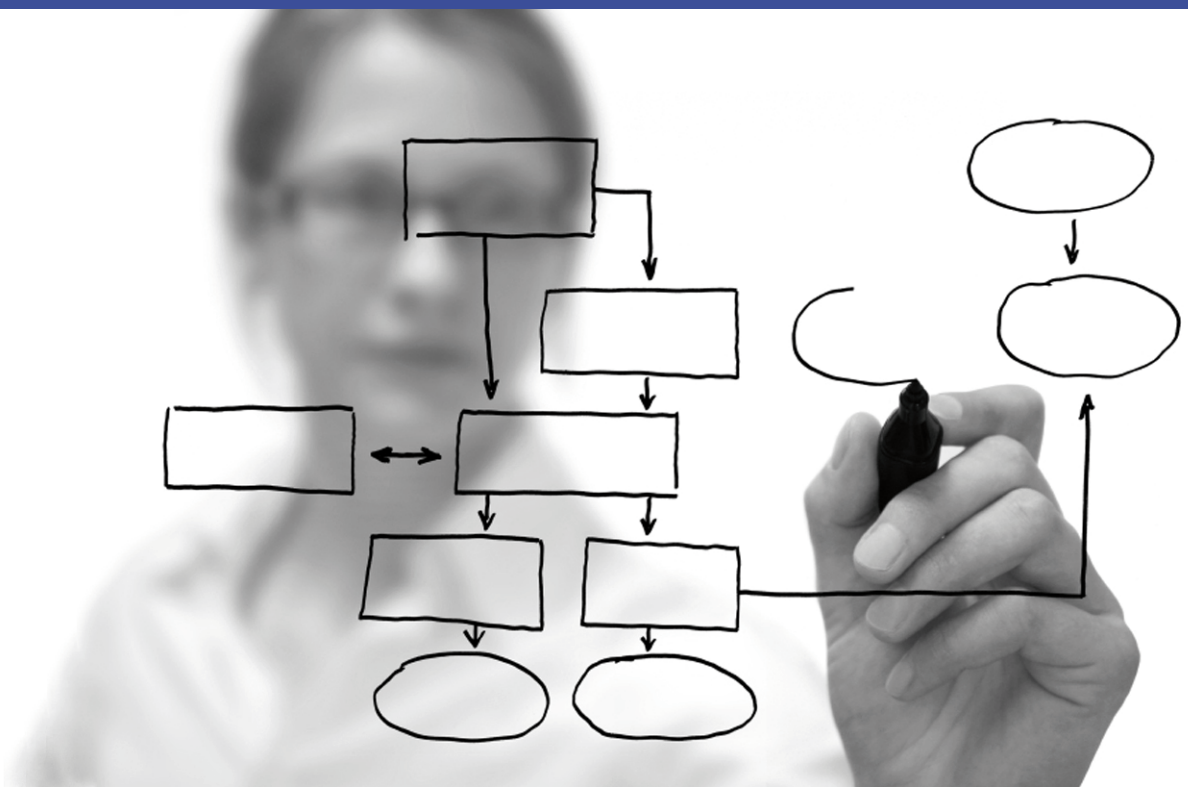
EUROREG

Raporty i analizy EUROREG 2/2009

Praktyka ewaluacji efektów programów rozwoju regionalnego - studium porównawcze



Karol Olejniczak

NARODOWA STRATEGIA SPÓJNOŚCI

MINISTERSTWO ROZWOJU REGIONALNEGO

UNIA EUROPEJSKA
EUROPEJSKI FUNDUSZ ROