic). To oznaczało, że w rzeczywistości badanie ewaluacyjne przypadło nie na połowę, a na pierwsze miesiące działania programu. Nie miał on żadnych produktów, a tym bardziej żadnych wymiernych efektów.

Wkład uczestników [segment W2 modelu]

Instytucją zlecającą ewaluację był SEUPB. Ponieważ był zaangażowany we wdrażanie programu PEACE, to jako zespół miał szerokie, całościowe spojrzenie na obszar graniczny, wraz z jego trudną historią. Badanie średniookresowe ewaluacyjne było traktowane jako wymóg techniczny i to dość uciążliwy. Personel SEUPB był całkowicie pochłonięty wprawieniem w ruch instytucjonalnej i proceduralnej machiny programu. Racjonalność badania programu, który fizycznie nie wystartował, a który był weryfikowany niecały rok wcześniej (w ewaluacji *ex ante*) wydawała się wątpliwa. Sam projekt Zakresu Zadań ewaluacji był kopią wymagań Komisji Europejskiej i wytycznych do nieco wcześniejszego kontraktu na ewaluację Programu PEACE. Zlecający powołał jednak specjalną Grupę Sterującą Ewaluacją (były to osoby zainteresowane ewaluacją, oddelegowane z Komitetu Monitorującego, reprezentujące

różne pola tematyczne, instytucje, technokratów i polityków, jak i obie strony granicy w równych proporcjach), której zadaniem był merytoryczny nadzór nad przyszłym kontraktem oraz streszczanie postępów przed Komitetem Monitorującym. GSE powstała już na etapie pisania zakresu zadań i miała decydujące słowo na temat jego formy i zakresu. Duży wkład w ostateczny zakres zadań miało też Wspólne Biuro Statystyczne Irlandii Północnej (weryfikujące wymagania dotyczące analizy statystycznej). Na realizację badania przeznaczono siedem miesięcy.

Wykonawcą kontraktu został północnoirlandzki (Belfast) oddział PricewaterhouseCoopers we współpracy z małą firmą doradczą z Dublina (Endocon). Lider konsorcjum, PwC, był firmą o globalnej, korporacyjnej kulturze konsultingowej. Podejście zaproponowane w ofercie bazowało na jej kontraktach z ewaluacji innych Inicjatyw Wspólnotowych w Irlandii Północnej (Equala, Leadera), uwypuklono jednak transgraniczność i zintegrowane podejście do badania programu. Cały zespół badawczy liczył pięć osób.

Analiza procesu ewaluacji

Organizacja procesu badawczego miała fazy bazujące na metodologii MEANS, podejściu sugerowanym w podręczniku DG-Budżet, lecz przede wszystkim wynikające ze standardowego podejścia wypracowanego przez korporację PwC. Badanie miało cztery etapy:

- Etap 1 – nawiązanie kontaktu z klientem, wyjaśnienie szczegółów kontraktu, zakresu zadań, oczekiwań;
- Etap 2 – zebranie i analiza dokumentów, przegląd baz danych, analiza społeczno-ekonomiczna;
- Etap 3 – badania terenowe oraz wywiady i badania ankietowe;
- Etap 4 – pisanie raportu, weryfikacja rekomendacji.

Zakres działań pierwszych dwóch etapów odpowiadał fazie strukturalizacji (i pod takim tytułem został przeanalizowany poniżej). Natomiast etapy trzeci i czwarty zostały przedyskutowane jako łączące obserwację, analizę i ocenę.

Strukturalizacja [segment D1 modelu]

Zestaw pytań postawionych w ewaluacji odpowiadał w całości liście tematów sugerowanych w wytycznych Komisji Europejskiej. W fazie strukturalizacji ewaluatorzy, po dyskusji ze zlecającymi, skoncentrowali jednak badanie na kwestiach zarządzania i monitoringu (a nawet, z powodu jego początkowego stanu, na dwóch pierwszych etapach wdrażania – promocji i procesie wyboru projektów). Był to wybór oczywisty – nie było efektów, które można by analizować. Z tych samych względów obrane podejście badawcze również było zorientowane na zarządzanie. Badanie skierowane było od początku do menadżerów programu i miało dostarczyć im technicznych informacji. Ze strony zlecającego wyszedł wymóg regularnych konsultacji między Grupą Sterującą Ewaluacją a ewaluatorami (co jest standardem kontraktów anglosaskich i wynika z szerszej kultury administracyjnej). Przybrało to formę comiesięcznych

spotkań roboczych. Ich plan, jak też harmonogram realizacji badania zaproponował i przedyskutował z GSE zespół ewaluacyjny. W dyskusji tej ewaluatorzy byli bardzo elastyczni. Relacje w fazie strukturalizacji można więc określić jako interaktywne.

Jak stwierdzili badani ewaluatorzy, pogłębiona metodologia badania nie odbiegała od standardowego podejścia stosowanego przez ich firmę do innych ewaluacji Inicjatyw Wspólnotowych. Opierała się na analizie dokumentów, wywiadach z zarządzającymi i ankietach z beneficjentami oceniającymi jakość procesów (szerzej o tym w dalszej części). W zasadzie, poza jednym elementem, nie pojawiły się specjalne techniki uwypuklające transgraniczność programu. Tym wyjątkiem był silny nacisk na prowadzenie analizy po obu stronach granicy, z uwzględnieniem szerokiego spektrum interesariuszy (choć i to podejście, korzystające z szerokich wywiadów, PwC stosował do Equala, gdzie badano szerokie grupy partnerstw).

Obserwacja, analiza i ocena [segmenty D2 i D3 modelu]

Nietypowym w skali wszystkich INTERREG-ów był fakt, iż program Irlandia – Irlandia Północna miał dobrze skonstruowaną, jasną bazę wskaźników monitoringu (*nota bene* zarządzaną przez prywatną firmę). Baza (będąca częścią systemu monitorowania wszystkich programów z funduszy strukturalnych) umożliwiała ewaluatorowi tworzenie szybkich raportów o stanie interwencji, zaawansowaniu finansowym i rzeczowym danych priorytetów i działań programu. Teoretycznie ewaluatorzy mieli więc możliwość dobrej analizy ilościowej efektów. W praktyce jednak nie było materiału do tejże analizy, gdyż program dopiero rozpoczął pracę. W tym wypadku okazało się, że głównymi metodami będą metody jakościowe i interaktywne (wywiady, ankiety, obserwacja uczestnicząca) nastawione na system zarządzania programem. Jak określił to jeden z badanych, „niewiele było w bazach, więc praca musiała opierać się na rozmowie z ludźmi". Wyzwaniem okazała się rozległość i stopień skomplikowania instytucji włączonych (czy potencjalnie zaangażowanych) w program. Udało się ewaluatorom uchwycić kilka perspektyw. Wywiady prowadzili z przedstawicielami Komisji Europejskiej, z Instytucją Zarządzającą będącą jednocześnie Sekretariatem (SEUPB), z przedstawicielami Komitetu Monitorującego i Komitetu Sterującego. Ponadto ewaluatorzy przeprowadzili wywiady ze wszystkimi dziewiętnastoma Agencjami Wdrażającymi, dodatkowo prowadząc obserwację uczestniczącą (po jednym dniu w każdej z instytucji). Zorganizowali również panel dyskusyjny z osobami oceniającymi projekty na temat procedury wyboru projektów. Badania ankietowe przeprowadzono z członkami Komitetów Monitorującego i Sterującego, z instytucjami które starały się o fundusze UE w programie INTERREG (58 ze 140 instytucji, stopa zwrotu 41%), a także, co nietypowe, z potencjalnymi beneficjentami (100 ankiet z 900 instytucjami, które mogły starać się o środki INTERREG). Część z pytań ankietowych, a także panel dyskusyjny dał interesariuszom możliwość oceny programu – byli więc oni włączeni w krytykę i formułowanie wniosków. Ponadto, już na etapie formułowania wniosków ewaluatorzy mieli comiesięcz-

ny kontakt z GSE, i jak zauważyli sami zlecający, podchodzili bardzo elastycznie do zmieniającej się sytuacji programu i zakresu kontraktu. Warto jednak zaznaczyć, że GSE, pozostając w stałej relacji z ewaluatorami, rzadko pracowała w komplecie. Wielu jej członków nie uczestniczyło w spotkaniach, ale dawało swoje komentarze w korespondencji e-mailowej. Część z nich była też mniej aktywna, gdyż tematy w których się specjalizowali, nie pojawiały się w ewaluacji średniookresowej INTERREG-u. W ostatniej fazie badania – przygotowywania wniosków i rekomendacji – ewaluatorzy skonsultowali racjonalność i operacyjność swoich pomysłów tak z SEUPB, jak i Grupą Sterującą Ewaluacją. Należy więc uznać, że rekomendacje zostały „przetestowane". Wprowadzono kilka poprawek, ale nie były one „dyktowane przez klienta" i nie naruszały w istotny sposób krytycznego wydźwięku poszczególnych spostrzeżeń.

Na podstawie powyższego, relacje między ewaluatorami a interesariuszami można określić jako „ograniczone interakcje". Ewaluatorzy odegrali rolę konsultantów zbierających znane informacje i zestawiających kilka perspektyw. Koncentrowali się na szybkości metod, identyfikowaniu problemów zarządzania i oczekiwaniach klienta.

Analiza produktów ewaluacji

Stworzone kontakty [segment P1 modelu]

Kontakt stworzony między ewaluatorami a zlecającymi był dobry. Dowodem na to, prócz opinii zlecających i ewaluatorów, jest fakt, że PwC kontynuowało również uaktualnienie ewaluacji średniookresowej (dwa lata później). Sekretariat programu (SEUPB), a także osoby z GSE były promotorami wyników ewaluacji na forum Komitetu Monitorującego. SEUPB, internalizując rekomendacje, sam przygotował ich *follow-up* (sprawdzenie postępów z realizacji). Ewaluatorzy nie zbudowali szerszej koalicji wokół rekomendacji z innymi interesariuszami (np. beneficjentami projektów czy Agencjami Wdrażającymi), ale wydaje się, że po prostu nie było takiej potrzeby. Wnioski i rekomendacje dotyczyły zarządzania na poziomie programu, jej jedynymi adresatami był SEUPB i Komitet Monitorujący, a więc instytucje, z którymi ewaluatorzy już współpracowali. Stworzone kontakty z szerszą grupą interesariuszy miały jednak inny aspekt. Przed ewaluacją interesariusze (przede wszystkim poszczególne dziewiętnaście instytucji będących Agencjami Wdrażającymi) postrzegały siebie jako niezależne, suwerennie funkcjonujące ciała. Dopiero za sprawą prowadzonej ewaluacji instytucje te zdały sobie sprawę z zależności oraz interakcji, jakich wymaga wdrażanie programu. Pytania ewaluatorów zwróciły ich uwagę na program (dotychczas koncentrowali się na krajowych funduszach strukturalnych i programie PEACE) oraz uświadomiły im ich rolę. Ponadto część potencjalnych beneficjentów dowiedziała się o programie dopiero z ankiet ewaluacyjnych. Ewaluacja odegrała więc rolę narzędzia promocji, budowania instytucjonalnej świadomości programu i współodpowiedzialności (jak określił to jeden z badanych: „sense of uni-programme responsibility").

Poza programem ewaluatorzy nie nawiązali żadnych kontaktów. Nie było porównań do innych programów INTERREG czy współpracy z innymi zespołami europejskimi, które w tamtym czasie wykonywały analogiczne badania.

Wiedza zawarta w raporcie [segment P3 modelu]

Wiedza płynąca z raportu skierowana była do kluczowych instytucji zarządzających programem (SEUPB i Komitet Sterujący) i okazała się mieć charakter czysto techniczny. Z uwagi na brak efektów programu wnioski koncentrowały się na technicznych obserwacjach dotyczących efektywności procesów zarządzania – podziału kompetencji, koordynacji między działaniami, procedury selekcji projektów, promocji programu i systemu płatności, poprawności zastosowanych wskaźników i metod gromadzenia danych czy wreszcie zastosowania zasad horyzontalnych UE w projektach (równe szanse, zrównoważony rozwój).

Podobnie rekomendacje (w sumie dwadzieścia jeden rekomendacji podzielonych na trzy grupy – strategii, celów i działań oraz mechanizmów wdrażania) były zorientowane na procesy adresowane głównie do SEUPB, poruszały one czysto techniczne kwestie zmian procedur, korekt w kompetencjach instytucji czy monitorowania poszczególnych działań. Nawet w grupie wniosków dotyczących strategii i głównych celów jedyny szerszy komentarz (odnoszący się do różnego rozumienia trasngraniczności przez aktorów programu, a więc mogący mieć charakter szerszej wiedzy na temat założeń i teorii) nie był konkluzywny. Identyfikował jedynie problem i rekomendował instytucjom programu wypracowanie jednolitego rozumienia tego szerokiego terminu (bez wskazania źródeł teoretycznych czy szerszej podstawy filozoficznej). Wiedza, którą dał raport, to *know-what* (co działa, a co nie) i *know-who* (kto ma działać w celu poprawy obecnego stanu). Incydentalnie raport sugerował też szersze unijne inspiracje (przykład list rezerwowych projektów – rozwiązanie angielsko-francuskie). W opinii samych odbiorców nie była to wiedza odkrywcza. Raport nie przyniósł nic, co byłoby zaskoczeniem dla zlecających. Służył raczej jako dokument formalizujący i kodyfikujący wiedzę SEUPB (o czym szerzej przy użyciu ewaluacji).

Sam raport postrzegany był przez odbiorców jako nieco przeładowany – sztucznie i nadmiernie rozbudowany, biorąc pod uwagę początkowe stadium programu. Niemniej streszczenie raportu uznano za bardzo funkcjonalne, zaś sam raport za przejrzysty (co ułatwiało poruszanie się w dokumencie i jego używanie).

Rozpowszechnianie wyników [segment P4 modelu]

Rozpowszechnienie raportu nastąpiło od razu, ale było ograniczone do głównych instytucji zarządzających – Komitetu Monitorującego i Instytucji Zarządzającej będącej jednocześnie Sekretariatem programu (SEUPB). Raport rozpowszechniano tradycyjnymi kanałami komunikacji między instytucjami programu (przekazanie do Komitetu Monitorującego, rozesłanie e-mailem).

Ewaluator był włączony w rozpowszechnianie wyników w ograniczonym stopniu – zaprezentował je przed wąskim gronem Grupy Sterującej Ewaluacją, a potem przed Komitetem Monitorującym. Raport opublikowano z czasem na stronie internetowej programu. Proces rozpowszechniania wyników należy uznać za ograniczony do wąskiej grupy odbiorców, zaś rolę ewaluatora w nim za pasywną.

Analiza użycia ewaluacji

Użycie ewaluacji [segment U1 modelu]

Raport nie przyniósł nic, co byłoby zaskoczeniem dla zlecających. Użycie miało głównie charakter instrumentalny. Ewaluacja uporządkowała techniczną wiedzę instytucji zarządzających o ograniczeniach i koniecznych korektach programu (przede wszystkim wiedzę SEUPB), wpływając na ich sposób myślenia o różnych kwestiach zarządzania programem (które pogłębiali wraz z postępami we wdrażaniu). Raport dał tym instytucjom impuls i podstawę do działania oraz formalny punkt odniesienia w dyskusjach z innymi interesariuszami, wzmocnił ich argumenty w tychże dyskusjach. Stał się też punktem odniesienia w mierzeniu własnych postępów. Dowodem instrumentalnego użycia było również wdrożenie większości rekomendacji (*follow up* wykonany w kilka miesięcy po badaniu przez SEUPB dał możliwość weryfikacji ich realizacji) (Special EU Programmes Body, 2004). Należy tu jednak zaznaczyć, że sytuacja w programie zmieniała się bardzo szybko i w momencie opublikowania raportu część działań była już wdrożona – kilka rekomendacji było więc nieaktualnych. Rekomendacji przedstawionych w raporcie użyto również jako punktu odniesienia aktualizacji ewaluacji średniookresowej (Pricewaterhousecoppers, 2005). Wreszcie za instrumentalne użycie należy uznać opisywany już impuls w budzeniu współodpowiedzialności i świadomości istnienia programu wśród całej rzeszy instytucji pośredniczących i potencjalnych beneficjentów.

Pewne znamiona użycia koncepcyjnego nosiło kontynuowanie przez SEUPB tematu geograficznej równości w programie (równe reprezentowanie północy i południa), kwestii poruszonej przez ewaluatorów, a następnie dyskutowanej na poziomie politycznym. Jak przyznali sami pracownicy SEUPB, raport pomógł im również w dalszym myśleniu o transgraniczności i współpracy nie tylko w kontekście wykorzystania środków UE, ale całego procesu zarządzania i relacji między instytucjami. Za użycie koncepcyjne można również uznać doświadczenie zdobyte przez ewaluatorów i zlecających. Ci pierwsi nauczyli się, iż ewaluacja INTERREG-ów powinna być prowadzona w mieszanym zespole (z ewaluatorami z drugiej strony granicy). W wywiadach podkreślili też, że zdali sobie sprawę z siły studiów przypadków i wizyt studyjnych – nawet przy ewaluacjach programów słabo zaawansowanych. Te elementy strategii badawczej (całościowe spojrzenie i wizyty studyjne) stały się standardem ich późniejszych badań. Badanie wzbogaciło również *port folio* firmy. Zlecający natomiast (przede wszystkim pracownicy SEUPB) zdali sobie sprawę z tego, że bardzo ważne jest stosunkowo wczesne zarysowanie

pytań badawczych w formie krótkiego, spójnego ToR, gdyż czas potrzebny na przygotowanie badania i jego prowadzenie był dużo dłuższy niż początkowo zakładali. Po drugie zaś, ciągły kontakt z ewaluatorami, interakcje ułatwiły płynne wprowadzanie rekomendacji. Te wnioski zostały wykorzystane dwa lata później przy kolejnej ewaluacji (aktualizacji *mid-term*). Nastąpił więc proces uczenia się, choć należy zauważyć, w obydwu przypadkach znów była to wiedza techniczna.

Wykorzystanie ewaluacji [segment U2 modelu]

Wykorzystanie ewaluacji było bardzo ograniczone. Rekomendacje z raportu nie były użyte w kolejnych edycjach programu, nie używano raportu jako źródła innych analiz. Jedynym, słabym sygnałem wykorzystania instrumentalnego było przedstawienie raportu w publikacji INTERACT (konkretnie kwestii prowadzenia badań potencjalnych beneficjentów) jako dobrej praktyki badań INTERREG. Niewątpliwie dowodem instrumentalnego wykorzystania tej ewaluacji był fakt, że PwC zdobyło kontynuację badania – kontrakt na uaktualnienie ewaluacji średniookresowej. Pewnym sygnałem wykorzystania koncepcyjnego była natomiast dyskusja nad sposobami mierzenia i egzekwowania geograficznej równości w programie (równego reprezentowania północy i południa), zapoczątkowana przez wyniki ewaluacji, a rozwijana zarówno na poziomie instytucji zarządzających, jak i w politycznej dyskusji na szczeblu międzynarodowym, międzyministerialnym. Ewaluacja pokazała również, że polityczna strategia integracji podzielonych społeczności (włączanie poprzez współdziałanie) odbywała się kosztem efektywności i prostoty struktur. Nie było to *novum* ani dla operatorów programu, ani dla polityków, ale badanie postawiło tę kwestię (swoisty wybór coś za coś) w sposób jasny. Słabe dowody wykorzystania ewaluacji można tłumaczyć jej czysto technicznym charakterem, zorientowaniem na szczegółowe procesy zarządzania, które to z kolei wynikało ze słabego stopnia zaawansowania programu.

Neutralne lub złe użycie [segmenty U3 i U4 modelu]

Po analizie kwestii neutralnego i złego użycia należy stwierdzić, na podstawie wywiadów i *follow up* rekomendacji, że pojawiły się pewne dowody użycia symbolicznego. W przypadku kilku kwestii (np. zmian w metodach selekcji projektów, ściślejszego monitoringu kwestii finansowych i działań partnerstw) zmiany były już zaplanowane przez Organ Zarządzający (SEUPB), a ewaluacja dała im tylko formalne wsparcie i legitymizację tych działań. Pojawił się też uzasadniony brak użycia. W trakcie trwania badania program dynamicznie się rozwijał, w rezultacie czego część zmian zaproponowanych w wyniku badania była już nieaktualna w momencie publikacji raportu. Co do kwestii pozornego użycia, badanie rzeczywiście rozpoczęto wyłącznie z powodu wymogów Komisji Europejskiej, jednak sam proces badania i jego efekty zinternalizowano i czynnie użyto w programie. Komisja Europejska nie przedstawiła zresztą żadnej reakcji na raport.

Wnioski zostały opublikowane w całości, a postęp w rekomendacjach śledził (i publikował dla innych instytucji) sam organ zarządzający. Nie doszło więc do nadużycia ani manipulacji wnioskami.

Wnioski ze studium przypadku

Raport **nie przyczynił się w żadnym stopniu do poprawy planowania** ani badanej, ani kolejnych interwencji. Fakt ten wytłumaczyć można tym, że badanie w praktyce nastąpiło nie w połowie programu, a na jego bardzo wczesnym stadium. Weryfikowanie i zmienianie strategicznych kierunków programu w rok po ewaluacji *ex ante* (której centralnym elementem jest właśnie weryfikacja racjonalności planowania) byłoby nieuzasadnione.

Siłą rzeczy, z uwagi na brak efektów, **badanie skoncentrowało się na procesach zarządzania**. W ocenie interesariuszy i w świetle późniejszych badań (*follow up* rekomendacji, uaktualnienie ewaluacji średniookresowej) raport okazał się pomocnym narzędziem w poprawie wdrażania interwencji. Dał impuls przede wszystkim do bieżących zmian analizowanego programu (poprawy procedur, koordynacji, promocji), i stanowił formalne uzasadnienie zmian już planowanych. Dał również słaby impuls do szerszej refleksji nad kwestiami transgranicznego zarządzania (równego reprezentowania stron nie tylko w podziale środków, ale i w procesach zarządzania) i związanej z tym kompleksowości rozwiązań instytucjonalnych.

Raport praktycznie **nie poruszał kwestii skuteczności i wydajności wykorzystania środków**, gdyż w momencie badania nie było jeszcze żadnych efektów (niewielka grupa projektów dopiero się rozpoczęła). Stąd nie miał wkładu we wzmacnianie odpowiedzialności sektora publicznego. Oczywiście informacja o słabym poziomie wykorzystania środków mogłaby świadczyć o słabej skuteczności instytucji publicznych. Jednak w tym wypadku była ona wytłumaczona czynnikami zewnętrznymi (całym systemem pomocy UE i opóźnieniami na poziomie UE) i skierowana do wąskiej grupy (operatorów programu), a nie szerokiej publiki (społeczności lokalnych). Była więc technicznym sygnałem na temat poprawy wdrażania programu (i uniknięcia zasady N + 2).

Funkcja uczenia się z tego badania ewaluacyjnego miała ograniczony charakter. Raport przyniósł jego odbiorcom (wąskiej grupie głównych operatorów programu) wiedzę techniczną, a nie kontekstową, czy szerszą refleksję nad mechanizmami interwencji, teorii leżących u jej podstaw. Jedyną kwestią wychodzącą poza bieżące zarządzanie (choć wciąż z nią powiązaną) była równość obu stron granicy i współzarządzanie programem. Ewaluacja jasno uświadomiła, że polityczna strategia integracji podzielonych społeczności (włączanie poprzez współdziałanie) odbywała się kosztem efektywności i prostoty struktur.

Procesy uczenia się z tego badania (choć znów bardzo techniczne) zaobserwowano jednak w polu umiejętności ewaluacyjnych. Ewaluatorzy zrozumieli siłę podejścia zintegrowanego (badań po obu stronach granicy, lokalnych zespołów ewaluacyjnych), a także znaczenie badań terenowych i studiów przypadków. Wykorzystali to w kolejnej ewaluacji. Zlecający natomiast nauczyli

się lepiej planować czasowo proces ewaluacji, formułować konkretne oczekiwania już na etapie zamówienia na badanie, a przede wszystkim przekonali się, iż włączanie innych instytucji i dialog z ewaluatorem znacznie usprawniają proces akceptacji i wdrażania rekomendacji.

Wreszcie przechodząc do ostatniej funkcji, należy uznać, że **ewaluacja przyczyniła się do wzmocnienia partnerstwa i poczucia współwłasności uczestników** interwencji publicznych. Przyczyną zmian nie był jednak sam raport, lecz proces prowadzenia badania. Szeroko zakrojone wywiady i obserwacja uczestnicząca we wszystkich instytucjach zaangażowanych we wdrażanie interwencji obudziła świadomość istnienia programu, uruchomiła poczucie współodpowiedzialności i współzależności między instytucjami dotychczas działającymi w przekonaniu o swej niezależności. Co więcej, badania ankietowe potencjalnych beneficjentów okazały się *de facto* promocją programu i zaktywizowały szerszą rzeszę interesariuszy (potencjalnych beneficjentów).

Z przeprowadzonego studium przypadku płyną następujące wnioski na temat **czynników, które wpłynęły na wykorzystanie ewaluacji**.

Kluczowym dla ograniczonego, technicznego wykorzystania ewaluacji był **stan programu** poddanego ewaluacji. Jego słabe zaawansowanie (a właściwie faza początkowa) skierowały uwagę badających na jedyną rzecz, którą można było badać – procedury i procesy. To sprawiło też, że wąską grupą odbiorców byli operatorzy programu – instytucje działające na najwyższym poziomie (bo tylko na poziomie programu toczyły się działania). Opóźnienie we wdrażaniu programu sprawiało też, że ewaluacja i postawione przez nią pytania okazały się w dużej mierze przedwczesne, zaś rekomendacje były mocno techniczne i ograniczone – dotyczyły embrionalnego, ale szybko zmieniającego się systemu. Natomiast skomplikowany układ instytucjonalny programu, ograniczona skala finansowa (a co za tym idzie mniejsze znaczenie polityczne) czy delikatna sytuacja polityczna (dogasający wieloletni konflikt w Irlandii Północnej) okazały się nie mieć wpływu na wykorzystanie ewaluacji.

Rzeczywiście **interaktywność procesu** rozumiana w tym przypadku jako regularne i formalne kontakty między zlecającym (i szerszą grupą interesariuszy nadzorujących badanie) a ewaluatorami, przełożyła się na szybkie instrumentalne użycie (przede wszystkim wdrożenie rekomendacji). Zaś interaktywność procesu badawczego (*de facto* zastosowane podejście badawcze) rozumiana jako użycie wywiadów, obserwacji uczestniczącej i paneli dyskusyjnych, umożliwiła nawiązanie kontaktów i współzależności między interesariuszami programu.

Rola ewaluatora, ograniczona do konsultanta, nie miała znaczącego wpływu na użycie ewaluacji. Ważnym natomiast okazała się **aktywność zlecających** (Grupy Sterującej Ewaluacją, ale przede wszystkim SEUPB – Instytucji Zarządzającego i Sekretariatu w jednym), który włączany w badanie przez ewaluatorów przyjął w dialogu aktywną rolę (bieżące komentarze postępów, informacje o uwarunkowaniach programu i ograniczeniach rekomendacji), a potem przyjął rolę promotora wyników.

W tym studium przypadku cechy produktów badania ewaluacyjnego (nawiązane kontakty, jakość raportu, sposób komunikacji wyników, etc.) nie mia-

ły wpływu na wykorzystanie przez wąską grupę operatorów programu. Wynikało to ze stanu programu. Rodzaj wiedzy zawartej w raporcie (techniczna wiedza na temat blokad, problemów i potrzebnych działań – *know-what & know-who*), jak i sposób jej komunikacji był konsekwencją tego stanu.

4.4.3. Przypadek 3: IIIB–Europa Północno-Zachodnia

Analiza wkładu w ewaluację

Warunki kontekstowe [segment W1 modelu]

Program INTERREG IIIB Europa Północno-Zachodnia (*North West Europe* – NWE) był największym programem INTERREG w latach 2000–2006 (655 mln euro, z czego 326 mln finansowane z budżetu UE). Swoim działaniem obejmował bardzo szeroki obszar – regiony Irlandii, Wielkiej Brytanii, Belgii, Francji, Holandii, Luksemburga, Niemiec, a także państwo spoza UE – Szwajcarię. W odróżnieniu od kilku innych programów komponentu B, operujących na regionach geograficznych o ukształtowanej tożsamości i długiej historii współpracy (np. IIIB-Region Morza Bałtyckiego czy IIIB-Przestrzeń Alpejska), NWE został stworzony sztucznie, bez wcześniejszej tradycji współpracy regionów tego terytorium. Dopiero pieniądze UE (transferowane od 1994 roku) stworzyły wspólnotę interesów i dały bodziec do współpracy[32].

Program IIIB NWE (North West Europe) był homogeniczny pod kątem kultury administracyjnej – skupiał bowiem państwa o tradycji anglosaskiej i kontynentalnej (uznawane za efektywne i nowoczesne administracje), regiony o podobnym, wysokim poziomie rozwoju[33], dobrej historii współpracy (jednak tylko współpracy bilateralnej) (porównaj: Bachtler et al., 2005: 27). Wszystkie te administracje miały też rozwiniętą kulturę użycia ewaluacji w administracji publicznej (choć nieco odmienne podejścia).

Z uwagi na wielkość budżetu i zasięg geograficzny obejmujący najbogatsze regiony Europy, waga polityczna programu była znacząca. Dawał on też państwom – płatnikom netto – możliwość korzystania ze wsparcia budżetu UE. Należy jednak zaznaczyć, że dla poszczególnych administracji krajowych ranga tego programu była różna, różny był też poziom zainteresowania udziałem (np. przedstawiciele Irlandii bardzo angażowali się w program, widząc w nim kolejną możliwość podkreślenia swojej obecności w UE, Francja zaś skutecznie lobbowała za zlokalizowaniem instytucji programu na swoim terytorium).

Tak jak wszystkie programy komponentu IIIB, NWE skoncentrował się na planowaniu przestrzennym – rozwinięciu wspólnej dla tego szerokiego regionu geograficznego strategii zrównoważonego rozwoju. Najważniejszymi priorytetami programu (z perspektywy wielkości alokacji finansowej) były P1: spójny system miast i regionów (planowanie obszarów metropolitarnych,

[32] NWE IIIB był bezpośrednią kontynuacją programu INTERREG IIC North West Metropolitan Area działającego na tym obszarze w latach 1994–1999.

[33] Obszar NWE w zasadzie pokrywa się z tradycyjnym centrum wzrostu Europy, siecią najbardziej dynamicznych regionów UE (*blue banana*).

policentryczny rozwój przestrzenny) oraz P3: zarządzanie zasobami wodnymi i bezpieczeństwo przeciwpowodziowe. Pozostałe tematy współpracy dotyczyły poprawy dostępności (mobilność, społeczeństwo informacyjne), ochrony zasobów przyrodniczych i kulturowych, integracji terytorialnej obszarów nadmorskich (NWE Joint Technical Secretariat, 2002). Specyfiką IIIB NWE było uwypuklenie działań na rzecz budowania nowoczesnej kultury i wspólnej wizji planowania przestrzennego tzw. *New planning culture & new strategic vision* – wpisane jako działanie horyzontalne programu (NWE Joint Technical Secretariat, 2002: 26).

W momencie podejmowania badania ewaluacyjnego program był słabo zaawansowany (co prawda 54% funduszy UE zostało rozdysponowanych, ale rzeczywiste płatności były na poziomie 0,40%)[34]. To sprawiało, że program wciąż był objęty ryzykiem reguły N + 2[35]. W praktyce w tamtym okresie realizowano czterdzieści dwa projekty (głównie zorientowane na studia i działania – np. plany biznesowe, strategie rozwoju – rzadko infrastrukturalne). Mała liczba projektów i ich słaby poziom zaawansowania w oczywisty sposób ograniczyły przy ewaluacji analizę pytań dotyczących efektów programu (gdyż takie jeszcze się nie pojawiły).

Z uwagi na bardzo szeroki obszar programu, jego system instytucjonalny (a co za tym idzie sieć interesariuszy) był rozbudowany. Typowymi dla innych programów INTERREG były instytucje takie jak Komitet Monitorujący (KM – złożony z przedstawicieli narodowych – szczebla centralnego i regionalnego), Komitet Sterujący (wybierający projekty), Wspólny Sekretariat Techniczny (WST), Instytucja Zarządzająca (IZ), Instytucja Płatnicza (IP). Trzy ostatnie instytucje, działające na stałej zasadzie, były zlokalizowane we Francji. Dodatkowymi instytucjami były krajowe punkty kontaktowe, Ciało Nadzorujące (ang. *Supervising Body* – węższa grupa członków KM nadzorująca na co dzień wdrażanie przez Sekretariat i IZ zaleceń KM) oraz Grupa Robocza ds. Wizji Przestrzennej (złożona z naukowców, analityków, polityków i prowadząca dyskusje nad nową strategiczną wizją zrównoważonego rozwoju przestrzennego tej części Europy) (INTERACT, 2005e). W sumie, wliczając w to instytucje zaangażowane w ówczesne czterdzieści dwa projekty, szeroka grupa interesariuszy programu i ewaluacji wynosiła około osiemset osób.

Z obecnej perspektywy trudno w sposób jednoznaczny zidentyfikować ewentualne konflikty i ówczesną presję reform. Niewątpliwie jednak ryzyko N + 2 tworzyło napięcie. Ponadto wybór projektów – podejmowany na Komitecie Sterującym – miał bardzo silne zabarwienie polityczne (właściwa dystrybucja geograficzna, nieformalne parytety narodowe), co w sposób oczywisty mogło rodzić i rodziło konflikty (nawet personalne – natknęli się na nie ewaluatorzy).

---

[34] Ten stan wynikał głównie z faktu, że program rozpoczął się dopiero w 2002 roku (proces przygotowania i późniejszych negocjacji CIP z Komisją Europejską wydłużył się).

[35] Zasada N+2 zakłada, że pieniądze UE przypisane danemu programowi powinny zostać wydatkowane i ostatecznie rozliczone w ciągu maksymalnie dwóch lat od przypisania środków (a więc w tym wypadku przekazane beneficjentom najpóźniej do 2006 roku, a rozliczone w 2008 roku). W przeciwnym razie pieniądze muszą zostać zwrócone do budżetu UE.

Wkład uczestników [segment W2 modelu]

Oczekiwania zlecających badanie nie odbiegały od wytycznych sformułowanych przez Komisję Europejską. Głównym celem było spełnienie wymogu KE. Ponadto zarządzający programem traktowali ewaluację średniookresową jako „okazję do swoistej autorefleksji refleksji nad skutecznością procedur i działań" (*opportunity to self-reflect*) i oczekiwali od ewaluatorów „świeżej, zewnętrznej i obiektywnej opinii na temat działania systemu". Od początku traktowali też ewaluację nie jako okazję do promocji, nagłośnienia spraw programu, ale jako wewnętrzne narzędzie („not really externalized").

Od strony zlecającego kwestiami technicznymi zajął się personel Sekretariatu (osoby o dużym wcześniejszym doświadczeniu w programie i w tematyce planowania przestrzennego). Specjalnie dla potrzeb procesu ewaluacyjnego stworzono Grupę Sterującą Ewaluacją (ang. *Evaluation Steering Group – ESG*) złożoną z przedstawicieli Instytucji Zarządzającej, Wspólnego Sekretariatu Technicznego, państw członkowskich. Była to główna platforma dyskusji merytorycznych między zlecającym a wykonawcą. Tym samym sformalizowano i wymuszono ciągłość kontaktów.

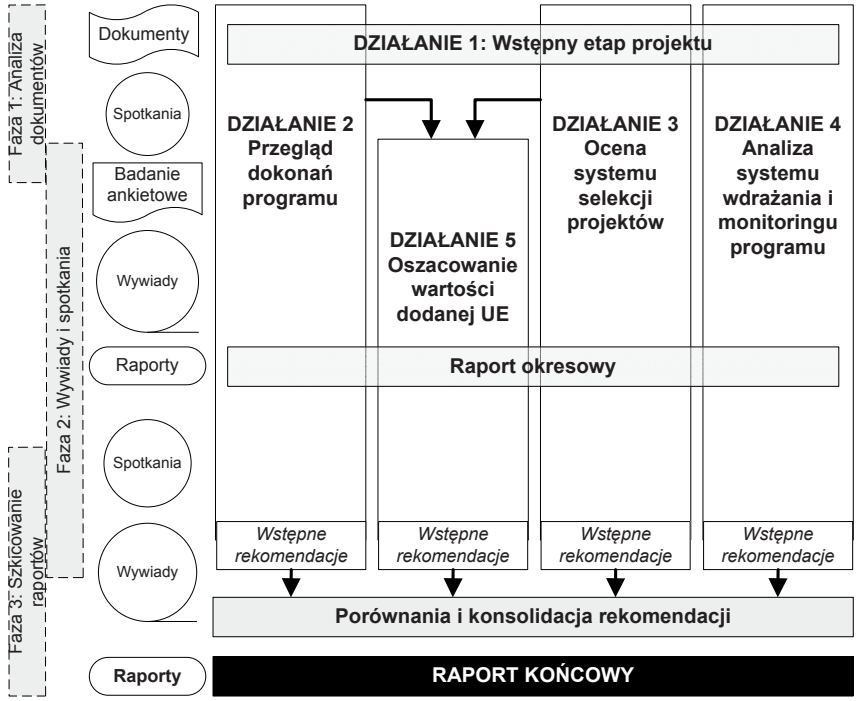
ToR (*Terms of Reference* – specyfikacja zamówienia) w całości pokrywał się z wytycznymi Komisji Europejskiej. Specyfiką było sformalizowanie nadzoru nad ewaluacją w postaci ESG oraz wymóg, by rekomendacje w raporcie końcowym miały określoną perspektywę czasową i koszty realizacji. Przy tworzeniu ToR pojawiły się wśród interesariuszy jedynie dyskusje natury technicznej (problemy z dostosowaniem przetargu do prawa francuskiego). Dokładny budżet badania nie odbiegał wielkością od średnich kontraktów badawczych UE. Czas badania określono na sześć miesięcy.

W przetargu wyłoniono wykonawcę – konsorcjum firm konsultingowych (z pewnym wsparciem akademickim, na przykład w polu wizji planowania przestrzennego). Liderem była nieduża (ośmio–dziesięcioosobowa) firma – Rupprecht Consult. Miała ona szerokie, wcześniejsze doświadczenia w ewaluacjach funduszy strukturalnych, ale żadnych doświadczeń w programach INTERREG. Paradoksalnie, zlecający uznał to za atut – gwarancję świeżego spojrzenia na program.

Analiza procesu ewaluacji

Organizację procesu badawczego, wraz z zarysem i sekwencją użytych metod, przedstawiono na poniższym schemacie.

Schemat 9. Organizacja badania ewaluacyjnego *mid-term* IIIB NWE



Źródło: Rupprecht Consult (2003: 10).

Strukturalizacja [segment D1 modelu]

Podejście badawcze zaproponowane przez ewaluatorów było połączeniem ewaluacji partycypatywnej z ewaluacją zorientowaną na zarządzanie. Na tą drugą wskazywał nacisk położony na badanie procedur i procesów oraz fakt, że głównymi adresatami i dyskutantami były osoby z personelu zarządzającego interwencją. Zaś o partycypatywnym charakterze badania świadczyło wybranie interaktywnych metod badań (grupy fokusowe, panele dyskusyjne, ankiety – słowem zbieranie opinii od szerokiej grupy interesariuszy) oraz ciągła dyskusja ewaluatorów z ESG.

Relacje między ewaluatorem a zlecającymi w fazie strukturalizacji były wysoce interaktywne. Ewaluatorzy zaproponowali więcej niż przewidywał ToR, metodologia była pogłębiona, a zakres badania szeroko zarysowany. Na forum ESG (Grupy Sterującej Ewaluacją) dyskutowano zakres zadań i – dostosowując się do sytuacji programu (brak widocznych efektów) – wybrano pytania priorytetowe dotyczące przede wszystkim procesów i procedur. Mimo dyskusji z ESG, to ewaluatorzy podejmowali ostateczne decyzje co do metodologii (np. skład grup fokusowych – odmienny od sugestii wysuwanych przez ESG).

Interesujący był też początek relacji między operatorami programu a ewaluatorami. W tej pierwszej fazie ewaluatorzy byli odbiorcami wiedzy – chłonę-

li *know-how* operatorów programu – zadawali pytania i „uczyli się programu" (obydwie strony silnie podkreślały szczególną intensywność tego procesu „pytań i odpowiedzi", obrazowały to też liczby spotkań i intensywność wymiany e-mailowej). W zrozumieniu specyfiki programu pomagała im, prócz wyjaśnień WST, obserwacja uczestnicząca (udział w spotkaniach Sekretariatu, Komitetu Monitorującego).

Obserwacja i analiza [segment D2 modelu]

W fazie obserwacji pojawił się problem danych wtórnych (z monitoringu). Wynikał z dwóch kwestii – niedopracowanego systemu wskaźników oraz słabego zaawansowania projektów (a więc braku widocznych efektów, które można by mierzyć). Badanie koncentrowało się więc na zbieraniu opinii.

Wybrane metody były wysoce interaktywne i o wyważonej mieszance podejścia ilościowo-jakościowego. Wstępny przegląd dokumentów był uzupełniany konsultacjami z operatorami programu. Przy metodach jakościowych sięgnięto po wywiady z przedstawicielami głównych instytucji programu (i najważniejszymi interesariuszami europejskimi), po warsztaty i grupy fokusowe. Użyto również metod ilościowych (szerokie badanie ankietowe na całej populacji interesariuszy).

Ewaluatorzy w tej fazie odegrali przede wszystkim rolę animatorów (spotkań, dialogu), a wnosząc nową perspektywę w prowadzone dyskusje do pewnego stopnia również edukatorów. Pewne napięcia między instytucjami i osobami wymagały również podjęcia roli mediatora, choć w zasadzie z perspektywy ewaluatorów ograniczało się to do dawania impulsu do dyskusji – nowych pomysłów i organizowania przestrzeni dyskusji. Należy podkreślić, że ewaluatorzy odegrali wyjątkowo aktywną rolę – interpretując rozszerzająco wymagania ToR (robiąc dużo ponad minimum), wykazując się własną inicjatywą oraz nie ograniczając się do powierzchownych badań, ale wchodząc w szczegółowe kwestie programu.

Relacje w tej kolejnej fazie ewaluacji były wysoce interaktywne (zważywszy na skalę prowadzonych spotkań i ich liczbę), z zaznaczeniem szerokich interakcji. Dowodem na to były przeprowadzone wywiady z przedstawicielami Komisji, programu ESPON (program-parasol zajmujący się tematyką planowania przestrzennego w UE), jak również bardzo ciekawa inicjatywa – konsultacje z ewaluatorami trzech innych programów IIIB (Region Morza Północnego, Region Bałtyku, Przestrzeń Alpejska) na temat metod, wspólnych problemów, wstępnych rekomendacji. W ramach programu interakcje odbywały się głównie w trójkącie Grupa Sterująca Ewaluacją – Sekretariat Programu – badacze. Miały one charakter ciągły (prócz oficjalnych trzech–czterech spotkań twarzą w twarz także kontakty robocze, wymiana informacji).

Ocena [segment D3 modelu]

Strategia zastosowana w fazie oceny była interaktywna – opierała się na silnym włączeniu interesariuszy w ocenę i wypracowywanie rekomendacji. Mimo

iż był to proces dwukierunkowy, to jednak ostateczny głos należał do badaczy. Ocenie poszczególnych aspektów programu posłużyły warsztaty i grupy fokusowe z przedstawicielami interesariuszy (dwadzieścia sześć osób z KM, KS, punktów kontaktowych, najbardziej aktywnych przedstawicieli beneficjentów, Grupy ds. Wizji Przestrzennej oraz, co ciekawe, ewaluatorów *ex ante* programu). Elementy oceny były też wkomponowane w badania ankietowe. Testowanie racjonalności wstępnych rekomendacji odbyło się natomiast w wąskim gronie Grupy Sterującej Ewaluacją (regularna wymiana opinii i komentarz do wstępnej wersji raportu). Warto jednak zaznaczyć, że poszukiwanie inspiracji i porównanie kierunków rekomendacji miało miejsce również na poziomie europejskim – w konsultacjach z zespołami ewaluacyjnymi innych programów INTERREG IIIB.

Na tym etapie badania ewaluatorzy spełnili rolę edukatorów – zbierając opinie, organizując dyskusje, wnosząc zewnętrzną perspektywę (doświadczenia innych programów) oraz prezentując syntetyczne opinie szerszej grupy interesariuszy – aplikujących o projekty. Ponadto w dwóch kwestiach sprowokowali żywiołową dyskusję wśród głównych interesariuszy programu (podsycaną narodowymi partykularyzmami), dotyczącą wizji planowania przestrzennego oraz kwestii nowego podejścia do oceny projektów. Nie podjęli tu jednak roli mediatorów, a ich rola jako animatorów była bardzo ograniczona (spotkania prowadzili przedstawiciele WST, oni też poszukiwali konsensusu w dyskusji). Z perspektywy późniejszych zmian w programie należy stwierdzić, że była to jednak rola reformatorska.

Analiza procesu tej ewaluacji pokazała też dwie dodatkowe kwestie nieobecne w modelu teoretycznym. Pierwszą była postawa zlecającego. Jego gotowość do wprowadzenia ewaluatora w specyfikę programu (cierpliwość w prezentowaniu informacji, wyjaśnianiu kontekstu proceduralno-instytucjonalnego) była ważnym czynnikiem płynnych relacji i interakcji. Kwestią drugą były sprawy techniczne w egzekucji kontraktu ewaluacyjnego. Poważne opóźnienia w płatnościach powodowały niepokój i frustrację po stronie konsultantów. Takie czysto techniczne, wydawałoby się drugorzędne kwestie, mogą mieć jednak znaczący wpływ na przebieg badania (np. opóźnienia w płatnościach mogą spowodować ograniczenie skali badania – zmniejszenie liczby wizyt studyjnych, spotkań, skali badania ankietowego). Ponadto działają demobilizująco na wykonawcę. Wreszcie takie trywialne, techniczne kwestie podważają klimat zaufania i w połączeniu z napiętym harmonogramem badania wprowadzają napiętą atmosferę na linii wykonawca-zlecający.

Analiza produktów ewaluacji

Stworzone kontakty [segment P1 modelu]

Ewaluatorzy stworzyli dobre kontakty ze zlecającymi – głównym narzędziem budowy zaufania były konsultacje w ramach ESG. Ten regularny kontakt i bazowanie na ocenach szerszej grupy interesariuszy stworzyły dobry grunt pod prezentację rekomendacji. Trudno jednak uznać, by kontakty te były na tyle silne, aby kluczowi interesariusze programu (z ESG) utożsamiali się

z rekomendacjami i w środowisku programu je promowali. Co więcej, należy podkreślić, że już w trakcie badania i podczas dyskusji wokół rekomendacji silnie dawały o sobie znać interesy narodowe. Ewaluatorzy odnotowali ten fakt i prowadzili dialog między grupami, nie koncentrowali się jednak na wypracowywaniu konsensusu.

Natomiast w odniesieniu do szerszej grupy interesariuszy zbudowane kontakty można scharakteryzować jako ograniczoną sieć (beneficjenci projektów wiedzieli o wykonanym badaniu ewaluacyjnym, ale nie brali udziału w dyskusji na temat wyników, nie byli zainteresowani promowaniem rekomendacji, choć z czasem uzyskali dostęp do raportu).

Dzięki ewaluatorom pogłębiono kontakty poza programem – z innymi programami IIIB oraz programem ESPON. Ten kontakt był raczej przejawem, a nie impulsem relacji rozwijanych przez program z szerszym środowiskiem europejskim.

Wiedza zawarta w raporcie [segment P3 modelu]

Raport, w ocenie samych operatorów programu, nie przyniósł większych zaskoczeń. Nie było w nim zupełnie nowych, wcześniej niebranych pod uwagę informacji o kontekście programu, wyzwaniach czy też zupełnie nowej perspektywie spojrzenia na program (ograniczone *know-about*).

Raport dał konkretną merytoryczną wiedzę na temat tego, co działa, a co nie działa w programie (*know-what*), a także wyjaśnienia, jakie są tego przyczyny (*know-why*). Nie były to informacje zupełnie zaskakujące dla operatorów. Były „raczej potwierdzeniem niż objawieniem" (*confirming rather than revealing*).

W raporcie znalazły się – prócz bardzo konkretnych technicznych informacji dotyczących tego, jak i kiedy wdrażać działania naprawcze (*know-how & know-when*) – ograniczone informacje o źródłach inspiracji. Zostały przytoczone dobre praktyki i pomysły z dwóch innych programów INTERREG IIIB (porównaj rekomendacja 7.3; Rupprecht Consult, 2003: 87). Sugestie te zostały wykorzystanie w kolejnym roku wdrażania pogramu.

Jeśli chodzi o rodzaje wiedzy, których dotyczył raport, należy zaznaczyć, że informacje na temat tego, kto ma wdrażać dane działania naprawcze (*know-who*), były bardzo konkretne. Każda z proponowanych rekomendacji została opatrzona szczegółową listą instytucji i aktorów odpowiedzialnych za jej rozważenie i ewentualne wdrożenie.

Rozpowszechnianie wyników [segment P4 modelu]

Końcową wersję raportu ewaluatorzy przekazali zlecającemu. Ich rola na tym etapie była pasywna – dostarczenie oczekiwanego produktu w ramach kontraktu (brak konferencji, sugerowania alternatywnych kanałów prezentacji wyników etc.) – mimo faktu, iż w formie raportu przewidzieli szerszych odbiorców (we wstępie przedstawiono mapę dla czytających, sugerującą odniesienie do różnych partii materiału w zależności od odbiorcy – KE, operatorów

projektów, narodowych przedstawicieli, instytucji zarządzającej programem). Niemniej warto zauważyć, że ewaluator rozpowszechniał raport jako dobrą praktykę w ramach swojej sieci kontaktów, prezentował go też na swojej stronie internetowej (raczej w celu autopromocji).

Ze strony zlecającego nie było specjalnej strategii komunikacji wyników. Odbiorcami raportu była wąska grupa głównych interesariuszy programu (instytucje zarządzające, przedstawiciele państw członkowskich), a wyniki rozpowszechniano międzyinstytucjonalnymi kanałami komunikacji. Reakcje na raport również były ograniczone (ze strony kilku państw członkowskich) – w formie korespondencji wewnętrznej. Raport opublikowano na stronie programu. Drogą elektroniczną dotarł on również do operatorów projektów i z czasem był rozpowszechniany już niezależnie – ich sieciami projektowymi (odwołania na stronach internetowych).

Analiza użycia ewaluacji

Użycie ewaluacji [segment U1 modelu]

W świetle informacji o wdrożonych rekomendacjach można uznać, że odbiorcy ewaluacji – głównie Komitet Monitorujący i Sekretariat – przyswoili informacje z raportu. Trudno jest określić w sposób obiektywny, na ile była to dla nich nowa wiedza. Z ocen zlecającego (trudno je uznać za reprezentatywne) wynika, że nie odbiegała ona w sposób znaczący od znanego już obrazu programu i od tego, co było wiadomo o jego problemach. Warto jednak pamiętać, że główni aktorzy byli na bieżąco, już w czasie ewaluacji, informowani o spostrzeżeniach. Element zaskoczenia przy raporcie końcowym był więc minimalny.

Ewaluatorzy niewątpliwie zdobyli nową, praktyczną wiedzę i doświadczenie (z zakresu prowadzenia ewaluacji partycypatywnej dla programów o dużej, zróżnicowanej grupie interesariuszy). Przyznali też, że była to okazja do dopracowania budowanego w firmie *know-how* (podejście wykorzystane w analizowanym badaniu opierało się na ich wcześniejszych doświadczeniach).

Użycie instrumentalne było stosunkowo łatwe do zidentyfikowania, gdyż uaktualnienie ewaluacji średniookresowej brało za punkt odniesienia rekomendacje z analizowanego badania, sprawdzając skalę ich wdrożenia i wyjaśniając ewentualne braki zmian (Research Voor Beleid International, 2005: 17–22). W perspektywie dwuletniej wdrożono połowę z zaproponowanych rekomendacji (głównie kwestie promocji, komunikacji, synergii między projektami i priorytetami). Część rekomendacji określonych jako konieczne do wdrożenia natychmiast (dotyczące zmian w strategicznej orientacji programu i dostosowań kompetencji międzyinstytucjonalnych) nie została wdrożona w krótkiej perspektywie, gdyż operatorzy uznali, że tak radykalne zmiany w trwającym dopiero rok programie mogłyby zaburzyć jego, i tak mocno opóźnione, działanie. Ten brak działania może być tłumaczony również ciążeniem instytucjonalnym i grą interesów (zmiany były znaczące i wymagałyby nowych negocjacji politycznych). Rekomendacje te częściowo wzięto pod uwagę przy projektowaniu kolejnej interwencji, w nowym układzie sił. Najbardziej dyskusyjne

i „delikatne" rekomendacje (mniej polityczny sposób wyboru projektów, zasada większościowego działania Komitetu Sterujacego) nie zostały uwzględnione. Część tematów była dalej pogłębiana w kolejnym badaniu (np. kwestia wskaźników). Wreszcie niektóre rekomendacje okazały się zbyt ambitne dla budżetu programu (pomysł stworzenia obserwatorium statystycznego dla obszaru programu).

Nie znaleziono żadnych dowodów, by badanie w znaczący sposób poszerzyło współpracę i kontakty interesariuszy (choć można zakładać, że rozmowy twarzą w twarz i warsztaty przyczyniły się do nawiązania kontaktów np. między operatorami projektów). Nie ma też dowodów na to, że nawiązane kontakty zewnętrzne zostały podtrzymane.

Warto zaznaczyć, że w wyniku ewaluacji pojawił się specyficzny rodzaj wiedzy (nie uwzględniony *explicite* w modelu), który zlecający wykorzystali w sposób praktyczny. Chodzi o doświadczenie w zlecaniu i nadzorowaniu badań ewaluacyjnych. Wiedza dotyczyła zarówno kwestii czysto technicznych (konstruowanie umowy, strona finansowa badania), jak i użytecznej dla zlecającego formy (interaktywne podejście spodobało się operatorom programu). Te doświadczenia zostały użyte przy kolejnej ewaluacji (*mid-term update*) – zlecający wymagał podejścia analogicznego (i takie zostało zastosowane).

Wykorzystanie ewaluacji [segment U2 modelu]

Część z rekomendacji (np. segment dotyczący roli punktów kontaktowych) została wykorzystana przy konstruowaniu INTERREG IVB NWE. Podobnie uwagi na temat systemu wyboru projektów oraz podziału kompetencji – usprawnień proceduralnych w informacji, promocji programu oraz generacji projektów – były używane przy nowej edycji programu na lata 2007–2013.

Ślady wykorzystania instrumentalnego raportu pojawiły się też w indeksach cytowań zewnętrznych publikacji. Raport średniookresowy był używany jako źródło informacji lub podawany jako przykład dobrych praktyk. Przykładem był zarówno sam raport ewaluacyjny – jego forma, podejście badawcze, jak i praktyki programu – praktyka informacyjna, praktyka kontaktów z beneficjantami, procedury rozliczeń (Bachtler et al., 2005; INTERACT, 2005e). Raport pojawił się też w pobocznych cytatach analiz zewnętrznych – regionów z obszaru objętego programem NWE. Kwestia problemów z rozpowszechnianiem wyników projektów pojawiła się w jednym z brytyjskich opracowań (Tym, Partners, 2005), a także na poziomie ogólnoeuropejskim – w studiach programu INTERACT (Fiorini et al., 2006; INTERACT, 2004). Do badania incydentalnie odwoływano się w kilku publikacjach naukowych – w kontekście relacji między ESDP a programami współpracy europejskiej (Stead, Waterhout, 2006) oraz w temacie europeizacji służby cywilnej w krajach UE (Steen, Schaap, 2004).

Ewaluatorzy, z uwagi na duże obciążenie kontraktami, nie starali się o kontrakt na uaktualnienie ewaluacji średniookresowej. Zdobyte doświadczenie wykorzystali w nowych studiach o nieco zbliżonym charakterze (duże programy o zróżnicowanej, międzynarodowej grupie beneficjentów), ale nie z zakresu współpracy terytorialnej.

Trudnym okazało się określenie w sposób jednoznaczny koncepcyjnego wykorzystania raportu w debatach dotyczących nowej polityki współpracy terytorialnej na lata 2007–2013. Był on na pewno jednym ze źródeł używanych przez interesariuszy INTERREG IIIB NWE zaangażowanych w projektowanie INTERREG IVB NWE. Były to jednak rozmowy natury technicznej, a nie dotyczące szerszej wizji. Z uwagi na fakt, że raport miał bardzo techniczny charakter i był adresowany głównie do operatorów badanego programu, należy wątpić w to, że wywołał zmiany w postrzeganiu współpracy terytorialnej.

Neutralne lub złe użycie [segmenty U3 i U4 modelu]

Mimo, iż większość omawianych w raporcie problemów była znana operatorom programu, to jednak dopiero w procesie ewaluacji wypracowano rozwiązania i przedyskutowano możliwe opcje. Rekomendacje były aktualne w momencie publikacji raportu, a część z nich zachowało swą aktualność nawet dla kolejnej edycji programu. Na tej podstawie należy stwierdzić, że raport nie został użyty w sposób symboliczny.

Raport nie był wyłącznie spełnieniem unijnego obowiązku. Z powyżej omówionych form użycia jasno wynika, iż miał praktyczną rolę dla zlecających go instytucji. Nie stwierdzono też manipulowania wnioskami. Raport w całości został udostępniony w Internecie, rekomendacje były opublikowane w krótkiej i dłuższej wersji – obydwie były przygotowane przez ewaluatora, bez ingerencji zlecającego.

Wnioski ze studium przypadku

Na podstawie przeprowadzonej analizy można stwierdzić, że **raport przyczynił się do poprawienia planowania interwencji publicznych** – choć nastąpiło to w przesunięciu czasowym – bardziej dla kontynuacji programu (na lata 2007–2013), niż w wersji programu poddanej ewaluacji. Przyczyną tego stanu był fakt słabego zaawansowania programu. Jego operatorzy nie chcieli zmieniać dopiero co rozpoczętego systemu w trakcie realizacji programu, bojąc się, że wywoła to dalsze opóźnienia. Nie bez znaczenia był tu też opór polityczny (swoiste ciążenie instytucjonalne). Postulowane zmiany w strukturze interwencji czy instytucjach łatwiej było uwzględnić w dyskusji nad nowym, tworzonym od podstaw programem.

Raport **spełnił funkcję poprawy zarządzania w wybranych polach wdrażania programu** (głównie promocji i komunikacji). Kontekst polityczny, ciążenie instytucjonalne i późny start programu ograniczyły skalę poprawek w sferze struktur instytucjonalnych i procedur. Pewne udoskonalenia z tego zakresu wprowadzono w kolejnym programie.

Funkcja wzmocnienia odpowiedzialności była realizowana w bardzo ograniczonym zakresie. Po pierwsze, wynikało to głównie ze słabego poziomu zaawansowania programu (ocena stopnia realizacji efektów była przedwczesna). Po drugie, techniczny, zorientowany na procesy charakter badania ograniczył grupę odbiorców przede wszystkim do operatorów programu. Kwe-

stia realizacji celów i racjonalności prowadzenia interwencji nie była w raporcie szerzej dyskutowana, a te kwestie, które podniósł zespół ewaluatorów, dotarły do wąskiego grona instytucji programu (traktujących to jako swoistą samoocenę, kontrolę stanu wdrażania) i Komisji Europejskiej. Skuteczność interwencji publicznej została dopiero nagłośniona pośrednio, za sprawą kolejnych badań porównawczych w skali europejskiej.

Raport **spełnił funkcję uczenia się głównie w zakresie budowania zdolności ewaluacyjnych** (raport i badanie promowane jako dobra praktyka w skali europejskiej, uczenie się na temat skutecznej ewaluacji samych uczestników procesu). Promował też w ograniczonym zakresie innowacyjne pomysły przenoszone z innych programów IIIB, w efekcie konsultacji z ewaluatorami tychże. Dla operatorów programu raport w małym stopniu przyniósł nową wiedzę o czynnikach warunkujących sukces i efekty interwencji. Dla aktorów zewnętrznych był jednak w tym temacie cennym źródłem informacji. Ewaluacja nie weryfikowała jednak, ani nie dawała wkładu w szersze teorie i założenia polityk współpracy terytorialnej.

Wydaje się, że **badanie przyczyniło się do wzmocnienia partnerstwa i poczucia współwłasności uczestników**. Trudno jednak ocenić skalę powiązań stworzonych wyłącznie dzięki badaniu (można raczej mówić o dodatkowym impulsie danym przez badanie, niż o wyłącznym jego wpływie). Badanie niewątpliwie zaktywizowało interesariuszy, choć najsilniej te instytucje, które już były obecne w procesie decyzyjnym (podniosło jednak jakość tych kontaktów).

Podsumowując **czynniki, które zdeterminowały wykorzystanie przeanalizowanego badania**, można przedstawić następujące wnioski.

Ważnym czynnikiem okazał się **stan programu (jego opóźnienie), a także skomplikowany układ instytucjonalny** – silna pozycja aktorów narodowych, wysoki stopień polityczności decyzji. W skali mikro ważnym okazała się **charakterystyka ewaluatora** jego wcześniejsze doświadczenia z ewaluacją partycypatywną, brak doświadczenia w programach INTERREG, zaś z drugiej strony przystępność i otwartość zlecającego, który w pierwszej fazie badania wprowadzał ewaluatora w meandry programu, a potem uczestnicząc w badaniu był promotorem wyników.

Interaktywność procesu (panele dyskusyjne, grupy fokusowe, warsztaty) oraz aktywna rola ewaluatora jako animatora i edukatora były wyraźnym czynnikiem wpływającym na użycie. Wysoka resposnywność zlecającego (jego otwartość na pytania, ciągłość dyskusji z ewaluatorem), a także sformalizowane, stałe relacje zlecającego i badaczy (w ramach Grupy Sterującej Ewaluacją) były dodatkowym czynnikiem.

W świetle tego studium przypadku niewątpliwie ważnym czynnikiem okazały się **nawiązane kontakty w relacji wykonawca-główni interesariusze**. Jakość i operacyjność zawartej wiedzy miała duże znaczenie przy prezentacji wyników, ale w praktyce wdrażania rekomendacji przeważyły kwestie szersze, kontekstowe, na przykład ciążenie instytucjonalne, kwestie polityczne, moment programu. Jakość ewaluacji (forma i komunikatywność raportu) i sposób komunikacji wyników (Internet jako medium) okazały się ważne przy ich szerszym wykorzystaniu przez aktorów poza programem.

5. WNIOSKI I REKOMENDACJE

Niniejszy rozdział prezentuje trzy grupy wniosków z przeprowadzonych analiz. Pierwszą są konkluzje dotyczące weryfikacji głównego pytania badawczego postawionego w niniejszej książce, czyli „jaki był wpływ badanych ewaluacji na poprawę zarządzania w sektorze publicznym". Drugą grupą są szersze wnioski, odnoszące obserwacje ze studiów przypadków do paradygmatów administracji i zarządzania sektorem publicznym. Wreszcie trzecią grupę stanowią wnioski praktyczne na temat dobrych praktyk, podejść, mechanizmów i procesów uczenia się, które można przenieść na grunt ewaluacji innych programów współpracy terytorialnej (i szerzej programów współfinansowanych przez UE). Część z tych spostrzeżeń jest adresowana do ewaluatorów, część zaś do osób i instytucji zlecających badania ewaluacyjne.

5.1. WNIOSKI Z BADANIA

Ewaluacje zrealizowały funkcję poprawy zarządzania…

…zaś w małym stopniu poprawy planowania.

Średniookresowe ewaluacje okazały się przydatne w zarządzaniu programami UE. Przyczyniły się do poprawy zarządzania, głównie w kwestiach technicznych (poprawa efektywności procedur i procesów bez wprowadzania znaczących zmian instytucjonalnych). Ich użytkownikami byli główni operatorzy programów. Użycie ewaluacji do kształtowania przyszłych programów (funkcja planowania) było bardzo ograniczone i wystąpiło tylko dzięki ciągłości instytucjonalnej i personalnej. Wyjaśnienia tego faktu należy upatrywać w dwóch kwestiach. Pierwszą jest zakres pytań ewaluacji średniookresowej, która z założenia koncentruje się na bieżącym zarządzaniu. Drugą kwestią okazał się późny start programów (w praktyce badanie nie w połowie, a w 1/3 wdrażania) i co za tym idzie brak ich namacalnych efektów w momencie podejmowania ewaluacji.

Ewaluacje wsparły uczenie się… ..ale głównie na temat kwestii technicznych.

Ewaluacje średniookresowe programów INTERREG wypełniły też – choć w różnym stopniu – inne funkcje ewaluacji (uczenia się, partnerstwa, odpowiedzialności). Funkcja uczenia się – poznawcza – została wsparta, ale odnosiła się tylko do niektórych rodzajów wiedzy. Główna wiedza płynąca z raportów dotyczyła tego, co działa (lub nie) w programie – a więc informacji technicznych (*know-what*). To zaś było logicznie powiązane z wnioska-

mi na temat działań zaradczych (*know-how*), ich terminów (*know-when*) i wykonawców (*know-who*). Zwykle nie była to wiedza nowa i zaskakująca dla operatorów programu. Potwierdzała i legitymizowała ich codzienne, pobieżne obserwacje. Nowa, interesująca dla użytkowników wiedza pojawiała się dopiero w momencie, gdy ewaluator pokazywał szerszy kontekst – na przykład wprowadzał porównania czy obserwacje z poziomu europejskiego (jak to było w przypadku programu INTERREG IIIB), albo wprowadzał nowe głosy i perspektywy patrzenia na program (np. przypadek irlandzki – punkt widzenia potencjalnych beneficjentów).

Nowa wiedza płynęła z porównań

Wkład do szerszych teorii funkcjonowania programów, refleksja nad podstawami polityki współpracy terytorialnej (*know-why*) praktycznie nie miały miejsca. Można to usprawiedliwiać technicznym charakterem studiów i koncentrowaniem się na palących opóźnieniach. Ten brak refleksji nad esencją programów INTERREG – mechanizmami współpracy ponad granicami – jest jednak zastanawiający. Programy potraktowano wyłącznie w kategoriach technicznych i nie sięgano po żadne koncepcje czy literaturę naukową. Jedno ze studiów (austriackie) dotknęło tego tematu, gdy po kilku miesiącach badań okazało się, że interesariusze odmiennie rozumieją „transgraniczność". To odkrycie nie pociągnęło jednak za sobą analizy głębszych przyczyn tego stanu rzeczy, a jedynie techniczne ćwiczenie skonstruowania wspólnej definicji „projektu ponadgranicznego". Szansę refleksji teoretycznej nad naturą współpracy europejskiej należy więc uznać za całkowicie niewykorzystaną.

Zabrakło jednak szerszej refleksji nad naturą współpracy terytorialnej UE

Funkcja „wzmocnienia partnerstwa i poczucia współwłasności uczestników" została spełniona we wszystkich trzech badaniach, choć w różnych proporcjach. Najsilniejsza była w badaniach austriackich – ich ciągła, trzyipółroczna forma zintensyfikowała dialog, współpracę i stałe kontakty między instytucjami programów. W przypadku programu irlandzkiego badanie obudziło świadomość istnienia programu i współzależności między aktorami. W przypadku INTERREG IIIB funkcja wzmacniania partnerstwa ograniczała się do intensyfikowania istniejących już relacji.

Ewaluacje w różnych proporcjach wsparły funkcję partnerstwa i współwłasności

Natomiast funkcja „wzmocnienia odpowiedzialności" (ang. *accountability*) została wypełniona w sposób symboliczny i pozorny w odniesieniu do Komisji Europejskiej, a nie szerszej opinii publicznej. Była to raczej kontrola – i do tego nieskuteczna. Raporty nie były w stanie poka-

Funkcja wzmacniania odpowiedzialności była pozorna i symboliczna

zać skuteczności i wydajności programów – gdyż te, będąc znacznie opóźnione (z przyczyn zewnętrznych – nie świadczących o stopniu efektywności personelu) nie przyniosły jeszcze konkretnych efektów. Sama Komisja, poza zgromadzeniem raportów – czysto administracyjnym ćwiczeniem, nie użyła ich w żaden widoczny sposób.

Przeprowadzone studia przypadków przynoszą także następujące wnioski odnośnie czynników warunkujących wykorzystanie ewaluacji.

Kluczowe czynniki dla wykorzystania ewaluacji: – stan programu i sytuacja polityczna

We wszystkich przypadkach ważnym dla użycia ewaluacji okazały się dwa czynniki kontekstowe: stan programów (ich wyjątkowe opóźnienia) i sytuacja polityczna programu (to, jaki miał priorytet w działaniach danych administracji). „Wielką politykę" szczególnie wyraźnie widać było w dwóch skrajnych przypadkach – austriackim (gdzie programy były kluczem kształtowania współpracy dobrosąsiedzkiej, a ewaluacja stała się głównym narzędziem integrowania nowych partnerów) oraz irlandzkim, gdzie od kwestii INTERREG-u uwagę odwracał dużo większy, strategiczny program PEACE.

– sposób podejścia badaczy do działania

Jeśli chodzi o wkład wnoszony przez uczestników, to ważnym było doświadczenie ewaluatorów i ich sposób podejścia do badania. Zorientowanie na potrzeby klienta, postrzeganie się w roli animatora dialogu (przypadek ewaluatora austriackiego), żywe uczestnictwo w procesach programu (np. obserwacja uczestnicząca w badaniu irlandzkim, warsztaty dla programu IIIB – NWE) okazały się nie do przecenienia.

– partycypatywne podejście

Wszystkie trzy przypadki dowiodły, że dla użycia ewaluacji średniookresowych, a co za tym idzie i bieżących (pokrywających analogiczne kwestie) kluczowy jest interaktywny proces. Ewaluacje partycypatywne ułatwiły szybką internalizację wniosków i rekomendacji (szczególnie dotyczących poprawy zarządzania i procedur) przez operatorów programu.

– rola ewaluatorów jako animatorów, doradców i edukatorów

Ewaluatorzy przyjmowali role aktywnego konsultanta (przypadek irlandzki), animatora dyskusji (przypadek IIIB North West Europe), a z czasem doradcy i edukatora (przypadek austriacki) budującego swój autorytet poprzez wprowadzanie głosów i informacji zewnętrznych. Interaktywność gwarantowała dobre użycie nawet w sytuacji, gdy badanie było początkowo postrzegane przez operatorów programu jako mało użyteczne (przypadek Irlandii). Dodatkowym wyjaśnieniem pozytywnej relacji między partycypatywnym podejściem a wysokim użyciem jest fakt, że użytkownicy włączani w proces oceny dużo szybciej chłonęli wiedzę i byli bar-

dziej otwarci na krytykę (gdyż nie była ona zewnętrzna i nie wychodziła na zewnątrz). Podejście partycypatywne wymagało jednak dużego zaufania w relacji ewaluator – operatorzy programu. Do tego potrzebny był czas oraz gwarancja, iż cały proces będzie wewnętrzny.

– zaufanie w relacjach ewaluatorzy – badan

Należy podkreślić, że relacje nie były jednak (jak zakładano na wstępie tej analizy) jednostronne. Ważnym okazał się „stopień responsywności interesariuszy", a przede wszystkim zaangażowanie i aktywność głównych instytucji (występujących jako promotorzy badania i dyscyplinujących inne instytucje do udziału). Dzięki temu swoistemu „mecenatowi", badania wykonywane były sprawnie i miały bezpośrednie przełożenie na rzeczywistość programów.

– zaangażowanie interesariuszy i „mecenat" zlecających

W przypadku analizowanych ewaluacji średniookresowych końcowa komunikacja wyników okazała się marginalnym czynnikiem tłumaczącym użycie. Ważna była raczej komunikacja bieżąca – w niej mieściły się również robocze dyskusje o wynikach programu, organizowane we wszystkich trzech ewaluacjach. Dzięki temu wiedza była przedstawiana i przyswajana na bieżąco, a końcowe wnioski nie były kontrowersyjne. Ewaluatorzy przygotowywali konsekwentnie grunt pod przyjęcie wniosków i proponowanych zmian, więc nie spotkały się one z ostrym oporem odbiorców.

Ważna rola bieżącej komunikacji wyników

Podsumowując wnioski oparte na modelu a płynące z analizy studiów przypadków, warto zwrócić uwagę na dwie kwestie.

Po pierwsze, wszystkie trzy przypadki jasno pokazały, że to nie sam raport ewaluacyjny, a cały proces ewaluacji był czynnikiem warunkującym wykorzystanie badania (przede wszystkim funkcje poprawy zarządzania, uczenia się, budowy partnerstwa). Raporty były tylko końcowym produktem partycypatywnego procesu badawczego. Uczestnicy czerpali i internalizowali wiedzę już w trakcie badania. Pozytywne zmiany wywołała nie sama treść raportów, ale poprzedzające ją działania i interakcje.

Proces badawczy jako źródło nauki

Po drugie, przeprowadzone badanie potwierdziło obecne w literaturze opinie, iż funkcja odpowiedzialności stoi w dużej mierze w sprzeczności z funkcjami poprawy jakości wdrażania i partnerstwa. Istnieje wyraźny wybór: „coś za coś" (*trade-off*). Ewaluacje średniookresowe, podobnie jak bieżące, koncentrują się zwykle na usprawnieniu zarządzania, na bieżącej pomocy uczestnikom programu, wzmocnieniu ich współpracy. Mają więc charakter operacyjny. Co za tym idzie, są skierowane „do wewnątrz

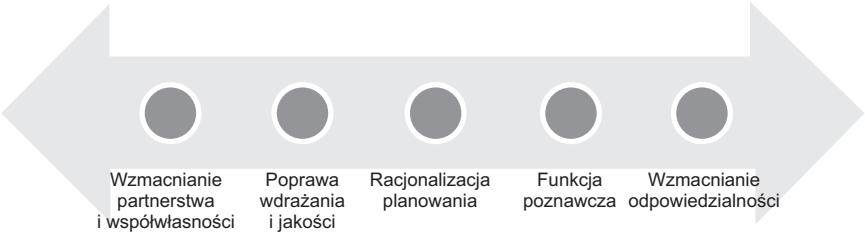
Konflikt funkcji odpowiedzialności z funkcjami poprawy wdrażania i partnerstwa

programu". Jak pokazały studia przypadków, do skutecznego wypełnienia tak sformułowanych celów konieczna była szczera dyskusja na temat problemów, prowadzona między interesariuszami. Klimat zaufania opierał się na małym dystansie między badanymi a ewaluatorami oraz na świadomości, że raport nie jest rozliczeniem, a jedynie roboczym dokumentem podsumowującym dyskusję.

Natomiast stawianie akcentu na odpowiedzialność kieruje uwagę na kwestie strategiczne, podsumowujące. Wnioski są adresowane „na zewnętrz programu" (pokazanie opinii publicznej skuteczności, efektów i sensowności założeń programu). Podobnie kwestie uczenia się – funkcja poznawcza – wymaga całościowej, pogłębionej analizy efektów. Nie można jej podjąć przy słabym stanie zaawansowania programu, w trakcie wdrażania, na podstawie pierwszych efektów. Jak pokazały przeprowadzone studia, takie próby kończą się fragmentarycznymi i przyczynkowymi wnioskami dotyczącymi tylko produktów, a nie oddziaływania.

Tę rozbieżność można podsumować w następujący sposób (patrz poniższy schemat)

Schemat 10. Dylematy wyboru funkcji ewaluacji



Źródło: opracowanie własne.

Wymóg pozornej kontroli blokował procesy naprawcze

W analizowanych przypadkach Komisja Europejska wymagała więc od ewaluacji średniookresowych pokrycia przeciwstawnych funkcji, w rezultacie ograniczając to, co było potrzebą chwili – pomoc we wdrażaniu opóźnionych programów. Co więcej, „odpowiedzialność" miała tak naprawdę charakter kontroli i to nie szerokiej – społecznej, a wąskiej – dokonywanaj przez Komisję Europejską. Była to też, jak się okazało, kontrola pozorna. Raporty z ewaluacji były zbierane, ale nie wykorzystywane, Komisja czerpała informacje o stanie programów z monitoringu finansowego. Z perspektywy uczestników programów świa-

domość tego kontrolnego aspektu badań blokowała ich w rozmowach na temat problemów. Wymóg Komisji nie był też dostosowany do faktycznego stanu programów.

Na tej podstawie można posunąć się do stwierdzenia, że zbyt wiele „odpowiedzialności", szczególnie egzekwowanej mechanicznie na wczesnym etapie rozwoju programu, może blokować jego wdrażanie. Można dalej spekulować, że tak silny nacisk na rozliczeniowy aspekt ewaluacji był elementem biurokratycznej „choroby kontroli", bądź próby przerzucania na innych aktorów – przede wszystkim przez Komisję – odpowiedzialności za ewentualne problemy (jak ujął to jeden z badanych: „blame transfering exercise").

Symptomy choroby kontroli?

5.2. SZERSZE WNIOSKI ODNOŚNIE DO ZARZĄDZANIA

Odnosząc analizowane przypadki do szerszych paradygmatów administracji i zarządzania sektorem publicznym należy stwierdzić, że we wszystkich trzech przypadkach rola ewaluacji średniookresowej wpisywała się w paradygmat menedżeryzmu. Identyfikowała potrzeby informacyjne głównych operatorów programu (głównych menedżerów) i dawała konkretne, techniczne rekomendacje dotyczące procesów i procedur. Wnioski z ewaluacji nie były odkrywcze, uwiarygodniały jednak i przedstawiały w sposób całościowy obserwacje menedżerów. Listy rekomendacji z ewaluacji były punktem odniesienia w kolejnych badaniach, jak również punktem odniesienia ich własnych działań.

Ewaluacje średniookresowe wpisały się w paradygmat menedżeryzmu

Studia przypadków pokazały w praktyce ograniczenia i wyzwania ewaluacji w menedżeryzmie. Znaczenie polityczne tych ewaluacji było znikome – nie budowały one szerszych wizji, nie weryfikowały (ani nie kwestionowały) strategii i założeń badanych programów. Lekcje przeniesione do programów kolejnej generacji miały charakter czysto techniczny. Wszystkie trzy przypadki (w sumie sześć ewaluacji) oceniały też produkty programów (liczbę projektów – a więc prosty wskaźnik efektywności), ale nie wychodziły już poza tę kwestię, ku analizie efektów czy sposobu podejścia do programów. Wreszcie trudną kwestią okazało się wypośrodkowanie między zorientowaniem na klienta (głównego zlecającego) i silnym włączaniem go w dyskusje w trakcie badania, a zachowaniem obiektywizmu i uniknięciem zawężania analizy wyłącznie do kwestii wskazanych przez główną instytucję programu.

Ograniczenia ewaluacji w menedżeryzmie:
– akcent na kwestie techniczne
– proste wskaźniki
– zorientowanie na dominującego aktora.

Mimo tego dominującego paradygmatu w poszczególnych przypadkach (szczególnie austriackich) można było

Elementy nowego paradygmatu – widoczne, lecz wciąż słabe…

również zauważyć elementy najnowszego paradygmatu (*New Public Governance* – NPG). Ewaluatorzy, mimo iż zorientowani na głównych operatorów programu, używali podejścia wielowymiarowego i systemowego (włączając w ocenę wiele grup i perspektyw). Silny był również element budowania sieci – nie tylko między bieżącymi aktorami programu, ale również z przyszłymi uczestnikami.

…i obarczone problemami metodologicznymi.

Pokonując ewaluację według nowego paradygmatu, ewaluatorzy nie uniknęli jednak problemów natury metodologicznej. Przy kompleksowym podejściu i napiętym harmonogramie używali głównie metod jakościowych i trudno im było utrzymać standardy warsztatu naukowego (np. reprezentatywność badanej grupy, stopę zwrotu ankiet, jasny klucz wyboru studiów przypadków etc.). Stosowany model analizy można było często określić wyrażeniem „szybkie i pobieżne" („quick and dirty approach").

Podsumowując, można uznać, że ewaluacja średniookresowa – czy analogiczne do niej ewaluacje bieżące – mają menedżerski charakter. Akcent jest stawiany na techniczną stronę procesów – szczególnie w słabo zaawansowanych programach. Elementy nowego podejścia do zarządzania sektorem publicznym (a co za tym idzie i roli ewaluacji), można wpleść poprzez ujęcie systemowe i włączanie szerokich grupy interesariuszy w cały proces badawczy.

Procesy uczenia się z ewaluacji są efektywne, gdy:

Przeprowadzone badanie zmusza również do odpowiedzi na pytanie o proces uczenia się instytucji publicznych oraz o rolę ekspertów zewnętrznych, jakimi są ewaluatorzy w tym procesie.

– są interaktywne

– istnieje ciągłość instytucjonalna

Analizowane w oparciu o model przypadki pokazały powszechną prawdę znaną z pola edukacji, iż włączanie uczestników we wspólną pracę aktywizuje ludzi i instytucje, wspomaga uczenie się i internalizowanie wiedzy. Jednak w przypadku instytucji warunkiem takiego budowania wiedzy i jej trwałości jest ciągłość instytucjonalna. Ewaluacje austriackie pokazały to najjaskrawiej – mimo wysokiej interaktywności procesu, doświadczenia ewaluatorów i dużego zaangażowania ze strony kluczowych instytucji wiedza na bieżąco „ginęła" po stronie nowych państw członkowskich (wysoka rotacja personelu). Rolę trwałości instytucji pokazały też pośrednio dwa pozostałe przypadki.

– instytucje mają sprecyzowane oczekiwania.

W procesie efektywnego uczenia się instytucji – odbiorców ewaluacji – nie bez znaczenia okazały się ich sprecyzowane oczekiwania. W dwóch przypadkach instytucje zlecające miały jasną wizję, do czego jest im potrzeb-

ne badanie. To umożliwiło im efektywne użycie ewaluacji do swoich potrzeb, mimo sprzecznych wymagań narzuconych przez Komisję Europejską.

Rozważając ostatnie zagadnienie – miejsce ewaluatorów w procesach uczenia się – należy stwierdzić, że ich rola w poznawaniu technicznej strony programów była bardzo ograniczona. Operatorzy programów dysponowali dużo dokładniejszą, ale i szerszą znajomością tematu (procedur, kompetencji, procesów). Znali też nieformalne uwarunkowania, które ewaluatorzy nie zawsze mieli okazję uchwycić. Wartość dodaną wniesioną do ewaluacji średniookresowej przez ekspertów można ująć w trzech punktach. Po pierwsze, ewaluatorzy myśleli i patrzyli na program w sposób świeży – nieobciążeni techniczną wiedzą o proceduralno-instytucjonalnych zawiłościach (tzw. myślenie *outside the box*). Zadawali często pytania o kwestie podstawowe i zasadnicze, które jednak zagubiły się w natłoku bieżących procedur. Po drugie, ewaluatorzy wprowadzali i łączyli w dyskusji nad programem różne, często nowe perspektywy (marginalizowane w codziennym zarządzaniu). Tym samym ewaluatorzy odgrywali rolę łączników między różnymi aktorami i dali instytucjom prowadzącym nowy sposób spojrzenia na znaną im rzeczywistość programu. Po trzecie, ewaluatorzy pokazali program w szerszym kontekście innych doświadczeń krajowych, sektorowych czy europejskich.

Wartości wniesione przez ewaluatorów: – świeże spojrzenie na program

– pokazanie nowych perspektyw i włączanie nowych głosów

– pokazanie kontekstu.

Podsumowując, wydaje się, że przy ewaluacji średniookresowej (i ewaluacjach bieżących zorientowanych na procesy zarządzania) najbardziej odpowiednią rolą ewaluatora jest rola animatora (dialogu między interesariuszami) i doradcy oraz rola edukatora – budowana w oparciu o wprowadzanie szerszej perspektywy (odniesień do innych programów, porównań czasowych, przykładów europejskich).

Rola ewaluatorów – animatorzy, doradcy i edukatorzy

5.3. SUGESTIE PRAKTYCZNE

Sugestie praktyczne zawarto w trzech grupach. Pierwszą są sugestie dotyczące wykorzystania metamodelu w praktyce. Pozostałe dwie grupy odnoszą się odpowiednio do zlecających badania i do ewaluatorów.

trzy rodzaje sugestii praktycznych

Poniższa tabela prezentuje możliwości wykorzystania metamodelu w praktyce. Skorygowany model przedstawiono na schemacie 6 w Aneksie 6.2.

1. Sugestie użycia modelu

Tabela 22. Możliwości praktycznego zastosowania metamodelu

Perspektywa instytucji zlecających	Perspektywa ewaluatorów
• sprecyzowanie potrzeb (celu ewaluacji, oczekiwanych efektów badania – funkcji); • całościowe określenie warunków brzegowych (czasu, budżetu, etc.) i realistyczne zaplanowanie kontraktu; • jasne określenie produktów badania (formy raportu, konkretnych rodzajów wiedzy, którą chcemy uzyskać); • zaplanowanie procesu wykorzystywania wyników (rozpowszechniania, śledzenia efektów raportu, jego użyteczności).	• stworzenie mapy oczekiwań klienta oraz ograniczeń kontekstowych badania; • świadomy wybór podejścia badawczego, metod i roli, jaką ewaluator będzie odgrywał w badaniu (swojej pozycji); • zaplanowanie produktów i pogrupowanie wniosków i rekomendacji wg rodzajów wytworzonej wiedzy; • ustawienie strategii badania i relacji w projekcie pod kątem pożądanych rodzajów wykorzystania badania.

Źródło: opracowanie własne.

2. Sugestie dla zlecających:

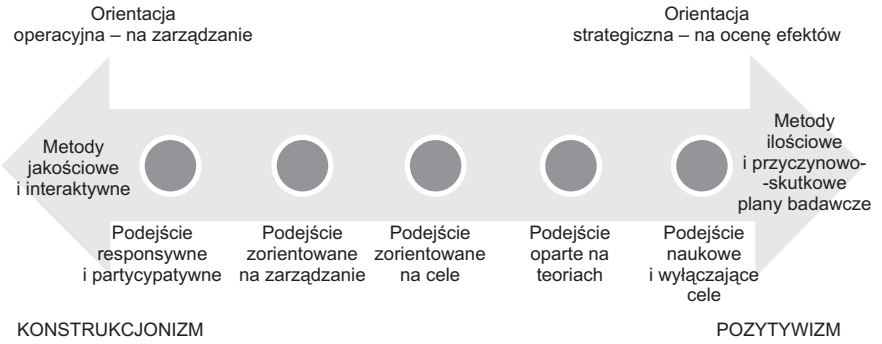
Opierając się na przeprowadzonych studiach przypadków można sformułować sześć szerszych **sugestii dla instytucji zlecających ewaluację**.

Świadomość wyboru „coś za coś"...

Po pierwsze, zlecający ewaluację powinni być świadomi wyboru „coś za coś" pomiędzy funkcjami, które spełniać może dane badanie i rozumieć powiązanie tych wyborów z typem czasowym ewaluacji. Dylemat ten został już przedstawiony na schemacie 10. Przypomnijmy, że jest to wybór pomiędzy orientacją na bieżące, operacyjne kwestie (zarządzanie, budowę partnerstwa) a refleksją strategiczną na temat rzeczywistego oddziaływania programu, natury mechanizmów, które wywołały bądź ograniczyły rzeczywiste efekty. Tę ostatnią podjąć można tylko wtedy, gdy program jest już bardzo zaawansowany we wdrażaniu, a najlepiej gdy został całkowicie zakończony – dopiero wtedy pojawiają się efekty.

Wybór funkcjonalnej orientacji badania determinuje jego dalszy kształt – możliwe podejścia badawcze, wachlarz metod i role ewaluatorów. Konsekwencje są jednoznaczne. Przedstawiono je na schemacie 11.

Schemat 11. Podejścia ewaluacyjne a orientacja badań



Źródło: opracowanie własne.

Przy orientacji operacyjnej ewaluatorzy sięgać będą (w idealnych warunkach, przy założeniu ich profesjonalizmu) po metody interaktywne i jakościowe, gdyż te umożliwią im interakcje, uchwycenie różnych głosów i perspektyw uczestników procesu wdrażania i kierowanie dyskusją. Ich filozofią działania będzie konstrukcjonizm. Przyjmować też będą rolę konsultantów i animatorów dialogu, z czasem wprowadzając nowe punkty widzenia i szersze porównania ewaluatorów. Natomiast przy orientacji na ocenę efektów filozoficzną podstawą będzie pozytywizm lub realizm. Z nim będą związane metody ilościowe i plany badawcze umożliwiające zidentyfikowanie relacji przyczynowo-skutkowych, znalezienie teorii najlepiej wyjaśniających sukces lub porażkę programu. Ewaluatorzy wcielą się więc w rolę naukowców i edukatorów (wobec szerszej publiki), jednak dla uczestników programu będą osobami kontrolującymi.

...pomiędzy orientacją operacyjną – na zarządzanie...

...a refleksją strategiczną, nad efektami.

Tak więc chęć pokrycia jednym badaniem efektów i procesów prowadzić będzie wprost do swoistej badawczej „schizofrenii", a w konsekwencji ograniczać skuteczność badania. Nie pomoże ono ani w procesach (interesariusze, czując się rozliczani, będą bać się szczerej dyskusji), ani tym bardziej w ocenie efektów (ta wymaga refleksji i czasochłonnych analiz związków przyczynowo-skutkowych wychodzących poza wywiady i samoocenę uczestników procesu wdrażania). Kazus ewaluacji średniookresowych wyraźnie to pokazał.

Po drugie, ewaluacje mające pomagać w zarządzaniu wymagają elastyczności czasowej i dostosowania się do sytuacji programu. Z perspektywy operatorów programu sprawdzają się badania „ciągłe" (przypadek austriacki), dające możliwość dostosowania do pojawiających się potrzeb informacyjnych oraz czas na zbudowanie zaufania.

Użyteczność ciągłej formy badań

W przypadku polskim zastosowanie takiego rozwiązania (kilkunastomiesięczne kontrakty ewaluacyjne, złożone z szeregu badań, na które trzeba się decydować *ad hoc*) może zwiększyć elastyczność reakcji na potrzeby informacyjne (głównie z zakresu zarządzania). Obecnie ta elastyczność została poważnie ograniczona przez roczne plany ewaluacyjne.

Po trzecie, przy rozpoczynaniu badania ewaluacyjnego trzeba jasno określić swoje potrzeby informacyjne. W przypadku programów współpracy terytorialnej *Zakres zadań* (ang. *Terns of Reference*) powinien uwzględniać potrzeby informacyjne aktorów ze wszystkich zaangażowanych krajów i traktować program całościowo. Należy

Jasno sprecyzowane oczekiwania

jednak podkreślić kwestię, o której zapominają instytucje polskie – postawione cele powinny być realistyczne, a zakres badania (tak pytania, jak i zakres przestrzenny) dostosowane do planowanego budżetu i danego czasu.

Przy poprawianiu zarządzania i partnerstwa sprawdza się dialog i mniej formalne raportowanie

Po czwarte, warunkiem wstępnym ewaluacji skoncentrowanej na funkcjach poprawy wdrażania i wzmacniania partnerstwa jest budowa zaufania i szeroki dialog. Kluczowe są dyskusja i nawiązane relacje między różnymi uczestnikami programu. Wybór wewnętrznego, mniej formalnego sposobu raportowania, połączony z interaktywnymi formami (warsztaty, panele dyskusyjne) daje komfort dyskusji i większą otwartość w rozmowach o problemach programu. Jest też bardziej konstruktywny przy poszukiwaniu rozwiązań. Dla programów współpracy terytorialnej ta interaktywność dodatkowo wzmacnia efekt współpracy i integracji.

Liczy się proces, nie tylko raport

Po piąte, przy rozpowszechnianiu pomysłów i wniosków ważniejsze od raportu końcowego okazują się dyskusje, proces dochodzenia do konkluzji i rekomendacji. Raport końcowy jest formalnością określającą granice kontraktu.

Wspieranie aktywnego uczestnictwa interesariuszy.

Wreszcie po szóste, partycypatywne podejście sugerowane w przypadku ewaluacji średniookresowej i bieżącej wymaga aktywności obydwu stron badania – tak ewaluatorów, jak i badanych, szczególnie zlecających kontrakt. Dzielenie się wiedzą o kwestiach technicznych programu, współpraca z ewaluatorami, informowanie o problemach i koncentrowanie się na wypracowaniu rozwiązań to elementy niezbędne w procesie interaktywnym. Analizowane przypadki pokazały również kluczową rolę „mecenatu" nad badaniem instytucji zlecającej. Badanie musi być robione dla kogoś, ktoś musi na nie czekać i interesować się jego przebiegiem, by ewaluator nie był pozostawiony sam sobie.

3. Sugestie dla ewaluatorów:

Szersze rekomendacje dla ewaluatorów płynące z niniejszego badania można podsumować w poniższych sześciu punktach.

Partycypatywność skuteczna przy *mid-term* i *on-going*

Po pierwsze, w przypadku ewaluacji średniookresowej (a także bieżącej) sprawdza się podejście interaktywne, partycypatywne, zorientowane na jasno określoną grupę odbiorców. W przypadku programów współpracy terytorialnej grupa operatorów programu jest relatywnie szersza i bardziej zróżnicowana, co stawia dodatkowe wyzwania.

Po drugie, należy zbudować trwałą relację – opartą na zaufaniu – ze zlecającym i operatorami programu. To wymaga od ewaluatora dużej aktywności. Zaufanie może być

Budowanie zaufania

wzmocnione dopuszczeniem zleceniodawców do roboczej strony badania („pokazanie ewaluacji od kuchni"). W procesie budowania zaufania kluczowe jest użycie metod interaktywnych – fokusów, warsztatów, mniej formalnych spotkań roboczych.

Wypracowanie wspólnych definicji

Po trzecie, prócz oczekiwań zlecających ważne jest zidentyfikowanie ich stanu wiedzy o programie, percepcji programu, a także pomoc w wypracowaniu wspólnych terminów (systemu definicji). Język, kultura oraz przyzwyczajenia z różnych systemów zarządzania programami UE determinują odmienną interpretację podstawowych terminów i procesów. Na tym etapie pomocnym mogą się okazać metody map mentalnych.

Uczenie się od operatorów programu

Po czwarte, ewaluator musi wejść w logikę programu, zrozumieć jego procedury, ograniczenia, zidentyfikować głównych aktorów, układ sił i zależności proceduralno-instytucjonalne, kontekst. W fazie strukturalizacji relacje są odwrócone, ewaluator jest uczącym się, operatorzy programu dzielą się swoją wiedzą i doświadczeniem. Ponadto przy programach współpracy terytorialnej sposobem sprostania wyzwaniu całościowego, transgranicznego spojrzenia jest mieszany zespół badawczy (badacze z różnych krajów).

Przyjmowanie aktywnej roli animatora i edukatora

Po piąte, w przypadku ewaluacji zorientowanych na procesy ewaluator na etapie analizy i oceny to przede wszystkim animator. Jego rola polega na sprzężeniu perspektyw pojedynczych aktorów, skoordynowaniu dialogu wokół kluczowych problemów/szans. Siła, z jaką ewaluator będzie pełnił rolę edukatora zależy od tego, jak często będzie wprowadzał do dyskusji nowe głosy, sięgał po porównania, pokazywał szerszą perspektywę.

Zachowanie dystansu i obiektywizm.

Wreszcie po szóste, wyzwaniem strategii interaktywnej jest zachowanie obiektywizmu i dystansu. Pomóc może budowanie (na etapie analizy) autorytetu edukatora, który pozwala mu się zdystansować od interesariuszy i zlecających badanie w momencie oceny programu. Paradoksalnie, partycypatywne podejście może pomóc w zachowaniu dystansu – włączenie różnych perspektyw minimalizuje ryzyko zdominowania oceny przez punkt widzenia jednej instytucji. Uczestnicy dialogu nawzajem kontrolują proporcje.

Kończąc rekomendacje praktyczne, warto powtórzyć wniosek, który przewijał się już kilkakrotnie – dobremu wykorzystaniu ewaluacji *mid-term* (średniookresowych) i *on-going* (bieżących) służy nie tylko produkt końcowy – raport, ale cały proces badawczy. Uwzględnianie różnych perspektyw, włączanie różnych interesariuszy, partycypacja i aktywność ze strony wykonaw-

cy, zlecających oraz odbiorców badania są warunkami efektywnego uczenia się instytucji. Zaś dla programów współpracy terytorialnej to całościowe podejście (obejmujące różnych aktorów, ze wszystkich zaangażowanych krajów i obydwie strony granicy) jest warunkiem *sine qua non*.

Systemowe, partycypatywne podejście do ewaluacji jest też kluczowym elementem Nowego Rządzenia Sektorem Publicznym (*New Public Governance*). Wykorzystywanie go w bieżących badaniach może więc pomóc polskim instytucjom publicznym w przejściu do nowoczesnych paradygmatów zarządzania.

Nie jest to wniosek odkrywczy. Wartość współuczestnictwa w procesie uczenia się podkreślał już Konfucjusz: „Powiedz mi, a zapomnę, pokaż, a zapamiętam, zaangażuj mnie, a zrozumiem". Ta myśl jest mottem dla wielu programów edukacyjnych od szkół podstawowych po specjalistyczne programy studiów. Powszechnym jest jednak traktowanie rzeczywistości administracji publicznej i programów UE jako materii zbyt skomplikowanej i poważnej, by stosować do niej edukacyjne porównania. Wydaje się, że fundusze wymagają wysublimowanych, skomplikowanych podejść i procedur, ezoterycznego języka.

Tymczasem, parafrazując słowa znanego analityka organizacji publicznych H. Mintzherga, „świat naszych skomplikowanych organizacji" jest tworzony i napędzany przez ludzi. To ludzka przenikliwość i zaangażowanie – osób i zespołów – jest podstawą rozwoju i dobrego funkcjonowania nawet najbardziej skomplikowanych programów publicznych. A mechanizmy uczenia się ludzi są zawsze podobne. Przy takim ujęciu „ewaluacji jako procesu uczenia się" pozbywamy się bagażu ewaluacji jako „nowej sztuki dla wtajemniczonych". Ewaluacja zaczyna jawić się jako nieustający powrót do „szkoły", szkoły otwartej i interaktywnej, opartej na ciągłym stawianiu pytań, dyskusji, krytycznej refleksji nad naturą naszych działań i samodoskonaleniu się. To podejście może się wydawać nieco górnolotne. Stawką w tym procesie uczenia się instytucji i ludzi jest jednak racjonalny rozwój społeczno-gospodarczy kraju.

6. ANEKSY

6.1. METODOLOGIA BADANIA

Pytania, hipotezy badawcze

Głównym pytaniem postawionym w badaniu było: „W jaki sposób wykorzystano raporty ewaluacyjne o wysokiej jakości w wybranych programach INTERREG III?”. Pytaniem uzupełniającym było natomiast: „Jakie czynniki i procesy wpłynęły na wykorzystanie poszczególnych ewaluacji?”.

Zbudowana w oparciu o te pytania **hipoteza główna** brzmiała: „**Raporty ewaluacyjne o wysokiej jakości były przydatne w zarządzaniu programem INTERREG – zarówno w poprawie bieżącego zarządzania, jak i w przygotowaniu przyszłych programów współpracy terytorialnej”**. Skoncentrowano się na raportach o wysokiej jakości, gdyż to jakość raportu jest podstawowym warunkiem jego pozytywnego użycia w zarządzaniu sektorem publicznym.

Powyższa hipoteza główna została wsparta przez dwie grupy hipotez pomocniczych.

Grupa pierwsza, **hipotezy pomocnicze 1–5**, uszczegóławiała sposoby wykorzystania ewaluacji, czyli funkcje, które dany raport mógł spełnić w sektorze publicznym.

Hipoteza pomocnicza 1: Dobry raport ewaluacyjny przyczynił się do poprawienia planowania interwencji publicznych.

Hipoteza pomocnicza 2: Dobry raport ewaluacyjny przyczynił się do poprawienia wdrażania interwencji publicznych.

Hipoteza pomocnicza 3: Dobry raport ewaluacyjny przyczynił się do wzmocnienia odpowiedzialności w sektorze publicznym (ang. *accountability*).

Hipoteza pomocnicza 4: Dobry raport ewaluacyjny wsparł procesy uczenia się w sektorze publicznym.

Hipoteza pomocnicza 5: Dobry raport ewaluacyjny przyczynił się do wzmocnienia partnerstwa i poczucia współwłasności uczestników interwencji publicznych.

Druga grupa hipotez, **hipotezy pomocnicze 6–8**, precyzowały czynniki warunkujące wykorzystanie badania ewaluacyjnego.

Hipoteza pomocnicza 6: Wykorzystanie ewaluacji jest warunkowane przez czynniki wyjściowe ewaluacji, takie jak kontekst badanego programu i jego charakter, charakterystykę uczestników badania.

Hipoteza pomocnicza 7: Wykorzystanie ewaluacji jest warunkowane przez cechy procesu badawczego – obrane podejście badawcze, interaktywność procesu i role uczestników.

Hipoteza pomocnicza 8: Wykorzystanie ewaluacji warunkowane jest przez cechy produktów badania ewaluacyjnego – nawiązane kontakty, jakość

raportu, rodzaj wiedzy zawartej w raporcie i wreszcie sposób komunikacji wyników.

Uzasadnienie zakresu badania

Inicjatywa Wspólnotowa INTERREG miała dotychczas trzy edycje (INTERREG I dla lat 1990–1993, INTERREG II w latach 1994–1999 i INTERREG III w latach 2000–2006). Badanie skoncentrowano na ostatniej edycji z kilku powodów. Po pierwsze, edycja ta dobiega końca – stanowiła więc już zamkniętą jednostkę analizy (liczba programów, cele i strategia wdrażania programów już się nie zmienią). Z drugiej strony wciąż jednak istniała możliwość dotarcia do osób i instytucji zaangażowanych zarówno we wdrażanie programów III edycji, jak i planujących INTERREG IV (co pozwoliło prześledzić ewentualny długookresowy wpływ wniosków z ewaluacji). Po drugie, dopiero przy INTERREG III prowadzono pełną dokumentację programów na poziomie europejskim oraz ujednolicono procedury zarządzania. Materiał badawczy był więc dużo szerszy, a kontekst poszczególnych programów bliższy teraźniejszości w porównaniu do dwóch poprzednich edycji. Po trzecie, istniała ciągłość między trzecią a czwartą edycją, co dało pole do ewentualnych wniosków praktycznych na przyszłość.

Prowadzone badanie skoncentrowano na ewaluacjach *mid-term* (średniookresowych), bo tylko w przypadku tego typu badania zostały przeprowadzone dla wszystkich programów INTERREG III mniej więcej w tym samym czasie (w połowie 2003 roku) i w dokładnie takim samym zakresie (lista pytań badawczych został narzucona przez Komisję Europejską). To zestandaryzowanie umożliwiło wyłączenie z analizy części zmiennych powszechnie uważanych za główne czynniki wpływające na wykorzystanie ewaluacji (np. zakres pytań badawczych, moment podjęcia badania), dało też unikalne pole do porównań w skali europejskiej. Ponadto wybór ewaluacji średniookresowych jako punktu analizy umożliwił śledzenie długookresowych efektów ewentualnego wykorzystania raportów (np. relacji między badaniami średniookresowymi a jego uaktualnieniami). Wreszcie tylko w przypadku ewaluacji średniookresowej znana i dostępna była praktycznie cała populacja raportów. Można więc było wybrać studia przypadków w sposób usystematyzowany i obiektywny.

Uzasadnienie podejścia badawczego

W naukach społecznych wyróżniane są dwa zasadnicze podejścia do rozumowania i logiki badań: dedukcyjne – traktujące jako punkt wyjścia teorię, która jest następnie weryfikowana w procesie badawczym oraz indukcyjne – gdy obserwacje i wnioski zostają uogólnione i prowadzą do stworzenia teorii (Bryman, 2004: 5–11)[1]. Walliman zaznacza możliwość łączenia rozu-

[1] Bryman zaznacza jednak, iż w praktyce badań podział ten często nie jest ostry. Należy go traktować raczej jako zarysowanie kierunku tendencji w myśleniu, a nie jednoznaczną, kontrastową typologię.

mowań, wyróżniając trzecią metodę – rozumowanie hipotetyczno-dedukcyjne (ang. *hypothetico-deductive metod*) – zdefiniowaną przez Karla Poppera. W tym rozumowaniu teoria poprzedza badania, jednak działają one na zasadzie sprzężenia zwrotnego, kolejne obserwacje i wnioskowania prowadzą do korekty elementów danego modelu.

Dla szerokich strategii badawczych podstawowy podział opiera się na rozróżnieniu między strategiami ilościowymi i jakościowymi różniącymi się pod względem kierunku rozumowania, przyjętej epistemologii i ontologii (Bryman, 2004: 19–21; Creswell, 1994: 145–159; Walliman, 2006: 36–37). Te pierwsze używają rozumowania dedukcyjnego – akcent stawiają na testowanie teorii i relacji przyczynowo-skutkowych. Opierają się na filozofii pozytywistycznej, używają praktyk i norm badań przyrodniczych (a więc głównie planu eksperymentalnego), zakładają też, że rzeczywistość społeczna jest obiektywną i zewnętrzną rzeczywistością. Natomiast strategie jakościowe używają raczej podejścia indukcyjnego, badają procesy i ich znaczenie, odrzucają pozytywizm, opierając się na indywidualnej interpretacji procesów społecznych, w sferze ontologii stosują zaś konstrukcjonizm (rzeczywistość społeczna jest postrzegana w sposób subiektywny i poprzez zinstytucjonalizowane interakcje między ludźmi; Frankfort-Nachmias, Nachmias, 2001).

Wreszcie w odniesieniu do planu badawczego (ang. *research design*) – czyli „projektu kierującego badaczem na etapie zbierania, analizy i interpretacji danych" (Frankfort-Nachmias, Nachmias, 2001: 605) – wyróżnić można cztery podstawowe kategorie (Bryman, 2004: 33–55)[2]:

- Klasyczny plan eksperymentalny lub plan *quasi*-eksperymentalny (ang. *experimental & quasi-experimental design*), w którym dwie porównywalne, równoważne grupy zostają poddane eksperymentowi – w grupie eksperymentalnej działa zmienna zależna, w grupie kontrolnej nie. Wyniki badania dają możliwość weryfikacji hipotezy na temat relacji przyczynowo-skutkowej;
- Plan przekrojowy (ang. *cross-sectional design*) używany głównie do badania związków między właściwościami a dyspozycjami, polega na zebraniu danych tak jakościowych, jak i ilościowych dotyczących kilku zmiennych, w jednym momencie czasowym, w odniesieniu do kilkunastu-kilkuset obiektów badawczych (najczęstszą metodą są badania sondażowe). Daje możliwość opisania relacji, ale trudno na tej podstawie ustalić zależność przyczynową;
- Plan porównań podłużnych (ang. *longitudinal design*) – strategia badania, w której głównym kryterium jest czas. Przedmioty badania są obserwowane wielokrotnie na przestrzeni lat. Ten plan badawczy służy zrozumieniu mechanizmów zmian i ich czynników;
- Plan oparty na studiach przypadków (ang. *case study design*) polega na dogłębnej i szczegółowej analizie jednego lub kilku przypadków i ich kontekstu. W badaniu zwykle używane są metody jakościowe, ale równie upraw-

[2] Tłumaczenia terminów angielskich i części definicji na podstawie: Frankfort-Nachmias, Nachmias (2001).

nione jest użycie metod ilościowych. Ten plan badawczy pozbawiony jest trafności zewnętrznej. Badanie koncentruje się więc nie na generalizowaniu wniosków w odniesieniu do szerszego kontekstu, ale na tym, jak sprawnie badacz wyprowadza teorię z dokonanych obserwacji (jakości procesu testowania lub budowania teorii). Część z autorów podkreśla jednak unikalną wartość studiów przypadków w budowaniu teorii – wcześniejsze koncepcje zostają poddane falsyfikacji na konkretnych przypadkach, dana jest możliwość dogłębnej analizy kontekstowej i osadzenia teorii w rzeczywistości społeczno-gospodarczej (Flyvbjerg, 2006). W studiach przypadków stosować można tak rozumowanie indukcyjne, jak i dedukcyjne (choć przeważa to pierwsze).

Przy jakościowych planach badań do tych kategorii dodawana jest też strategia teorii bazowej (ang. *grounded theory*). W tym podejściu badacz zaczyna od ogólnego pytania badawczego, zbiera dane i porządkując je, wyprowadza z nich hipotezy oraz kategorie pojęciowe (tzw. kodowanie). Następnie zbiera kolejne dane i na ich podstawie wprowadza korekty do kategorii wyjściowych i tworzonej teorii. Ten cykl dopracowywania kategorii jest wielokrotnie powtarzany. Strategię teorii bazowej stosuje się w badaniach terenowych. Johnson (1998: 94–95), wzorując się na logice tej strategii badawczej, proponuje w odniesieniu do badań koncepcji ewaluacji podejście metamodelowe (ang. *metamodeling*). Metamodelowanie, tak jak strategia teorii bazowej, polega na poszukiwaniu (w oparciu o dane) kategorii pojęć, porządkowaniu ich w formie modelu, wzbogacaniu go kolejnymi danymi. Jednak metamodelowanie, w odróżnieniu od teorii bazowej, ma swoje zakorzenienie nie w obserwacji terenowej, a w przeglądzie literatury przedmiotu[3].

W zaprezentowanym badaniu zastosowana została hipotetyczno-dedukcyjna metoda rozumowania z jakościową strategią badania opartą na połączeniu metamodelowania i studiów przypadków.

Zastosowanie opisanego powyżej podejścia i planu badawczego wsparły dwie grupy przesłanek: negatywne – wykluczające alternatywne podejścia i plany badawcze, pozytywne – wspierające obraną strategię.

Jako przesłanki negatywne należy wymienić:

- wielkość populacji będącej przedmiotem zainteresowania badawczego (raporty z siedemdziesięciu dwóch międzynarodowych programów) oraz specyfika źródeł danych (zróżnicowanie języków, dostępności do dokumentów, brak ciągłości instytucjonalnej i personalnej) uniemożliwiła z przyczyn obiektywnych przeprowadzenie badań ilościowych;
- skala badanej populacji i jej złożoność (kilkadziesiąt osób w każdym z kilkudziesięciu programów) uniemożliwiła również zastosowanie planu prze-

[3] Drugim źródłem inspiracji są modele przyczynowe (ang. *causality models*). Metamodele stosują te same trzy kryteria budowania związków przyczynowych: sprawdzanie w celu ustalenia relacji między zmiennymi, ustalanie kolejności następstwa zmiennych (zwykle na podstawie wcześniejszych teorii), eliminowanie zmiennych i alternatywnych wyjaśnień w oparciu o obserwację i analizę teoretyczną. Metamodelowanie kładzie jednak dużo słabszy akcent na testowanie modelu, a koncentruje się na jego tworzeniu i budowaniu.

krojowego, zaś brak ciągłości danych i instytucji, jak również horyzont czasowy niniejszego badania – planu porównań podłużnych;
- z uwagi na wieloaspektowy rodzaj relacji, kompleksowość czynników i procesów (ludzkie, organizacyjne, techniczne, prawne) oraz środowiska (instytucje z różnych krajów, środowisko programu i poziom europejski, krótka i długa perspektywa czasowa), którego dotyczyła analiza wykorzystania ewaluacji, niemożliwym było kontrolowanie zmiennych, a co za tym idzie zastosowanie podejścia ilościowego i planu *quasi*-eksperymentalnego;
- wyjątkowe wewnętrzne zróżnicowanie (każdy program jest kombinacją co najmniej dwóch kultur administracyjnych i szeregu odmiennych zestawień instytucjonalnych) sprawiły, że trudnym byłoby wyłonienie reprezentatywnej grupy programów i generalizowanie na podstawie ich obserwacji.

Natomiast główne pozytywne argumenty wspierające wybrane podejście i plan badawczy były następujące:
- hipotetyczno-dedukcyjne, metamodelowe podejście dało możliwość wykorzystania szerokiej literatury przedmiotu, przeniesienia i logicznej adaptacji (w tym przetestowania) wypracowanych już koncepcji i doświadczeń nauki światowej na grunt europejski i krajowy;
- jakościowa strategia pozwoliła badać procesy, interakcje i ich postrzeganie przez różnych aktorów – te elementy są identyfikowane jako rdzeń procesów uczenia się i ewaluacji;
- strategia jakościowa dała możliwość badań terenowych – ważnych z perspektywy uchwycenia instytucjonalnych i personalnych determinantów wykorzystania ewaluacji, wykorzystano też techniki opisowe, które w najpełniejszy sposób pozwoliły uchwycić wieloaspektową rzeczywistość wykorzystywania ewaluacji w programach UE;
- nie istnieje spójna teoria wykorzystania ewaluacji – w związku z tym uzasadnionym było podejście metamodelowe porządkujące, kategoryzujące i integrujące zróżnicowane elementy, koncepcje i wyniki wcześniejszych badań;
- plan badawczy oparty na studiach przypadków posłużył zaprezentowaniu jakości procesu testowania i falsyfikowania koncepcji, dobrze współgrał więc z celami niniejszego badania, które koncentrowało się na weryfikacji i udoskonaleniach modelu bazującego na wcześniejszych doświadczeniach;
- studia przypadków umożliwiły dogłębną analizę kontekstu, aktorów i relacji, były więc najbardziej odpowiednim planem badania wyjątkowej mieszanki czynników jednostkowo-organizacyjnych i kulturowych będących specyfiką programów współpracy terytorialnej;
- połączenie podejścia metamodelowego ze studiami przypadków osadziło rozważania teoretyczne w ramach europejskich doświadczeń, dało unikalną możliwość weryfikacji koncepcji w praktyce, odniesienia ich i przystosowania do kompleksowego zarządzania programami europejskimi.

Plan badawczy

Punktem wyjścia dla prezentowanego badania była więc teoria. Ponieważ jednak nie istnieje spójna teoria wykorzystania ewaluacji – można tu raczej

użyć terminu „koncepcji" – zastosowano więc podejście metamodelowe do wypracowania spójnego modelu wykorzystania ewaluacji. Model ten następnie zastosowano i wzbogacono w badaniu empirycznym opartym na studiach przypadków.

Pierwszym etapem badania było zebranie i analiza koncepcji wykorzystania ewaluacji zarysowanych lub sformułowanych *explicite* w obszernej literaturze przedmiotu, opartych na rozważaniach teoretycznych bądź empirii. Na tej podstawie wypracowano metamodel wykorzystania ewaluacji. Opisał on mechanizm wykorzystania ewaluacji: zarówno zmienne niezależne – czynniki wpływające i warunkujące wykorzystanie, jak i zmienne zależne – formy wykorzystania ewaluacji i funkcje, które może ona spełniać w sektorze publicznym. Model określił także sekwencję czasową zmiennych.

W drugim etapie badania metamodel zastosowano do analizy trzech studiów przypadków. Było to sześć raportów wysokiej jakości z ewaluacji średniookresowej (*mid-term*) dotyczących sześciu programów europejskiej współpracy terytorialnej – INTERREG III. Skoncentrowano się na raportach o wysokiej jakości, gdyż jest ona warunkiem *sine qua non* jego pozytywnego użycia w zarządzaniu sektorem publicznym. Dopiero spełnienie tego warunku wyjściowego pozwala analizować dalsze procesy i ewentualne czynniki wpływające na zakres i sposób wykorzystania ewaluacji (czyli jej efekty). Analiza dobrych przypadków dała też pole do konstruktywnych wniosków praktycznych. Wybór konkretnych przypadków oparto na wynikach wcześniejszych badań jakości raportów ewaluacyjnych (metaewaluacja ewaluacji średniookresowych programów INTERREG III; Bachtler et al., 2005). Badanie tego, jak przebiegł proces wykorzystania ewaluacji w danym przypadku objęło analizę dokumentów programowych, wywiady z kluczowymi aktorami procesu (przedstawicielami instytucji zlecających badanie oraz głównymi ewaluatorami), analizę szerszych źródeł europejskich, a także wywiady przekrojowe – prowadzone z osobami zaangażowanymi w politykę współpracy terytorialnej UE na poziomie europejskim.

W trzecim etapie badania, po przeprowadzeniu powyższych analiz, wyciągnięto trzy grupy wniosków: dotyczące weryfikacji głównego pytania badawczego i postawionych hipotez badawczych, wnioski dotyczące szerszych paradygmatów administracji i zarządzania sektorem publicznym oraz praktyczne spostrzeżenia dotyczące zaobserwowanych dobrych praktyk, podejść, wsparcia procesów uczenia się, które można przenieść na grunt ewaluacji innych programów współpracy terytorialnej oraz szerzej – programów publicznych.

Metodologia studiów przypadków

Wszystkie trzy studia przypadków były poprowadzone według tej samej metodologii. Jako trzy główne zasady badania przyjęto:

- Podejście jakościowe – umożliwiło ono badanie procesów, interakcji, a jednocześnie zapewniło elastyczność i otwartość badania (reakcję na nowe dane i fakty nieprzewidziane w modelu);

- Zastosowanie triangulacji – obiektywizm badania zwiększyła kombinacja trzech głównych źródeł, to znaczy opinii osób z instytucji zlecającej, opinii ewaluatorów oraz analizy oficjalnej i roboczej dokumentacji związanej z badaniami;
- Zastosowanie zróżnicowanych źródeł czasowych – analiza dokumentów i kolejnych raportów (z dwóch, trzech lat po badaniu) dała możliwość zaobserwowania ewentualnego długoterminowego oddziaływania raportu.

Studia przypadków były analizowane z użyciem metamodelu wykorzystania ewaluacji stworzonego w pierwszym etapie badania. Każde ze studiów objęło:

- Analizę treści raportu ewaluacyjnego ewaluacji średniookresowej i ewentualnych dokumentów roboczych;
- Analizę raportu z uaktualnienia ewaluacji średniookresowej (raporty z 2005 roku);
- Analizę dokumentów programowych – Programu Operacyjnego (ang. *Community Indicative Programme* – CIP), Uzupełnienia Programu Operacyjnego (ang. *Programme Complement* – PC), roboczych dokumentów wewnętrznych;
- Analizę planów kontynuacji programu – projektów dokumentów na lata 2007–2013;
- Ustrukturalizowany wywiad z przedstawicielami instytucji zlecającej ewaluację – osobami nadzorującymi kontrakt ewaluacyjny;
- Ustrukturalizowany wywiad z członkiem zespołu ewaluacyjnego – kluczowym ewaluatorem.

Ponadto, w celu zidentyfikowania ewentualnego oddziaływania raportu w dłuższej perspektywie, poza programem przeprowadzono badania na poziomie europejskim:

- Wywiad z przedstawicielem europejskiej sieci tematycznej – INTERACT;
- Przegląd źródeł europejskich – studiów związanych tematycznie z INTERREG (głównie publikacje INTERACT) pod kątem ewentualnych odniesień do badanych raportów;
- Analizę treści stron internetowych (wykonane w styczniu 2007 roku).

Należy zaznaczyć, że struktury wywiadów ze zleceniodawcami ewaluacji i ewaluatorami były analogiczne. Dało to możliwość triangulacji – analizy danych kwestii z dwóch perspektyw. Co więcej, struktura przeprowadzonych wywiadów odzwierciedlała strukturę metamodelu. Spoiło to silniej część koncepcyjną pracy z częścią empiryczną. Lista wywiadów oraz ich struktury z instytucją zlecającą i ewaluatorami znajdują się w dalszej części aneksu.

Scenariusz wywiadu z przedstawicielami instytucji zlecającej ewaluację

Issue 1: Evaluation input

1.1. Were there any additional specific reasons (apart from EC requirements) for launching the MTE evaluation?

1.2. How MTE ToR was prepared (choice of evaluation question, approach requirements)?

Issue 2: Evaluation process

2.1. How the relations between you and evaluators looked like during the evaluation process?

2.2. What is your general assessment of evaluation process (professionalism of evaluators, transparency, smoothness, etc)?

- What were the most difficult issues/biggest problems
- What were the main successes?

2.3. From the present time perspective, what are your main lessons learnt?

Issue 3: Evaluation product

3.1. In general, how would you assess the Report Recommendations (their usefulness for you)?

3.2. In general, how would you assess the Report Findings (most positive, negative, and surprising)?

3.4. How did you disseminate the report?

3.5. What were the general reactions to the report & recommendations?

Issue 4: Evaluation use & assessment of its functions

4.1. How many recommendations were implemented?

4.2. Who, from present time perspective, was the main user of this report?

4.3. Did MTE helped you realized the implementation problems that you were not aware of?

4.4. Did you make changes in Prog strategy on the basis of this evaluation?

4.5. Did this report change your perspective/improve your understanding on cross-border cooperation mechanisms?

4.6. In general, was it useful experience? Did you learn anything new? As person/as organisation

4.7. Did the report findings were…:

- ...used by your organisation on EU level (conferences, policy debate)?
- ...used by any of the stakeholders for policy debate (SF debate)?
- ...used during the work on the New Programme debate?
- ...used for new evaluations (MTE Update)

Issue 5: Contacts to the key actors & documents

5.1. Who else, from main stakeholders, was really involved in MTE process?

5.2. Could you provide me with following documents/information?

- Copy of Terms of Reference for Mid-term Evaluation Tender
- Report from Mid-term Evaluation Update
- Other eventual evaluation reports & research done after MTE
- Links to sources, documents, that to your knowledge quote or base on MTE research

Scenariusz wywiadu z ewaluatorami

Issue 1: Evaluation input

1.1. How you were preparing the offer?

1.2. What was the composition of the evaluators team?

Issue 2: Evaluation process

2.1. How the relations between evaluators' team and stakeholders looked like during the evaluation process?

2.2. What is your general assessment of evaluation process?

- What were the most difficult issues/biggest problems
- What were the main successes?

2.3. From the present time perspective, what are your main lessons learnt?

Issue 3: Evaluation product

3.1. How you come up with the final form of the report?

3.2. How you were working on making the Recommendations operational and practical?

3.3. In general, how would you assess the Report Findings (most positive, negative, and surprising)?

3.4. How the dissemination of the results and report look like in practice?

3.5. What were the general reactions to the report & recommendations?

Issue 4: Evaluation use & assessment of its functions

4.1. Who, from present time perspective, was the main user of this report?

4.2. From the present time perspective how would you assess MTE impact?

- Was it mainly used for management purposes?
- Was it use for planning (present and new programme, national strategies?)
- Did it change the stakeholders' perspective/understanding of cbc issues in general?
- Was it useful for institutional networking/internal know-how?

4.3. Do you know if the report findings were…:

- ...used at national level (conferences, policy debate)?
- ...used at EU level (SF debate, conferences)?
- ...used during the work on the New Programme debate?
- ...used for new evaluations (MTE Update)

4.4. In general, was MTE exercise an useful experience? Did you learn anything new – as person/as organization?

Issue 5: Contacts to the key actors & documents

5.1. Who else, from main stakeholders, was really involved in MTE process?

5.2. Could you provide me with any additional documents or references? (follow up analysis, quotation of MTEs, etc)?

Lista wywiadów

Osoba	Funkcja	Data, miejsce	Studia przypadku
Esben Poulsen Moray Gilland Eveline Petrat Niels Bjerring Hansen	Urzędnicy Komisji Europejskiej (*Desk Officers*) DG-Regio, Wydział ds. INTERREG	06.2004, Bruksela	Wywiad przekrojowy
Matnia Agnetti	p.o. Dyrektora Sekretariatu INTERACT	29.05.2006, Wiedeń	Wywiad przekrojowy
Georg Schadt Aleksandra Demel Andrea Schwecherl	Koordynator INTERREG w Federalnym Urzędzie Kanclerskim Instytucja Zarządzająca AT-SK JTS wszystkich INTERREG III dla Austrii	01.06.2006, Wiedeń	Przypadek 1
Richard Hummelbrunner	Ewaluator programów INTERREG dla Austrii	03.06.2006, Graz	Przypadek 1
Karen Lennon	Ewaluator IIIA Irlandia – Irlandia Północna	03.07.2006, Belfast	Przypadek 2
Nuala Cormican Sean Mcanualla	JTS IIIA Irlandia – Irlandia Północna (SEUPB)	05.07.2006, Monoghan	Przypadek 2
Paul Stephenson	JTS IIIB NWE	01.08.2006, Internet	Przypadek 3
Monika Tanaka	JTS IIIB NWE	18.08.2006, Internet	Przypadek 3
Frank Wefering	Ewaluator IIIB NWE	26.09.2006, Internet	Przypadek 3
John Burton	Ambasador UE przy prezydencie Stanów Zjednoczonych, były premier Irlandii	15.12.2006, Waszyngton	Wywiad przekrojowy

6.2. MODELE WYKORZYSTANIA EWALUACJI

Schemat 1. Ramy analityczne użycia ewaluacji autorstwa J. Bradley Cousins, Kenneth A. Leithwood (źródło: Cousins, Leithwood, 1986: 348)

Schemat 2. Model procesu ewaluacji autorstwa J. Bradley Cousins, Kenneth A. Leithwood (źródło: Cousins, Leithwood, 1993 za Johnson 1998: 99)

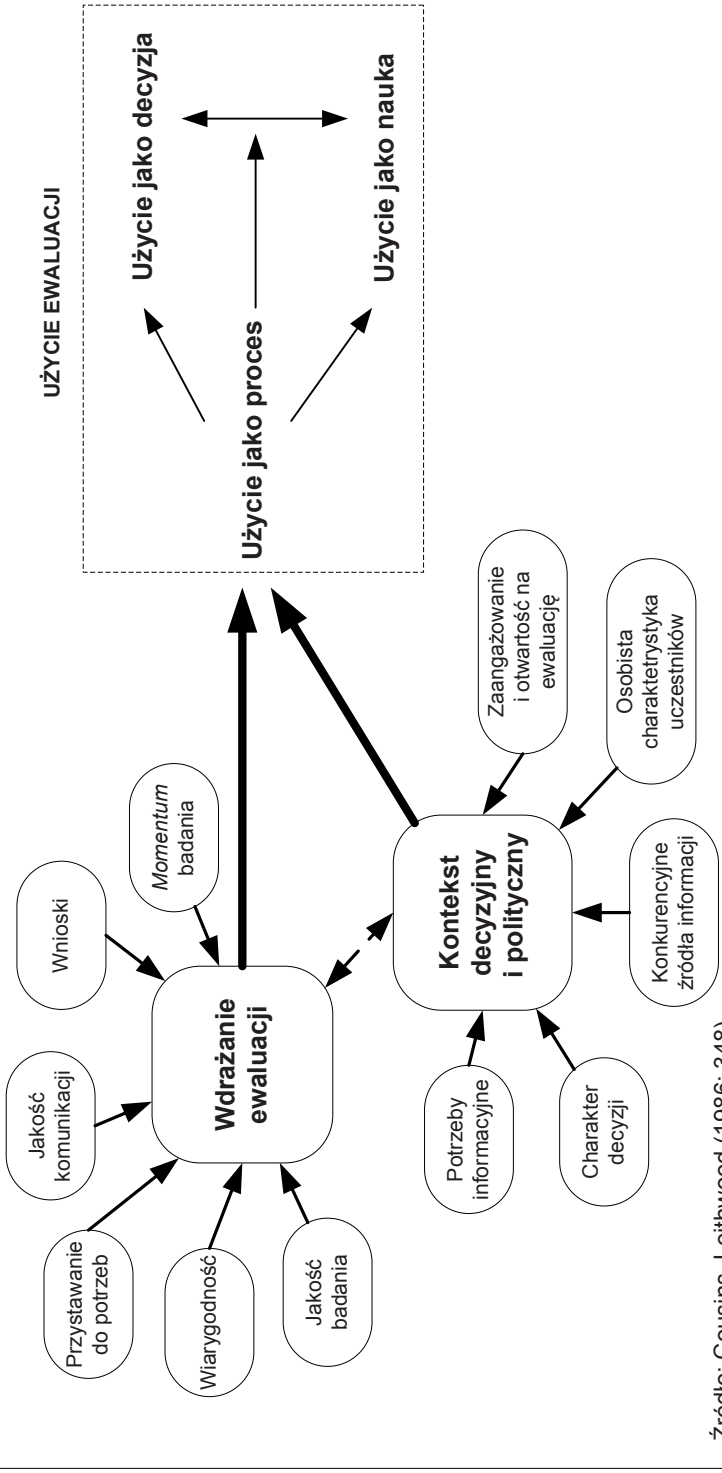
Schemat 3. Teoretyczny model wykorzystania ewaluacji autorstwa R. Burke Johnson (źródło: Johnson, 1998: 104)

Schemat 4. Model logiczny ewaluacji autorstwa Bradley J. Cousins rozbudowany przez Melvin M. Mark, Gary T. Henry (źródło: Mark, Henry, 2004: 38)

Schemat 5. Metamodel analizy wykorzystania ewaluacji (źródło: opracowanie własne na podstawie Cousins, 2003; Mark, Henry, 2004; Johnson, 1998; EC, 1999; Alkin, Taut, 2003; Batterbury, 2006)

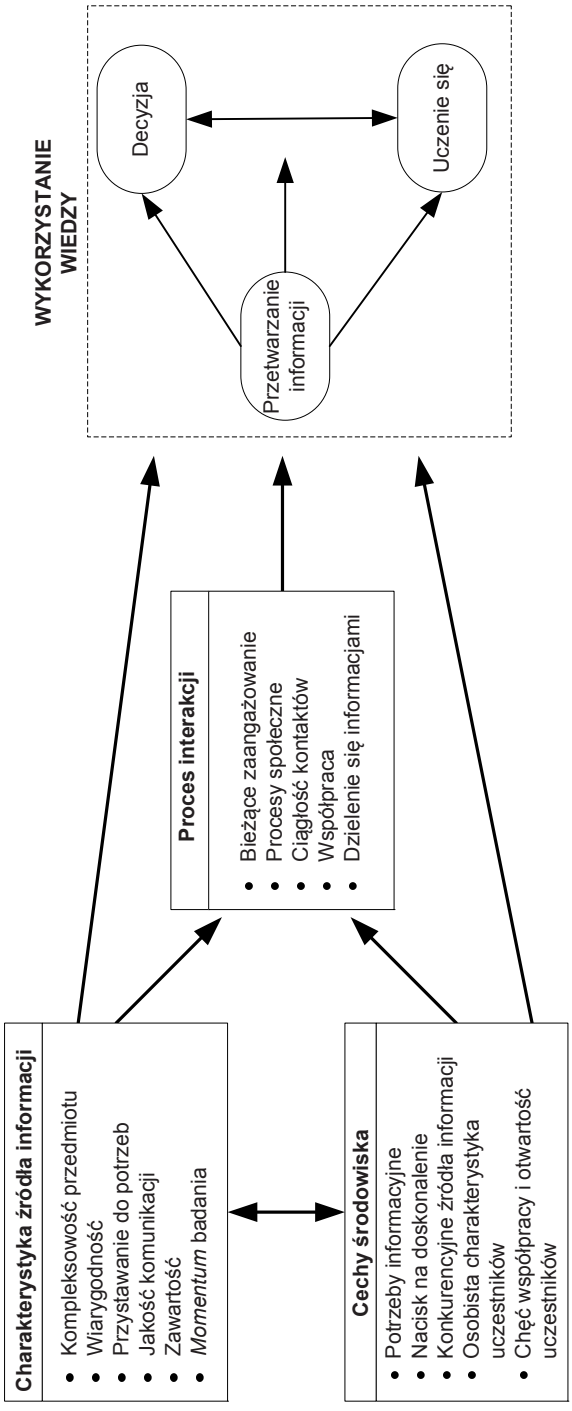
Schemat 6. Skorygowany metamodel analizy wykorzystania ewaluacji (źródło: opracowanie własne na podstawie Cousins, 2003; Mark, Henry, 2004; Johnson, 1998; EC, 1999; Alkin, Taut, 2003; Batterbury, 2006)

Schemat 1. Ramy analityczne użycia ewaluacji



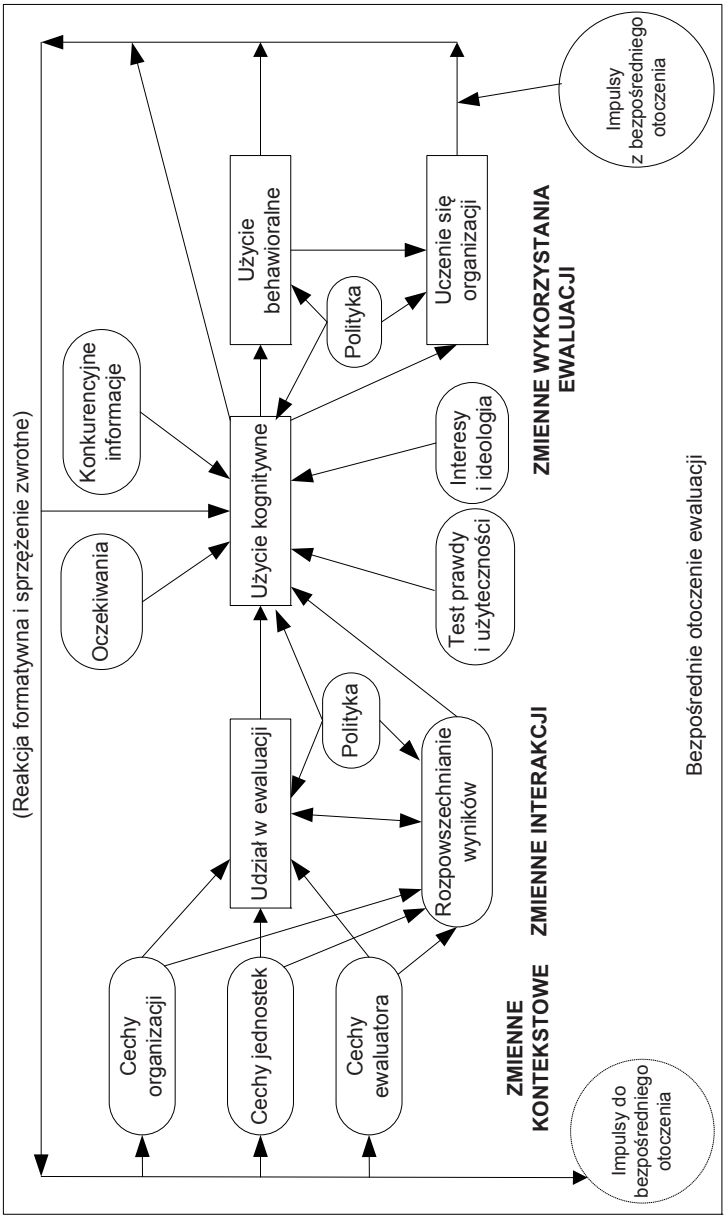
Źródło: Cousins, Leithwood (1986: 348)

Schemat 2. Model procesu ewaluacji



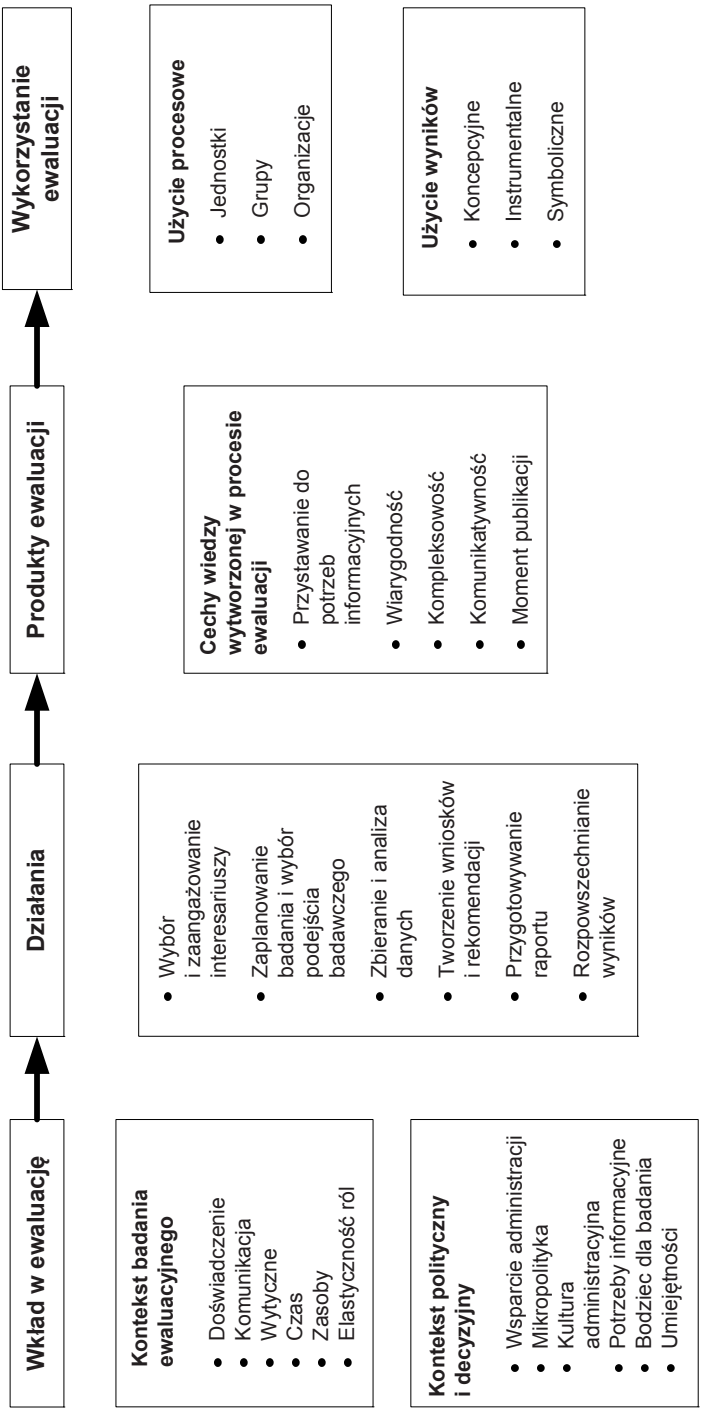
Źródło: Cousins, Leithwood (1993) za Johnson (1998: 99).

Schemat 3. Teoretyczny model wykorzystania ewaluacji



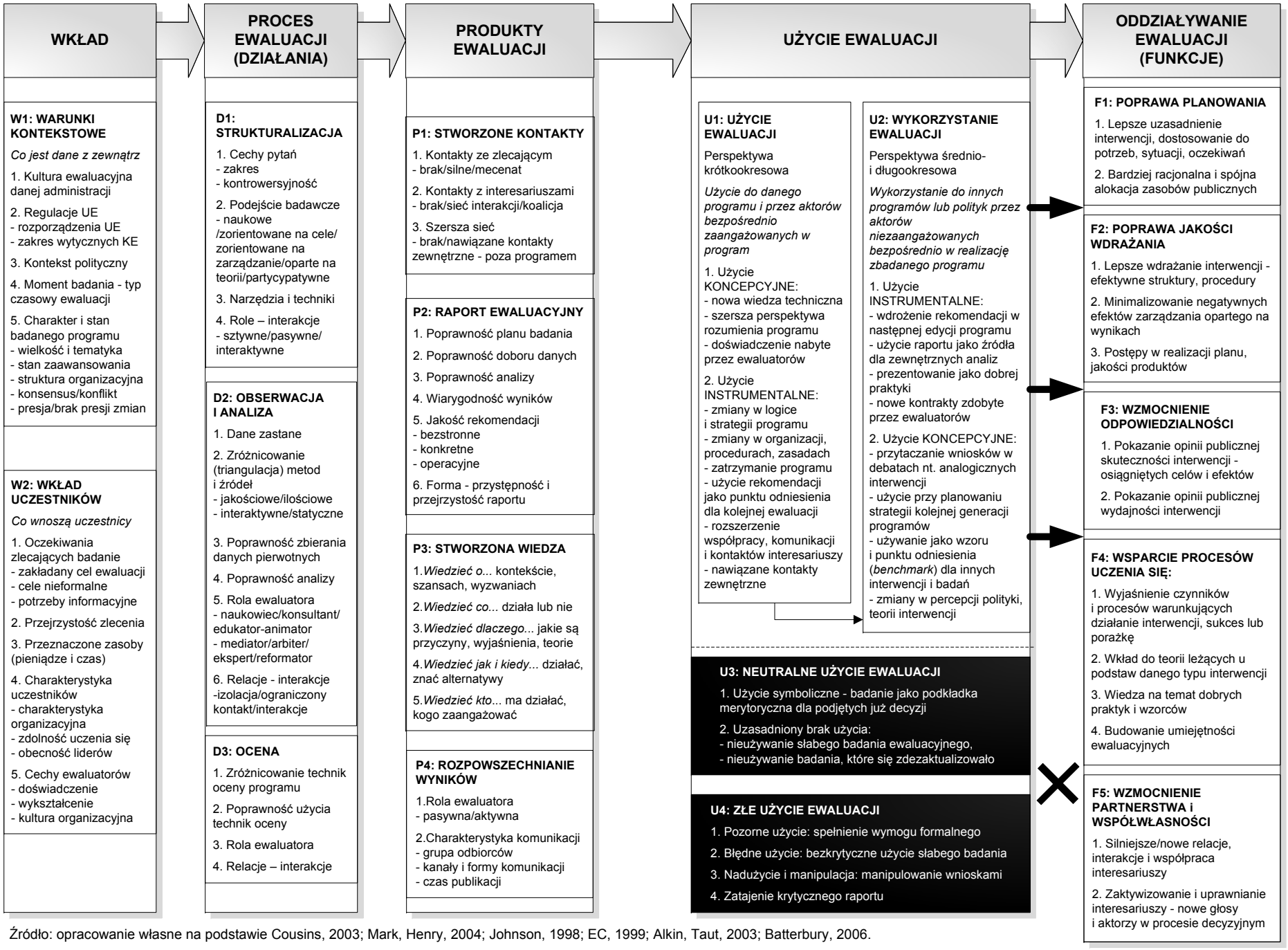
Źródło: Johnson (1998: 104).

Schemat 4. Model logiczny ewaluacji



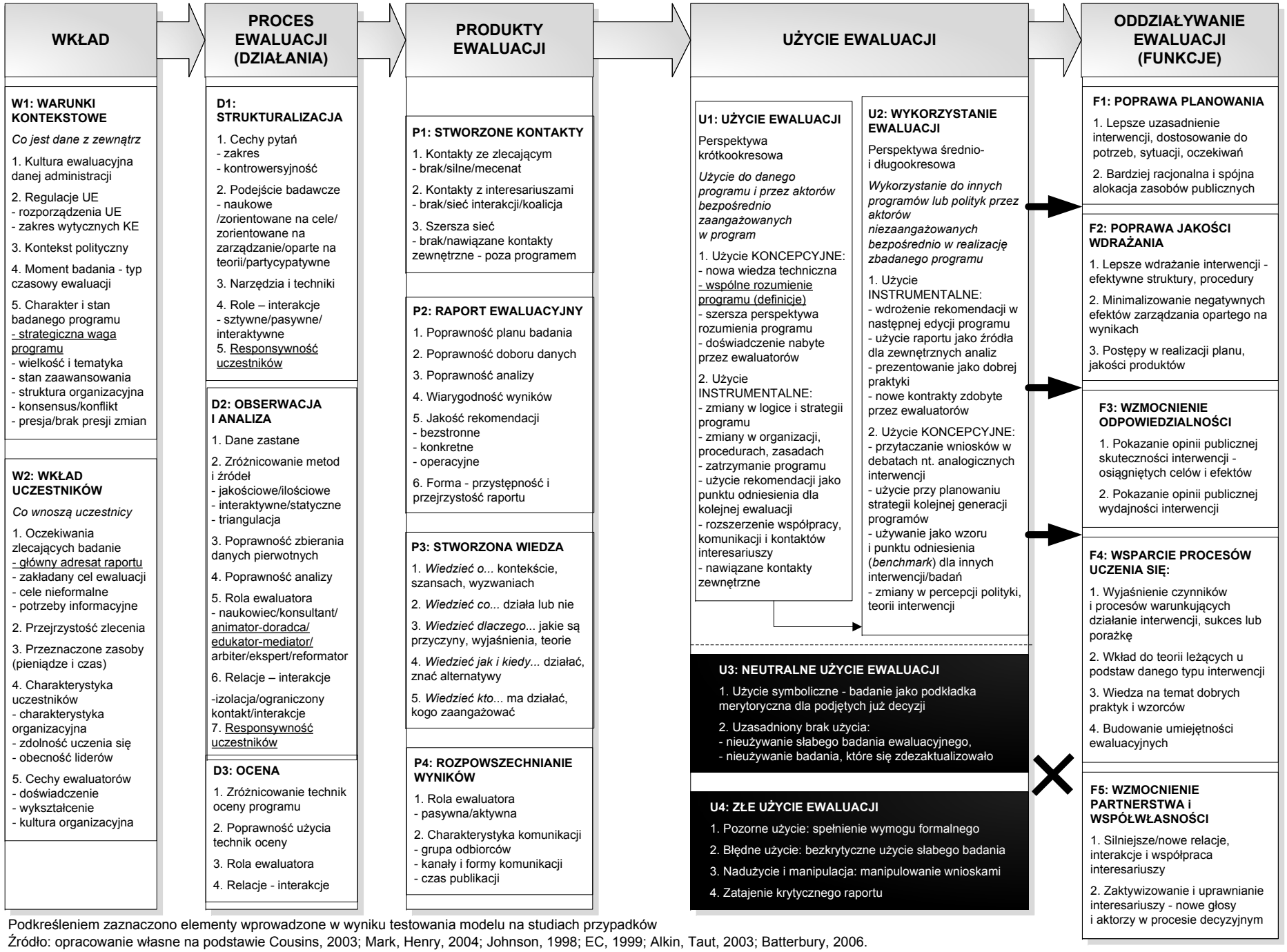
Źródło: Cousins (2003) rozbudowany przez Mark, Henry (2004: 38).

Schemat 5. Metamodel analizy wykorzystania ewaluacji



Źródło: opracowanie własne na podstawie Cousins, 2003; Mark, Henry, 2004; Johnson, 1998; EC, 1999; Alkin, Taut, 2003; Batterbury, 2006.

Schemat 6. Skorygowany metamodel analizy wykorzystania ewaluacji



Podkreśleniem zaznaczono elementy wprowadzone w wyniku testowania modelu na studiach przypadków

Źródło: opracowanie własne na podstawie Cousins, 2003; Mark, Henry, 2004; Johnson, 1998; EC, 1999; Alkin, Taut, 2003; Batterbury, 2006.

6.3. LISTY I MAPY PROGRAMÓW INTERREG III

Tabela 1. Programy INTERREG III A 2000–2004 i 2004–2006

Nr	Nazwa programu	Państwa (obszar przygraniczny – NTS3)
1	SE/FI Skärgĺrden	Finlandia, Szwecja
2	FI/SE/NO Kvarken – Mittskandia	Finlandia, Szwecja, Norwegia
3	DE/AT Bavaria/Austria	Niemcy, Austria
4	***AT/CZ Austria/Czech Republic***	***Austria, Czechy***
5	***AT/SI Austria/Slovenia***	***Austria, Słowenia***
6	***AT/HU Austria/Hungary***	***Austria, Węgry***
7	***AT/SK Austria/Slovakia***	***Austria, Słowacja***
8	SE/NO Sweden/Norway	Szwecja, Norwegia
9	DE/NL Ems-Dollart Region	Niemcy, Holandia
10	AU/DE/CH/LI Alpenrhein/Bodensee/Hochrhein	Austria, Niemcy, Szwajcaria, Lichtenstein
11	DE/PL Saxony/Lower Silesia	Niemcy, Polska
12	DE/CZ Saxony/Czech Republic	Niemcy, Czechy
13	NL/DE Euregio, Euregio Rhine-Waal -Meuse-North	Niemcy, Holandia
14	DE/PL Brandenburg/Lubuskie	Niemcy, Polska
15	IT/AT Italy/Austria	Włochy, Austria
16	IT/FR Alcotra	Włochy, Francja
17	IT/FR Islands	Włochy, Francja
18	IT/SI Italy/Slovenia	Włochy, Słowenia
19	***IE/UK Ireland/Northern Ireland***	***Wielka Brytania, Irlandia***
20	IE/UK Ireland/Wales	Wielka Brytania, Irlandia
21	FR/DE Pamina	Francja, Niemcy
22	DE/FR/CH Oberrhein Mitte-Süd – Rhin Supérieur Cent	Niemcy, Francja, Szwajcaria
23	DE/CZ Bavaria/Czech Republic	Niemcy, Czechy
24	DK/DE Fyn/K.E.R.N.	Dania, Niemcy
25	DK/DE Sřnderjylland/Schleswig	Dania, Niemcy
26	DK/DE Storstrøm/Ostholstein-Lübeck	Dania, Niemcy
27	DE/LU/BE Germany/Luxembourg/Belgium	Niemcy, Luksemburg, Belgia
28	DE/FR Saarland-Mosel/Lorraine- Western Palatinat	Francja, Niemcy
29	ES/PT Spain/Portugal	Hiszpania, Portugalia
30	ES/MA Spain/Morocco	Hiszpania, Maroko
31	IT/CH Italy/Switzerland	Włochy, Szwajcaria
32	SE/DK Öresund region	Dania, Szwecja
33	EL/AL Greece/Albania	Grecja, Albania
34	EL/MK Greece/FYROM	Grecja, Macedonia
35	EL/BG Greece/Bulgaria	Grecja, Bułgaria
36	EL/CY Greece/Cyprus	Grecja, Cypr

Nr	Nazwa programu	Państwa (obszar przygraniczny – NTS3)
37	DE/PL Mecklenburg-Vorpommern/Brandenburg/ Western Pomerania	Niemcy, Polska
38	BE/DE/NL Euregio Maas-Rhein	Niemcy, Holandia, Belgia
39	FI/RU Euregio Karelia	Finlandia, Rosja
40	FI/RU South East Finland/Russia	Finlandia, Rosja
41	FR/CH France/Switzerland	Francja, Szwajcaria
42	FR/ES France/Spain	Francja, Hiszpania
43	FI/SE/NO/RU Nord	Finlandia, Szwecja, Norwegia, Rosja
44	FI/EE Finland/Estonia	Finlandia, Estonia
45	BE/NL Grensregio Vlaanderen-Nederland	Belgia, Holandia
46	BE/FR/LU Wallonia/Lorraine/Luxemburg	Francja, Belgia, Luksemburg
47	FR/UK Franco – British programme	Wielka Brytania, Francja
48	UK/MA Gibraltar/Marocco	Wielka Brytania, Maroko
49	FR/BE France/Wallonia-Flanders	Belgia, Francja
50	IT/AL Italy/Albania	Włochy, Albania
51	EL/IT Greece/Italy	Grecja, Włochy
52	EL/TR Greece/Turkey	Grecja, Turcja
53	IT/AL/BA/HR/SCG Italy/Eastern Adriatic Countries	Włochy, Albania, Chorwacja, Bośnia i Hercegowina, Serbia i Czarnogóra
54	**CZ/PL Czech Republic/Poland**	**Czechy, Polska**
55	**PL/SK Poland/Slovakia**	**Polska, Słowacja**
56	**SK/CZ Slovakia/Czech Republic**	**Słowacja, Czechy**
57	**PL/UA/BY Poland/Ukraine/Belarus**	**Litwa, Ukraina, Białoruś**
58	**LT/PL/RU Lithuania/Poland/Russia**	**Litwa, Polska, Rosja**
59	**HU/SK/UA Hungary/Slovakia/Ukraine**	**Węgry, Słowacja, Ukraina**
60	**HU/RO/SCG Hungary/Romania/ Serbia & Montenegro**	**Węgry, Rumunia, Serbia i Czarnogóra**
61	**SI/HU/HR Slovenia/Hungary/Croatia**	**Słowenia, Węgry, Chorwacja**
62	**IT/MT Italy/Malta**	**Włochy, Malta**
63	**EE/LV/RU Estonia/Latvia/Russia**	**Estonia, Łotwa, Rosja**
64	**LV/LT/BY Latvia/Lithuania/Belarus**	**Łotwa, Litwa, Białoruś**

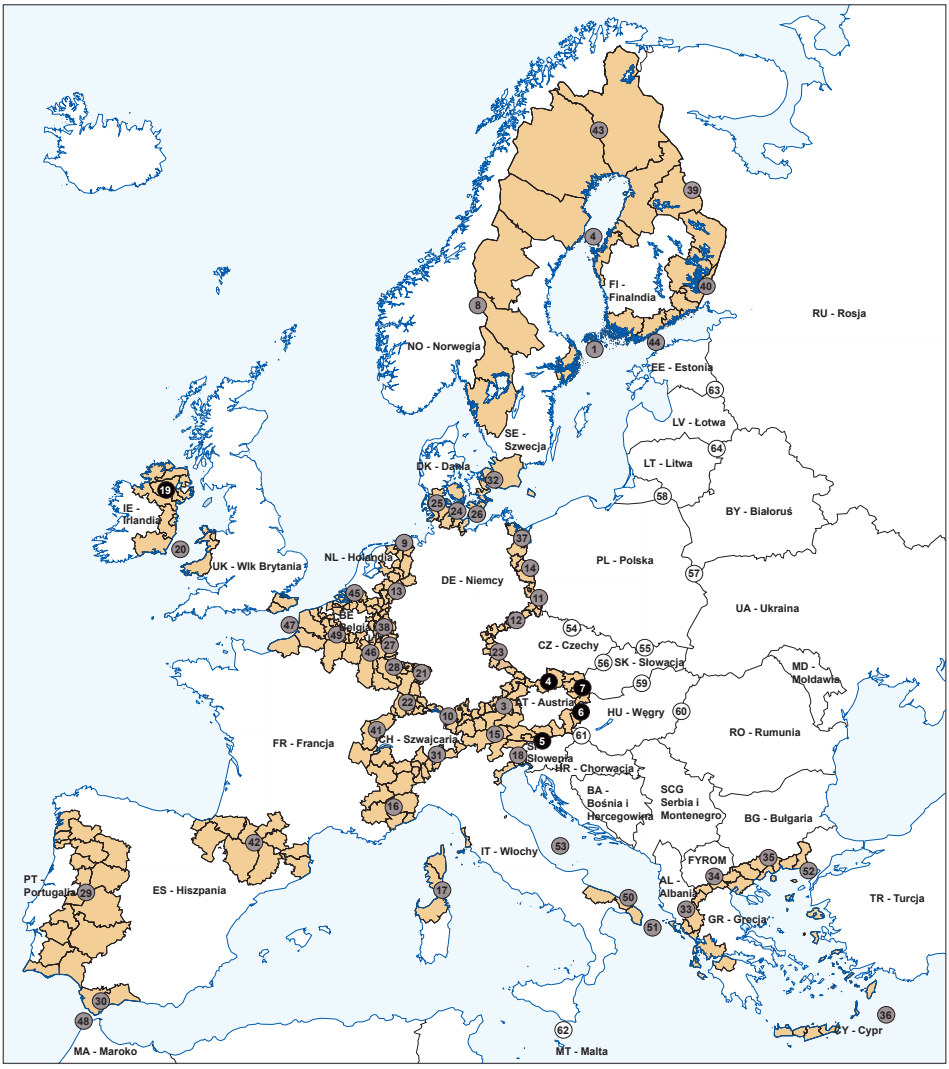
Wytłuszczeniem i kursywą zaznaczono programy wybrane objęte studiami przypadku.
Wytłuszczeniem zaznaczono programy dodane w 2004 roku.
Numery porządkowe odpowiadają numerom programów na Mapie 1.
Źródło: opracowanie własne na podstawie danych DG-Regio i programu INTERACT.

Tabela 2. Programy INTERREG III B 2000–2004 i 2004–2006

Nr	Nazwa programu	Współpracujące państwa/regiony
1	**Apline Space**	**Austria, Francja, Niemcy, Włochy, Słowenia, Szwajcaria, Liechtenstein.**
2	**Archimed**	**Włochy, Grecja, Cypr, Malta, Turcja, Liban, Syria, Autonomia Palestyńska, Izrael, Jordania, Egipt, Libia**
3	Atlantic Sea	Francja, Irlandia, Portugalia, Hiszpania, Wielka Brytania
4	**Balic Sea Region**	**Dania, Niemcy, Szwecja, Finlandia, Estonia, Litwa, Łotwa, Polska Norwegia, Rosja, Białoruś**
5	**CADSES**	**Austria, Grecja, Niemcy, wschodnie regiony Włoch, Polska, Czechy, Słowacja, Węgry, Słowenia Rumunia, Bułgaria, Chorwacja Bośnia-Hercegowina, Serbia I Czarnogóra, Macedonia, Albania, Mołdawia, Ukraina**
6	Northern Periphery	Finlandia, Szkocja, Szwecja, Norwegia, Grenlandia
7	North Sea Region	Belgia, Dania, Niemcy, Holandia, Szwecja, Wlk. Brytania Norwegia
8	South West Europe	Hiszpania, Francja, Portugalia, Wlk. Brytania
9	***North West Europe***	***Belgia, Francja, Niemcy, Irlandia, Luksemburg, Holandia, Wlk. Brytania Szwajcaria***
10	**Western Mediterranean**	**Region Włoch, Francji, Portugali, Hiszpanii, Wlk. Brytania (Gibraltar), Grecja, Malta**
11	Acores-Madeira-Canaires	Portugalia – Hiszpania (Azory – Madera – Wyspy Kanaryjskie)
12	Reunion	Reunion
13	Carribiean Space	Gwadelupa, Martynika, Gujana (francuska)

Wytłuszczeniem i kursywą zaznaczono program wybrany do studium przypadku.
Wytłuszczeniem zaznaczono programy rozszerzone w 2004 roku.
Źródło: opracowanie własne na podstawie danych DG-Regio.

Mapa 1: Programy INTERREG IIIA w latach 2000-2004 i 2004-2006



LEGENDA:

Regiony UE15 - NTS3 objęte Inicjatywą INTERREG IIIA

50 Program INTERREG IIIA, którego raport został wybrany do studium przypadku

18 Program INTERREG IIIA 2000-2006

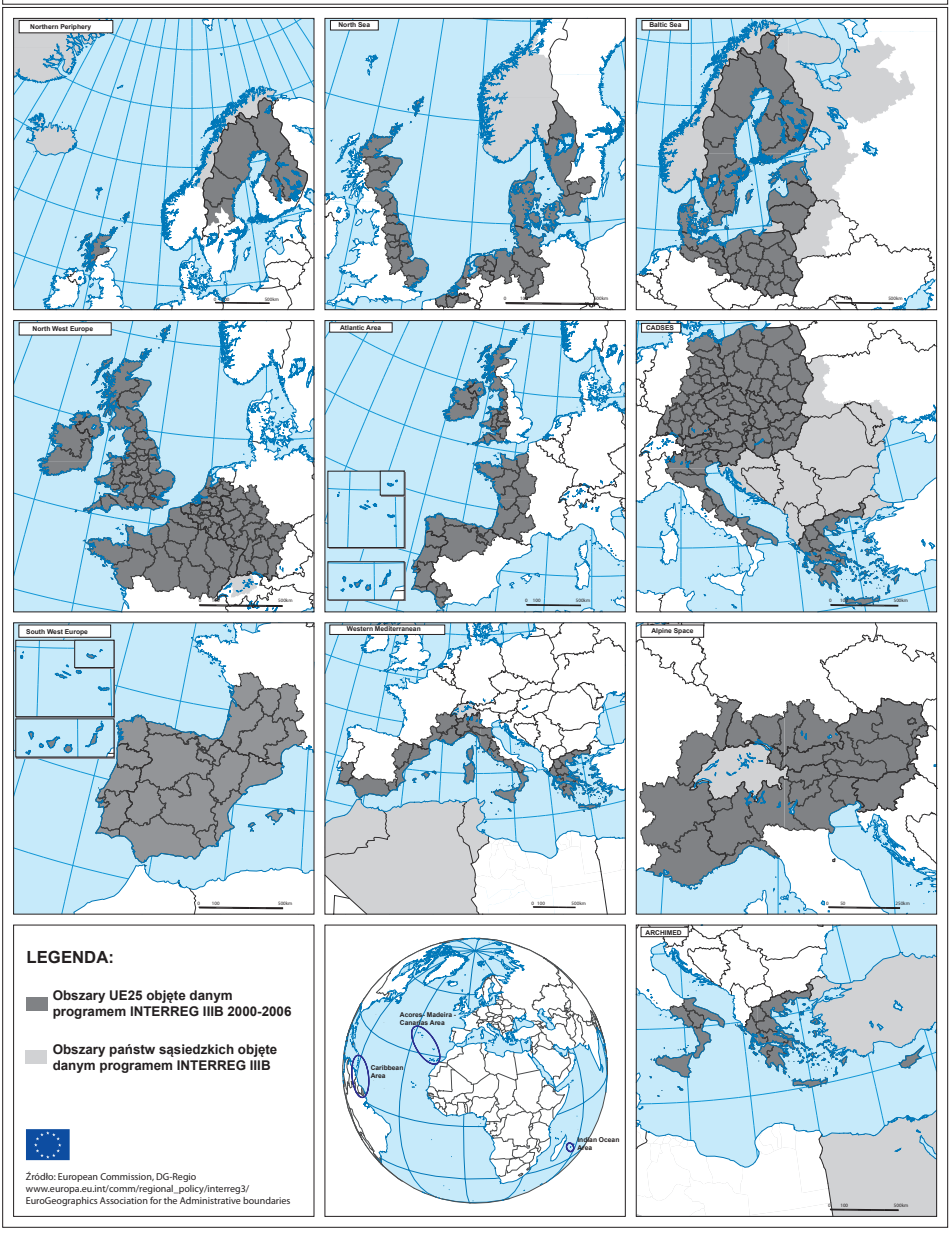
56 Program INTERREG IIIA uruchomiony po rozszerzeniu UE w 2004 r.

Opracowanie K.Olejniczak na podkładzie:
www.europa.eu.int/comm/regional_policy/interreg3/

SIG16

0 100 500km

Mapa 2: Programy INTERREG IIIB w latach 2000-2006



BIBLIOGRAFIA

Alkin M.C., 2004, *Evaluation roots: tracing theorists' views and influences*, Thousand Oaks, CA-London: Sage.

Alkin M.C., 2005, „Utilization of evaluation", w: S. Mathison (red.), *Encyclopedia of evaluation* (s. 143, 434–436). Thousand Oaks, CA-London: Sage.

Alkin M.C., Christie C.A., Rose M., 2006, „Communicating evaluation", w: I.Shaw, J.C. Greene, M.M. Mark (red.), *Handbook of evaluation: policies, programs and practices*, London: Sage, 384–403.

Alkin M.C., Coyle K., 1988, „Thoughts on evaluation utilization, misutilization and non-utilization", *Studies in Educational Evaluation*, nr 14(3), 331–340.

Alkin M.C., Taut S.M., 2003, „Unbundling evaluation use", *Studies in Educational Evaluation*, nr 29, 1–12.

Allen D., 2000, „Cohesion and the Structural Funds", w: H. Wallace, W. Wallace (red.), *Policy-making in the European Union* (s. 243–266). Oxford: Oxford University Press.

Amstrong H., 1993, „Community Regional Policy", w: J. Lodge (red.), *The European Community and the challenge of the future*, London: Pinter.

Bachtler J. (2000) *The evaluation of regional policy in the European Union – main report*, European Policies Research Centre, Glasgow.

Bachtler J., 2001, „Quod erat demonstrandum? Ewaluacja polityki regionalnej", *Studia Regionalne i Lokalne*, nr 4(7).

Bachtler J., Poverari L., Taylor S., 2000, „Culture, Commitment and Capacity. A comparative review of Structural Funds Evaluation in the European Union", *Edinburgh Conference on Evaluation*, Edinburgh, United Kingdom.

Bachtler J., Taylor S., 1999, *Objective 2: experiences, lessons and policy implications*, Glasgow: European Policies Research Centre, University of Strathclyde.

Bachtler J., Taylor S., Olejniczak K., 2005, *A Study of the Mid Term Evaluations of INTERREG programmes for the programming period 2000 until 2006*, Vienna: INTERACT Programme Secretariat.

Bachtler J., Wren C., 2006, „Evaluation of European Union Cohesion Policy: Research Questions and Policy Challenges", *Regional Studies*, nr 40(2), 143–153.

Balme R., 1997, „Regional policy and European Governance", w: M. Keating, J.P. Loughlin (red.), *The political economy of regionalism* (s. 64–67). London: Frank Cass.

Balthasar A., 2006, „The effects of Institutional Design on the Utilization of Evaluation. Evidenced using Qualitative Comparative Analysis (QCA)", *Evaluation*, nr 12(3), 353–371.

Batterbury S., 2006, „Principles and Purposes of European Union Cohesion policy Evaluation", *Regional Studies*, nr 40(2), 179–188.

Bauer-Wolf S., 2005, *Update of the Mid-Term Evaluation of the Community Initiative Programme INTERREG IIIA Austria-Hungary*, Vienna-Budapest: Managing Authority of INTERREG IIIA Austria-Hungary Programme.

Behrens P., Smyrl M., 1999, „The conflict of rationalities: EU regional policy and the Single Market", *Journal of European Public Policy*, nr 6(3).

Berk R.A., Rossi P.H., 1999, *Thinking about program evaluation*, Thousand Oaks, CA-London: Sage Publications.

Bingham R.D., Felbinger C.L., 2002, *Evaluation in practice: a methodological approach*, New York-London: Seven Bridges Press.

Bouckaert G., 2006, „Modernising Government: The way forward – a comment", *International Review of Administrative Sciences*, nr 72(3), 327–332.

Boyle R., Lemaire D., 1998, *Building effective evaluation capacity: lessons from practice*. New Brunswick, N.J.: Transaction Publishers.

Bryman A., 2004, *Social research methods*, Oxford: Oxford University Press.

Campbell D.T., Stanley J.C., 1967, *Experimental and quasi-experimental designs for research*, Chicago: Rand McNally College Publishing Company.

Caralcelli V.J., 2000, „Evaluation Use at the Threshold of the Twenty-First Century", *New Directions for Evaluation*, nr 88, 99–111.

Castells M., 2000, *The rise of the network society*, Oxford: Blackwell.

Castells M., Cardoso G., 2006, *The network society: from knowledge to policy*, Washington, D.C.: Center for Transatlantic Relations, Paul H. Nitze School of Advanced International Studies, Johns Hopkins University.

Chelimsky E., 1985, *Program evaluation: patterns and directions*, Washington, D.C.: American Society for Public Administration.

Chelimsky E., 1997, „The Coming Transformation in Evaluation", w: E. Chelimsky, W.R. Shadish (red.), *Evaluation for the 21st century: a handbook* (s. 1–26). Thousand Oaks, CA-London: Sage.

Chen H.-T., 1990, *Theory-driven evaluations*, Newbury Park, CA: Sage Publications.

Chen H.-T., 2005, „Theory-driven evaluation", w: S. Mathison (red.), *Encyclopedia of evaluation* (s. 415–419). Thousand Oaks, CA-London: Sage.

Chen H.-T., Rossi P.H., 1983, „Evaluating with sense: The theory-driven approach", *Evaluation Review*, nr 7, 283–302.

Christie C.A., 2007, „The use of evaluation – latest empirical inside from US", *Seminars in the framework of Academy of Evaluation – Koźmiński School of Management*, Warsaw, December 2007.

Christie C.A., Alkin M.C., 1999, „Further reflections on evaluation misutilization", *Studies in Educational Evaluation*, nr 25, 1–10.

COUNCIL OF THE EUROPEAN ECONOMIC COMMUNITY, 1988, *Council Regulation (EEC) No 2052/88 of 24 June 1988 on the tasks of the Structural Funds and their effectiveness and on coordination of their activities between themselves and with the operations of the European Investment Bank and the other existing financial instruments*, 2052/1988, Official Journal of the European Communities L 185, 15/07/1988, p.0009–0020.

COUNCIL OF THE EUROPEAN ECONOMIC COMMUNITY, 1993, *Council Regulation (EEC) No 2082/93 of 20 July 1993 amending Regulation (EEC) No 4253/88 laying down provisions for implementing Regulation (EEC) No 2052/88 as regards coordination of the activities of the different Structural Funds between themselves and with the operations of the European Investment Bank and the other existing financial instruments*, 2082/1993, OJ L 193, 31/07/1993 pp. 0020–0033.

COUNCIL OF THE EUROPEAN UNION, 1999, *Council Regulation (EC) No 1260/1999 of 21 June 1999 laying down General Provisions on the Structural Funds*, 1260/1999, OJ L161, 26.6.1999.

COUNCIL OF THE EUROPEAN UNION, 2002, *Council Regulation (EC, Euratom) No 1605/2002 of 25 June 2002 on the Financial Regulation applicable to the general budget of the European Communities*, OJ L 248, 16.09.2002.

COUNCIL OF THE EUROPEAN UNION, 2005, *Proposal for a Council Regulation laying down General Provisions on the European Regional Development Fund, the European Social Fund and the Cohesion Fund*, Interinstitutional File 2004/0163 (AVC), CADREFIN 269, Brussels 21 December 2005.

COUNCIL OF THE EUROPEAN UNION, 2006, *Council Regulation (EC) No 1083/2006 of 11 July 2006 laying down General provisions on the European Regional Development Fund, the European Social Fund and the Cohesion Fund*, 1083/2006, OJ L210, 31.7.2006.

Cousins B.J., 2003, „Utilization effects of participatory evaluation", w: T. Kellaghan, D.L. Stufflebeam, L.A. Wingate (red.), *International handbook of educational evaluation*, s. 2 v. (ix, 1061 p.). Dordrecht-London: Kluwer Academic.

Cousins B.J., 2004, „Commentary: Minimizing evaluation misuse as principled practice", *American Journal of Evaluation*, nr 25(3), 391–397.

Cousins B.J., Leithwood K.A., 1986, „Current empirical research in evaluation utilization", *Review of Educational Research*, nr 56(3), 331–364.

Cousins B.J., Leithwood K.A., 1993, „Enhancing Knowledge Utilization as a Strategy for School Improvement", *Science Communication*, nr 14(3), 305–333.

Cousins B.J., Shulha L.M., 2006, „A comparative analysis of evaluation utilization and its cognate fields of inquiry: current issues and trends", w: I. Shaw, J.C. Greene, M.M. Mark (red.), *Handbook of evaluation: policies, programs and practices* (s. 266–91). London: Sage.

Cousins J.B., Whitmore E., 1998, „Framing participatory evaluation", *New Directions for Evaluation*, nr 80, 5–23.

Creswell J.W., 1994, *Research design: qualitative & quantitative approaches*, Thousand Oaks, CA: Sage Publications.

Cronbach L.J., Associates, 1981, *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco: Jossey-Bass Publications.

Cronbach L.J., Shapiro K., 1982, *Designing evaluations of educational and social programs*, San Francisco-London: Jossey-Bass.

CSF Evaluation Unit, 1998, *Review of Ongoing Evaluation Function in the Community Support Framework (CSF) for Ireland 1994–1999*, Dublin: CSF Evaluation Unit.

Dahler-Larsen P., 2005, „Evaluation in Public Management", w: E. Ferlie, L.E. Lynn, C. Pollitt (red.), *The Oxford handbook of public management* (s. 615–639). Oxford: Oxford University Press.

Davidson E.J., 2005, *Evaluation methodology basics: the nuts and bolts of sound evaluation*, Thousand Oaks, CA: Sage Publications.

Derlien H.U., 1990, „Genesis and Structure of Evaluation Efforts in Comparative Perspective", w: R.C. Rist (red.), *Program evaluation and the management of government: patterns and prospects across eight nations*, s. ix, 184. New Brunswick, N.J.-London: Transaction Publishers.

Dinan D., 1999, *Ever closer union: An introduction to European integration*, Basingstoke: Macmillan.

Dunsire A., 1995, „Administrative theory in the 1980s: a viewpoint", *Public Administration*, nr 73, 17–40.

Ekblom P., 2002, „From the Source to the Mainstream is Uphill: The Challenge of Transferring Knowledge of Crime Prevention Through Replication, Innovation and Anticipation", *Crime Prevention Studies*, nr 13, 131–203.

EPEC, 2005, *Study on the use of evaluation results in the Commission. Final report*, European Policy Evaluation Consortium (EPEC), Paris.

Eser T.W., Nussmueller E., 2006, „Mid-term Evaluations of Community Initiatives under European Union Structural Funds: a process between accounting and common learning", *Regional Studies*, nr 40(2), 249–258.

European Commission, 1993, *Project Cycle Management Manual. Integrated apporach and logical framework*, Evaluation Unit.

European Commission, 1995a, *MEANS Handbook 1 – Organising intermediate evaluation in the context of Partnership*, DG XVI, Brussels: Commission of the European Communities.

European Commission, 1995b, *MEANS Handbook 2 – Identifying the structuring effects of community interventions*, DG XVI, Brussels: Commission of the European Communities.

European Commission, 1995c, *MEANS Handbook 3 – Measuring the employment effects of Community Structural Interventions*, DG XVI, Brussels: Commission of the European Communities.

European Commission, 1995d, *MEANS Handbook 4 – Applying the multi-criteria method to the evaluation of Structural Programmes*, DG XVI, Brussels: Commission of the European Communities.

European Commission, 1995e, *MEANS Handbook 5 – Quality assessment of evaluation reports: a framework*, DG XVI, Brussels: Commission of the European Communities.

European Commission, 1995f, *MEANS Handbook 6 – The ex post evaluation of Structural Funds Programmes: a methodological framework*, DG XVI, Brussels: Commission of the European Communities.

European Commission, 1995, *MEANS Handbook 7 – The ex ante evaluation of Structural Funds Programmes: a methodological framework*, DG XVI, Brussels: Commission of the European Communities.

European Commission, 1997, *Evaluating EU expenditure programmes: A guide for ex post and intermediate evaluation*, DG-Budget, Brussels: Office for Official Publications of the European Communities.

European Commission, 1999a, *Evaluation in the Commission. Annual Evaluation Review 1998*, DG-Budget, http://ec.europa.eu/budget/evaluation/Key_documents/evalguide_study_en.htm.

European Commission, 1999b, *MEANS collection: evaluating socio-economic programmes*, DG-Regio, Luxembourg: Office for Official Publications of the European Communities.

European Commission, 2000a, *Communication from the Commission to the Member States of 28 April 2000 laying down guidelines for a Community initiative concerning trans-European cooperation intended to encourage harmonious and balanced development of the European territory*, 2000/C143.

European Commission, 2000b, *Evaluation in the Commission. Annual Evaluation Review 1999*, DG-Budget.

European Commission, 2000c, *Focus on Results: Strengthening Evaluation of Commission Activities*, SEC2000/1051.

European Commission, 2000d, *Implementing Activity-Based Management in the Commission*, European Commission.

European Commission, 2000e, *Methodological Working Paper 3 – Indicators for Monitoring and Evaluation: An indicative methodology*, DG-Regio, Luxembourg: The New Programming Period 2000–2006 Methodological Working Papers Office for Official Publications of the European Communities, http://ec.europa.eu/regional_policy/sources/docoffic/working/sf2000_en.htm.

European Commission, 2000f, *Methodological Working Paper 7 – Ex Ante Evaluation and Indicators for INTERREG (Strand A and B)*, DG-Regio, Luxembourg: The New Programming Period 2000–2006 Methodological Working Papers Office for Official Publications of the European Communities, http://ec.europa.eu/regional_policy/sources/docoffic/working/sf2000_en.htm.

European Commission, 2001a, *Evaluation in the Commission. Annual Evaluation Review 2000*, DG-Budget.

European Commission, 2001b, *A guide to evaluation procedures and structures currently operational in the Commission's External Co-operation programmes*, Europeaid.

European Commission, 2001c, *Project Cycle Management Manual*, Europeaid.

European Commission, 2001d, *Project Cycle Management Training Courses Handbook*, Europeaid.

European Commission, 2001e, *Putting Strategic Evaluation into practice within Commission*, C2001/3661.

European Commission, 2001f, *White Paper on European Governance – Report of the Working Group Evaluation & Transparency*, Working Group 2B, White Paper on European Governance – Working Area no2 Handling the Process of Producing and Implementing Community Rules, http://europa.eu.int/comm/governance/index_en.htm.

European Commission, 2002a, *Evaluation in the Commission. Annual Evaluation Review 2001*, DG-Budget.

European Commission, 2002b, *Methodological Working Paper 8a – The Mid Term Evaluation of Structural Fund Interventions – INTERREG III*, DG-Regio, Luxembourg: The New Programming Period 2000–2006 Methodological Working Papers Office for Official Publications of the European Communities, http://ec.europa.eu/regional_policy/sources/docoffic/working/sf2000_en.htm.

European Commission, 2003, *Evaluation in the Commission. Annual Evaluation Review 2002*, DG-Budget.

European Commission, 2004a, *Evaluating EU activities – A practical guide for the Commission services*, DG-Budget, Luxembourg: Office for Official Publications of the European Communities.

European Commission, 2004b, *Project Cycle Management guidelines,* t. 1, Europeaid, Luxembourg: Office for Official Publications of the European Communities, http://ec.europa.eu/europeaid/evaluation/methodology/index_en.htm.

European Commission, 2004c, *Proposal for a Regulation of the European Parliament and of the Council laying down general provisions establishing a European Neighbourhood and Partnership Instrument*, COM(2004)628.

European Commission, 2005a, *Evaluation in the Commission. Annual Evaluation Review 2004*, DG-Budget.

European Commission, 2005b, *Overview of completed evaluations per ABB activity for the years 2002, 2003 and 2004*, DG-Budget, http://ec.europa.eu/budget/evaluation/Key_documents/evalguide_study_en.htm.

European Commission, 2006, *Evaluation in the Commission. Annual Evaluation Review 2005*, DG-Budget, Luxembourg: Office for Official Publications of the European Communities.

European Commission, 2008, *Evaluation quality assessment form and guide on scoring the criteria*, DG-Budget, http://ec.europa.eu/budget/documents/evaluation_en.htm#standards.

Ferlie E., Lynn L.E., Pollitt C., 2005, *The Oxford handbook of public management*. Oxford: Oxford University Press.

Fetterman D.M., 1996, *Empowerment evaluation: knowledge and tools for self-assessment and accountability*, Sage.

Fink A., 2005, *Conducting research literature reviews: from the Internet to paper. Arlene Fink*. Thousand Oaks, CA-London: Sage.

Fiorini A., Buscemi V., Angori F., 2006, *Study on Indicators for Monitoring Transnational and Interregional Cooperation Programmes*, Vienna: INTERACT Programme Secretariat.

Flyvbjerg B., 2006, „Five Misunderstandings About Case-Study Research", *Qualitative Inquiry*, nr 12(2), 219–245.

Forss K., Rebien C.C., Carlsson J., 2002, „Process use of evaluations. Types of use that procede lessons learned and feedback", *Evaluation*, nr 8(1), 29–45.

Forunier D.M., 2005, „Evaluation", w: S. Mathison (red.) *Encyclopedia of evaluation* (s. 139–140). Thousand Oaks, CA-London: Sage.

Frankfort-Nachmias C., Nachmias D., 2001, *Metody badawcze w naukach społecznych*, Warszawa: Wydawnictwo Zysk i S-ka.

Frechtling J.A., 2007, *Logic modeling methods in program evaluation*, San Francisco, California: Jossey-Bass.

Furubo J.-E., Rist R.C., Sandahl R., 2002, *International atlas of evaluation*, New Brunswick: Transaction Publishers.

GHK Consultants, 2006, *The Evaluation of Socio-Economic Development. The Evalsed Guide*, (2nd edition), www.evalsed.com, accessed 10.10.2006.

Goldsmith S., Eggers W.D., 2004, *Governing by network: the new shape of the public sector*. Washington, D.C.-Bristol: Brookings Institution-University Presses Marketing.

Gorzelak G., 1995, *Transformacja systemowa a restrukturyzacja regionalna*, Warszawa: Uniwersytet Warszawski.

Gorzelak G., Kozak M., Olejniczak K., Smętkowski M., 2006, *Doświadczenia Programu Współpracy Przygranicznej PHARE Polska-Niemcy 1999–2001. Lekcje dla Wschodu*, Warszawa: Władza Wdrażająca Program Współpracy Przygranicznej PHARE.

Gorzelak G., 2007, „Rozwój polskich regionów a polityka spójności Unii Europejskiej", w: G. Gorzelak (red.), *Polityka regionalna i lokalna w świetle badań EUROREG-u* (s. 12–34), Warszawa: Wydawnictwo Naukowe Scholar.

Górniak J., 2007, „Ewaluacja w cyklu polityk publicznych", w: S. Mazur (red.), *Ewolucja funduszy strukturalnych, perspektywa regionalna* (s. 11–24), Kraków: Instytut Ekonomiczny, Małopolska Szkoła Administracji Publicznej.

Górniak J., 2005, „Ewaluacja jako czynnik doskonalenia zarządzania strategicznego w administracji publicznej: projekt SAPER", *I konferencja ewaluacyjna – Ewaluacja programów o charakterze społeczno-gospodarczym finansowanych z funduszy strukturalnych*, Warszawa: Ministerstwo Gospodarki i Pracy & Polska Agencja Rozwoju Przedsiębiorczości, 16 września 2005, 83–101.

Greene J.C., 2003, Understanding Social Programs through Evaluation, w: N.K. Denzin, Y.S. Lincoln (red.), *Collecting and interpreting qualitative materials*, s. 590–617. Sage, Thousand Oaks, Calif.; London.

Greene J.C., 2006, „Evaluation, democracy, and social change", w: I.F. Shaw, J.C. Greene, M.M. Mark (red.), *The SAGE Handbook of Evaluation* (s. 118–140). Thousands Oaks-London: Sage Publications Ltd.

Gruening G., 2001, „Origin and theoretical basis of New Public Management", *International Public Management Journal*, nr 4, 1–25.

Guba E.G., Lincoln Y.S., 1989, *Fourth generation evaluation*, Newbury Park: Sage.

Gulick L., 1937, „Notes on theory of organization", w: L. Gulick, L. Urwick (red.), *Papers on the science of administration*, New York: Columbia University.

Haber A. (red.), 2007, *Ewaluacja ex post. Teoria i praktyka badawcza*, Warszawa: Polska Agencja Rozwoju Przedsiębiorczości.

Hart C., 1998, *Doing a literature review: releasing the social science research imagination*, London: Sage.

Heinrich C.J., 2003, „Measuring Public Sector Performance", w: G. Peters, J. Pierre (red.), *Handbook of Public Administration* (s. 25–37). London: Sage Publications.

Henry G.T., 2000, „Why Not Use?", *New Directions for Evaluation*, nr 88, 85–98.

Henry G.T., 2001, „How modern democracies are shaping evaluation and the emerging challenges for evaluation", *American Journal of Evaluation*, nr 22(3), 419–429.

Henry G.T., 2005, „Realist evaluation", w: S. Mathison (red.), *Encyclopedia of evaluation* (s. 359–362). Thousand Oaks, CA-London: Sage.

Henry G.T., Julnes G., 1998, „Values and Realist Evaluation", *New Directions for Evaluation*, nr 78, 53–71.

Henry G.T., Julnes G., Mark M.M., 1998, „Realist evaluation: An emerging theory in support of practice", *New Directions for Evaluation*, nr 78.

Henry G.T., Mark M.M., 2003, „Beyond use: understanding evaluation's influence on attitudes and actions", *American Journal of Evaluation*, nr 24(3), 293–314.

Herman J.L., Morris L.L., Fitz-Gibbon C.T., 1987, *Evaluator's handbook*, Newbury Park: Sage & University of California Los Angeles – Center for the Study of Evaluation.

Hood C., 1991, „A public management for all seasons?", *Public Administration*, nr 69(1), 3–19.

Hood C., Jackson M., 1991, „The New Public Management: A recipe for disaster", *Canberra Bulletin of Public Administration*, 1964 (May), 16–24.

House E., 2001, „Unfinished business: Causes and values", *American Journal of Evaluation*, nr 22, 309–15.

Huberman M., Thurler G.M., 1991, *De la recherche a la practique: Elements de base*. Berne: Peter Lang.

Hummelbrunner R., 2000, „A systems approach to Evaluation. Applications of Systems theory and Systems thinking in Evaluations", *4th European Evaluation Society Conference*, Lausanne, 12–14th of October 2000.

Hummelbrunner R., 2004, *On-going Evaluation INTERREG III A Programmes AT-CZ, AT-HU, AT-SI, AT-SK. Synthesis Cross-programme topics*, Graz: Federal Chencellery.

Hummelbrunner R., 2006, *Process Monitoring of Impacts. Proposal for a new approach to monitor the implementation of "Territorial Co-operation" programmes*, Graz: INTERACT-Managing Transition Point.

Hummelbrunner R., Omersel-Weeks B., 2005, *Update of the Mid-Term Evaluation of the Community Initiative Programme INTERREG IIIA Austria-Slovenia*, Graz-Ljubljana: Managing Authority of INTERREG IIIA Austria-Slovenia Programme.

INTERACT, 2004, *Handbook on Financial Support to Draft INTERREG Project Applications. A tool for better use of European Regional Development Funds commitment*, Vienna: INTERACT Point Tool Box.

INTERACT, 2005a, *Programme Summary of the Austria-Czech Republic INTERREG IIIA Programme*, Maastricht: INTERACT Point Tool Box.

INTERACT, 2005b, *Programme Summary of the Austria-Hungary INTERREG IIIA Programme*, Maastricht: INTERACT Point Tool Box.

INTERACT, 2005c, *Programme Summary of the Austria-Slovakia INTERREG IIIA Programme*, Maastricht: INTERACT Point Tool Box.

INTERACT, 2005d, *Programme Summary of the Austria-Slovenia INTERREG IIIA Programme*, Maastricht: INTERACT Point Tool Box.

INTERACT, 2005e, *Programme Summary of the INTERREG IIIB North West Europe (NWE) Programme*, Vienna: INTERACT Point Tool Box.

Izdebski H., 2007, „Od administracji publicznej do *public governance*", *Zarządzanie Publiczne*, nr 1(1), 7–20.

Hausner J., 2007, „Czy należy pomagać regionom słabo rozwiniętym?", *Sympozjum Regional Studies Association – Sekcja Polska*, Warszawa, 18.04.2007.

Johnson B., 1998, „Towards a theoretical model of evaluation utilization", *Evaluation and Program Planning*, nr 21, 93–110.

Jones L.R., Thompson F., 1997, „The five Rs of the New Public Management", *Advances in International Comparative Management*, Suplement 3, 15–45.

Kardorff E.V., 2004, „Qualitative Evaluation Research", w: U. Flick, E.V. Kardorff, I. Steinke (red.), *A Companion to Qualitative Research* (s. 138–142). London: Sage.

Kettl D.F., 1997, „The global revolution in Public Management: driving themes, missing links", *Journal of Policy Analysis and Management*, nr 16(3), 446–462.

Kettl D.F., 2006, „Modernising Government: The way forward – a comment", *International Review of Administrative Sciences*, nr 72(3), 313–317.

Kierzkowski T., 2002a, „Evaluation of socio-economic programmes in Poland – current situation and future challenges", *Annual Conference of Regional Studies Association "Evaluation and EU Regional Policy"*, Aix en Province: Regional Studies Association, June 2002.

Kierzkowski T., 2002b, *Ocena (ewaluacja) programów i projektów o charakterze społeczno-gospodarczym w kontekście przystąpienia Polski do Unii Europejskiej*, Warszawa: Polska Agencja Rozwoju Przedsiębiorczości.

King J.A., 1998, „Making sense of participatory evaluation practice", *New Directions for Evaluation*, nr 80, 7–92.

King J.A., 2005, „Participatory evaluation", w: S. Mathison (red.), *Encyclopedia of evaluation* (s. 291–294). Thousand Oaks, CA.-London: Sage.

Kirkhart K.E., 2000, „Reconceptualizing Evaluation Use: An Integrated Theory of Influence", *New Directions for Evaluation*, nr 88, 5–25.

Kojan M., Navratilova K., 2005, *Update of the Mid-Term Evaluation of the Community Initiative Programme INTERREG IIIA Austria-Czech Republic*, Vienna-Praha: Managing Authority of INTERREG IIIA Austria-Czech Republic Programme.

Korporowicz L. (red.), 1997, *Ewaluacja w edukacji*, Warszawa: Oficyna Naukowa.

Korporowicz L., 2000, „Ewaluacja – zaproszenie do rozwoju", *Edukacja i Dialog*, nr 6(119).

Kozak M., 2007, „Managing European Regional Policy in Poland in the initial period after the accession", *Regional and Local Studies*, numer specjalny, 67–89.

Krajowa Jednostka Oceny, 2005, *Ewaluacja Narodowego Programu Rozwoju i Programów Operacyjnych w Polsce. Poradnik*, Warszawa: MGiP – KJO.

Langbein L., Felbinger C.L., 2006, *Public Program Evaluation. A statistical guide*. New York-London: M.E.Shape.

Lavis J., Ross S., McLeod C., Gildiner A., 2003, „Measuring the impact of health research", *Journal of Health Services Research and Policy*, nr 8(3), 165–170.

Leeuw F.L., 2002, „Evaluation in Europe 2000. Challenges to the Growth Industry", *Evaluation*, nr 8(1), 5–12.

Leeuw F.L., Rist R.C., Sonnichsen R.C., 1994, *Can governments learn?: comparative perspectives on evaluation & organizational learning*, New Brunswick-London: Transaction Publishers.

Leviton L.C., 2003, „Evaluation use: advances, challenges and applications", *American Journal of Evaluation*, nr 24(4), 525–535.

Leviton L.C., Hughes E.F.X., 1981, „Research on the utilization of evaluations: a review and synthesis", *Evaluation Review*, nr 5(4), 525–548.

Lundvall B.-Ĺ., Johnson B., 1994, „The Learning Economy", *Journal of Industry Studies*, nr 1(2), 23–42.

Lynn L.E., 2003, „Public Management", w: B.G. Peters, J. Pierre (red.), *Handbook of Public Administration* (s. 14–23). London: Sage.

Malan J., 2004, „Benchmarking Evaluation Capacity in the New Member States", *Open Days, Session RG22, 'Evaluation and Monitoring'*, Brussels.

Mark M.M., Henry G.T., 2004, „The mechanisms and outcomes of evaluation influence", *Evaluation*, nr 10(1), 35–57.

Mark M.M., Henry G.T., Julnes G., 2000, *Evaluation: an integrated framework for understanding, guiding, and improving policies and programs*, San Francisco: Jossey-Bass.

Martin R., 1998, *Regional Policy in the EU – Economic Foundations and Reality*, Brussels: Centre for European Policy Studies.

Mathison S., 2005, *Encyclopedia of evaluation*, Thousand Oaks, CA-London: Sage.

Mazur S., 2006, „Historia administracji publicznej", w: J. Hausner (red.), *Administracja publiczna* (s. 41–62). Warszawa: PWN.

Mazur S. (red.), 2007, *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Kraków: Akademia Ekonomiczna w Krakowie – Małopolska Szkoła Administracji Publicznej.

Metcalfe L., Richards S., 1990, *Improving public management*, London: Sage.

Ministerstwo Rozwoju Regionalnego, 2007, *Ewaluacja* ex ante *– podsumowanie doświadczeń administracji polskiej*, Ministerstwo Rozwoju Regionalnego.

Mizerek H. (red.), 1997, *Ewaluacja w szkole*, Olsztyn: Wydawnictwo MG.

Morabito S.M., 2002, „Evaluator roles and strategies for expanding evaluation process influence", *American Journal of Evaluation*, nr 23(3), 321–330.

Morgan G., 2006, *Images of organization*, Thousand Oaks-London: Sage.

Morgan M.G., Fischhoff B., Bostrom A., Atman C.J., 2002, *Risk communication: a mental models approach*, Cambridge: Cambridge University Press.

Morris L.L., Freeman M.E., Fitz-Gibbon C.T., 1987, *How to communicate evaluation findings*, Newbury Park-London: Sage.

Moussis N., 2003, *Access to European Union. 12 edition*, Brussels: European Study Service.

Nowak A.Z., 2002, *Integracja europejska. Szansa dla Polski?*, Warszawa: Polskie Wydawnictwo Ekonomiczne.

Nowak A.Z., 2005, „Fundusze Strukturalne Unii Europejskiej jako czynnik mobilizacji gospodarki polskiej", w: A. Adamczyk, J. Borkowski (red.), *Regionalizm, polityka regionalna i Fundusze Strukturalne w Unii Europejskiej* (s. 69–78). Warszawa.

Nutley S., Walter I., Davies H.T.O., 2003, „From Knowing to Doing. A Framework for Understanding the Evidence-Into-Practice Agenda", *Evaluation*, nr 9(2), 125–48.

NWE Joint Technical Secretariat, 2002, *Community Initiative Programme. Sustainable Territorial Development in the North West of Europe – Towards Long Term Transnational Co-operation. INTERREG IIIB Community Initiative North West Europe*, Lille: NWE Joint Technical Secretariat.

OECD, 1995, *Governance in transition: public management reforms in OECD countries*, OECD.

OECD, 1997, *Managing across levels of government*, OECD.

OECD, 1998, *Best practice guidelines for evaluation*, PUMA, OECD.

OECD, 1999, *Improving Evaluation Practices. Best Practice Guidelines for Evaluation and Background Paper*, PUMA, OECD.

OECD, 2005, *Modernising government: the way forward*, Paris: OECD, 2005.

ÖIR-Managementdisenste GMBH, 2001a, *Austria-Czech Republic INTERREG IIIA. Community Initiative Programme*, Vienna: Joint Secretariat-ÖIR.

ÖIR-Managementdisenste GMBH, 2001b, *Austria-Hungary INTERREG IIIA. Community Initiative Programme*, Vienna: Joint Secretariat-ÖIR.

ÖIR-Managementdisenste GMBH, 2001c, *Austria-Slovakia INTERREG IIIA. Community Initiative Programme*, Vienna: Joint Secretariat-ÖIR.

ÖIR-Managementdisenste GMBH, 2001d, *Austria-Slovenia INTERREG IIIA. Programme Complement*, Vienna: Joint Secretariat-ÖIR.

ÖIR-Managementdisenste GMBH, 2002, *Austria-Slovakia INTERREG IIIA. Programme Complement*, Vienna: Joint Secretariat-ÖIR.

ÖIR-Managementdisenste GMBH, 2004a, *Austria-Czech Republic INTERREG IIIA. Programme Complement*, Vienna: Joint Secretariat-ÖIR.

ÖIR-Managementdisenste GMBH, 2004b, *Austria-Hungary INTERREG IIIA. Programme Complement*, Vienna: Joint Secretariat-ÖIR.

Olejniczak K., 2002, „Towards the evaluation culture in Poland – experiences and prospects", *Annual Conference of Regional Studies Association "Evaluation and EU Regional Policy"*, Aix en Province: Regional Studies Association, June 2002.

Olejniczak K., 2004, *Proces przeprowadzania ewaluacji ex ante i on-going w ramach NPR z uwzględnieniem etapów prac: planowanie, projektowanie i realizacja ewaluacji – wkład do procedur operacyjnych ewaluacji w Polsce*, Krajowa Jednostka Oceny, http://www.funduszestrukturalne.gov.pl/EWALUACJA+funduszy+w+Polsce/Dokumenty+ewaluacyjne/.

Olejniczak K., 2005, „Teoria i praktyka ewaluacji w świetle doświadczeń europejskich", *Studia Regionalne i Lokalne*, nr 4(22).

Olejniczak K., 2006, „Territorial Cooperation in the European Union – INTERREG lessons for the New Programming Period", w: R. Michie (red.), *IQ-Net Ten: 1996–2006 – Ten years of improving the quality of structural funds programming through exchange of experiences* (s. 29–32). Glasgow: European Policies Research Centre, http://www.eprc.strath.ac.uk/iqnet/default.cfm.

Olejniczak K., 2007, „Ewaluacja jako narzędzie zarządzania w sektorze publicznym", w: G. Gorzelak, A. Tucholska (red.), *Rozwój, region, przestrzeń* (s. 361–96). Warszawa: Ministerstwo Rozwoju Regionalnego.

Osborne D., Gaebler T., 1992, *Reinventing government: how the entrepreneurial spirit is tansforming the public sector*. Reading, MA.:Addison-Wesley Pub. Co.

Osborne S.S., 2006, „The New Public Governance?", *Public management review*, nr 8(3), 377.

Ostrom E., Ostrom V., 1971, „Public choice: a different approach to the study of public administration", *Public Administration Review*, nr 31, 203–216.

Owen J., Lambert F., 1998, „Evaluation and the information needs of organizational leaders", *American Journal of Evaluation*, nr 19, 355–365.

Owen J.M., 2007, *Program evaluation: forms and approaches. 3rd edition*, London: The Guilford Press.

Owen J.M., Rogers P.J., 1999, *Program evaluation: forms and approaches*, London: Sage.

Oxford University Press, 2002, Oxford English Dictionary, Oxford: Oxford University Press.

Paddock S.C., 1998, „Evaluation", w: J.M. Shafritz (red.), *International Encyclopedia of Public Policy and Administration* (s. 818–823). Boulder, CO: Westview Press.

Patton M.Q., 1997, *Utilization-focused evaluation: the new century text* ,Thousand Oaks, CA: Sage Publications.

Patton M.Q., 2004, „Evaluation Research", w: M.S. Lewis-Beck, A. Bryman, T.F. Liao (red.), *The Sage encyclopedia of social science research methods* (s. 337–340). Thousand Oaks, CA: Sage.

Patton M.Q., 2005, „Utilization-focused evaluation", w: S. Mathison (red.), *Encyclopedia of evaluation* (s. 429–32). Thousand Oaks, CA-London: Sage.

Pawson R., 2006, *Evidence-based policy: a realist perspective*, London: Sage.

Pawson R., Tilley N., 1997, *Realistic evaluation*, London: Sage.

Pawson R., Tilley N., 2005, „Realistic evaluation", w: S. Mathison (red.), *Encyclopedia of evaluation* (s. 362–7). Thousand Oaks, CA-London: Sage.

Peters T.J., Waterman R.H., Austin N., 1992, *Excellence: In search of excellence lessons from America's best-run companies*, New York: Random House, Inc.

Pollitt C., 1998, „Evaluation in Europe. Boom or bubble?", *Evaluation*, nr 4(2), 214–224.

Pollitt C., Bouckaert G., 2003, „Evaluating public management reforms: An international perspective", w: H. Wollmann (red.), *Evaluation in public-sector reform: concepts and practice in international perspective* (s. 12–35). Cheltenham: Edward Elgar.

Premfors R., 2006, „Modernising Government: The way forward – a comment", *International Review of Administrative Sciences*, nr 72(3), 333–335.

Preskill H., Torres R., 2000, „The learning dimension of evaluation use", *New Directions for Evaluation*, nr 88, 25–37.

Pricewaterhousecoppers, 2003, *Mid-Term Evaluation of the Community Initiative Programme INTERREG IIIA Ireland-Northern Ireland*, Monaghan: Special EU Programmes Body.

Pricewaterhousecoppers, 2005, *Update of the Mid-Term Evaluation of the Community Initiative Programme INTERREG IIIA Ireland-Northern Ireland*, Monaghan: Special EU Programmes Body.

Provus M.M., 1971, *Discrepancy evaluation for educational program improvement and assessment*, Berkeley, CA: McCutchan Pub. Corp.

Rainey H.G., Chun Y.H., 2005, „Public and Private Management Compared", w: E. Ferlie, L.E. Lynn, C. Pollitt (red.), *The Oxford handbook of public management* (s. 72–102). Oxford: Oxford University Press.

Reiner K., Karasz P., 2005, *Update of the Mid-Term Evaluation of the Community Initiative Programme INTERREG IIIA Austria-Slovakia*, Vienna-Bratislava: Managing Authority of INTERREG IIIA Austria-Slovakia Programme.

Research Voor Beleid International, 2005, *Update of the Mid-term Evaluation of the INTERREG IIIB North-West Europe Programme*, Leiden: Monitoring Committee of INTERREG IIIB North-West Europe Programme.

Rich R.F., 1977, „Uses of social science information by federal bureaucrats: knowledge for action versus knowledge for understanding", w: C.H. Weiss (red.), *Using social research in public policy making* (s. 199–211). Lexington, MA: D.C. Heath.

Rist R.C., 1990, *Program evaluation and the management of government: patterns and prospects across eight nations*, New Brunswick, N.J.-London: Transaction Publishers.

Rossi P.H., Freeman H.E., Lipsey M.W., 1999, *Evaluation: a systematic approach. 6th edition*, Thousand Oaks, CA-London: Sage Publications.

Rossi P.H., Lipsey M.W., Freeman H.E., 2004, *Evaluation: a systematic approach*. Thousand Oaks, CA-London: Sage Publications.

Rupprecht Consult, 2003, *Mid-Term Evaluation of the INTERREG IIIB North-West Europe Programme*, Köln: Monitoring Committee of INTERREG IIIB North-West Europe Programme.

Ryan K.E., Destefano L. (red.), 2000, *Evaluation as democratic process: Promoting inclusion, dialogue, and deliberation*, San Francisco: Jossey-Bass.

Scriven M., 1976, ”Prose and cons about goal-free evaluation”, w: G.V. Glass (red.), *Evaluation studies review annual*, t. 1, Bevery Hills, CA: Sage.

Scriven M., 1991, *Evaluation thesaurus*, Newbury Park-London: Sage.

ShadishW.R., Cook T.D., Campbell D.T., 2001, *Experimental and quasi-experimental designs for generalized causal inference*, Boston: Houghton Mifflin.

ShadishW.R., Cook T.D., Leviton L.C., 1991, *Foundations of program evaluation: theories of practice*, Newbury Park, CA: Sage Publications.

Shafritz J.M., Russell E.W., 2003, *Introducing public administration –3rd edition*. New York-London: Pearson Longman.

Shafritz J.M., Russell E.W., Borick C.P., 2007, *Introducing public administration*, New York-London: Pearson Longman.

Shulha L.M., Cousins B.J., 1997, „Evaluation Use: Theory, Research, and Practice Since 1986”, *Evaluation Practice*, nr 18(3), 195–208.

Smith M.K., 2005, Donald Schon: learning, reflection and change & Chris Argyris: theories of action, double-loop learning and organizational learning. The encyclopedia of informal education.

Smith P., 1995, „On the unintended consequences of publishing performance data in the public sector”, *International Journal of Public Administration*, nr 18(2&3), 277–310.

Special EU Programmes Body, 2001, *Ireland-Northern Ireland INTERREG IIIA. Operational Programme*, Monaghan: Special EU Programmes Body.

Special EU Programmes Body, 2004, *Follow up of the recommendations from MTE Ireland-Northern Ireland INTERREG IIIA*, Monaghan: Special EU Programmes Body.

Stake R., 2000, „A modest commitment to the promotion of democracy”, *New Directions for Evaluation*, nr 85, 97–107.

Stake R., 2005, „Responsive evaluation”, w: S. Mathison (red.), *Encyclopedia of evaluation* (s. 376–9). Thousand Oaks, CA-London: Monaghan.

Stead D., Waterhout B., 2006, „The influence of the ESDP on European inter-regional cooperation programmes and projects: A retrospective and prospective assessment”, *Annual Conference of Regional Studies Association „Shaping EU Regional Policy: Economic Social and Political Pressures”*, Leuven: Regional Studies Association.

Steen T., Schaap D., 2004, „Europeanization of the civil service: towards a European identity?”, *EGPA Congress, Study Group on Public Personnel Policies*, Ljubljana: European Group of Public Administration.

Stern E., 2005, „What do we know about the utilization of evaluation?”, *I konferencja ewaluacyjna – Ewaluacja programów o charakterze społeczno-gospodarczym finansowanych z funduszy strukturalnych*, Warszawa: Ministerstwo Gospodarki i Pracy & Polska Agencja Rozwoju Przedsiębiorczości, 16 września 2005, 7–29.

Stewart J., Ranson S., 1994, „Management in the Public Domain”, w: D. McKevitt, A. Lawton (red.), *Public sector management: theory, critique and practice* (s. 54–70). London: Sage, Open University.

Stufflebeam D.L., 1994, „Empowerment Evaluation, Objectivist Evaluation, and Evaluation Standards: Where the future of evaluation should not go and where it needs to go", *Evaluation Practice*, nr 15(3), 321–338.

Stufflebeam D.L., 2000a, „The CIPP model in program evaluation", w: D.L. Stufflebeam, G.F. Madaus, T. Kellaghan (red.), *Evaluation models: viewpoints on educational and human services evaluation*, Boston-London: Kluwer Academic Publishers.

Stufflebeam D.L., 2000b, „Lessons in Contracting for Evaluations", *American Journal of Evaluation*, nr 21(3), 293–314.

Stufflebeam D.L., 2001a, „Evaluation Models", *New Directions for Evaluation*, nr 89, 7–92.

Stufflebeam D.L., 2001b, „The Metaevaluation Imperative", *American Journal of Evaluation*, nr 22(2), 183–209.

Stufflebeam D.L., 2005, „CIPP Model", w: S. Mathison (red.), *Encyclopedia of evaluation* (s. 60–5). Thousand Oaks, CA-London: Sage.

Swiss J.E., 1992, „Adapting Total Quality Management (TQM) to government", *Public Administration Review*, nr 52(4), 356–362.

Szlachta J., 1997, *Programowanie rozwoju regionalnego w Unii Europejskiej*, Warszawa: Wydawnictwo Naukowe PWN.

Szlachta J., 2000, „Uwarunkowania wyboru modelu polityki rozwoju regionalnego związane z Unią Europejską", w: A. Pyszkowski, J. Szlachta, J. Szomburg (red.), *Model instytucjonalny polityki rozwoju regionalnego w Polsce*, Gdańsk-Warszawa: Instytut Badań nad Gospodarką Rynkową.

Szlachta J., 2007a, „Wprowadzenie", w: A. Haber (red.), *Ewaluacja* ex post*. Teoria i praktyka badawcza*, Warszawa: Polska Agencja Rozwoju Przedsiębiorczości.

Szlachta J., 2007b, „System ewaluacji funduszy strukturalnych w Polsce", w: S. Mazur (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, (s. 75–89), Kraków: Uniwersytet Ekonomiczny – Małopolska Szkoła Administracji Publicznej.

Tavistock Institute, GHK, IRS, 2003, *The Evaluation of Socio-Economic Development. The Evalsed Guide (1st edition)*, www.evalsed.com, downloaded 21.08.2006.

Torres R., Preskill H., 2001, „Evaluation and organizational learning: past, present and future", *New Directions for Evaluation*, nr 22(3), 387–395.

Toulemonde J., 2000, „Evaluation culture(s) in Europe: differences and convergence between national practices", *The Danish Evaluation Society Annual Conference*.

Tym R., Partners, 2005, *Impact of the EU INTERREG III Programme in the South West Region*, Cornvall: Tym Consults.

Valovirta V., 2002, „Evaluation utilization as argumentation", *Evaluation*, nr 8(1), 60–80.

Walliman N.S.R., 2006, *Social research methods*, London: Sage.

Weiss C.H., 1998a, *Evaluation: methods for studying programs and policies*. Upper Saddle River, N.J.: Prentice Hall.

Weiss C.H., 1998b, „Have We Learned Anything New About the Use of Evaluation?", *American Journal of Evaluation*, nr 19(1), 21–33.

Weiss C.H., 2005, „Evaluation for decisions: is anybody there? Does anybody care?", w: E. Stern (red.), *Evaluation research methods* (s. 287–299). London: Sage Publications.

Weiss C.H., Bucuvalas M.J., 1980a, *Social science research and decision-making*. New York-Guildford: Columbia University Press.

Weiss C.H., Bucuvalas M.J., 1980b, „Truth Tests and Utility Tests: decision-makers' frame of reference for social science research", *American Journal of Sociological Review*, nr 45, 302–313.

Weiss C.H., Murphy-Graham E., Birkeland S., 2005, „An Alternate Route to Policy Influence How Evaluations Affect D.A.R.E.", *American Journal of Evaluation*, nr 26(1), 12–30.

Widmaier B., 2000, „Knowledge, innovation systems and knowledge-based economy in Europe", w: A. Kukliński (red.), *The knowledge-based economy. The European Challenges of the 21st century* (s. 39–46). Warsaw: State Committee for Scientific Research.

Williams K., de Laat B., Stern E., 2002, *The Use of Evaluation in the Commission Services. Final Report*, Paris: Technopolis France.

Wilson W., 1887, „The study of Administration", *Political Science Quarterly*, nr 2(2), 481–506.

Wolf A., 2000, „Trends in public administration – a practitioner's view", *International Review of Administrative Sciences*, nr 66, 689–96.

Wollmann H., 2003, „Evaluation in public-sector reform: Towards a 'third wave' of evaluation?", w: H. Wollmann (red.), *Evaluation in public-sector reform: concepts and practice in international perspective* (s. 1–11). Cheltenham: Edward Elgar.

Worthen B.R., 1990, „Program evaluation", w: H.J. Walberg, G.D. Haertel (red.), *The International encyclopedia of educational evaluation* (s. 42–47). Oxford-New York: Pergamon Press.

Zawicki M., 2007, „Trzy płaszczyzny ewaluacji – z praktyki polskiej administracji publicznej", *Zarządzanie Publiczne*, nr 2(2), 113–124.

Ewaluacja jest standardem nowoczesnego, efektywnego zarządzania w sektorze publicznym. Jest też wymogiem i powszechną praktyką programów funduszy strukturalnych. W niniejszej książce autor podejmuje problem rzeczywistego wykorzystania i użyteczności badań ewaluacyjnych. Mechanizmy uczenia się dzięki ewaluacji i determinanty tego procesu Karol Olejniczak weryfikuje w empirycznym badaniu ewaluacji średniookresowej programów współpracy terytorialnej (INTERREG). Wnioski i rekomendacje płynące z tej analizy mogą okazać się bardzo przydatne dla rozwijającej się praktyki badań ewaluacyjnych w Polsce.

Oprócz analizy mechanizmów wykorzystania badań ewaluacyjnych Czytelnik znajdzie w książce:

- systematyczny przegląd definicji i typologii ewaluacji,
- omówienie jej genezy i funkcji w sektorze publicznym,
- bogatą syntezę koncepcji wykorzystania ewaluacji,
- przegląd praktyk ewaluacji w Unii Europejskiej.

Jest to jedna z pierwszych w Polsce publikacji przedstawiających w sposób kompleksowy problematykę ewaluacji programów publicznych. Książka z pewnością zainteresuje pracowników administracji publicznej – szczególnie instytucji zajmujących się funduszami strukturalnymi, naukowców i prywatnych konsultantów wykonujących badania ewaluacyjne, a także studentów administracji publicznej, gospodarki przestrzennej i europeistyki.

ISBN 978-83-7383-306-7

9 788373 833067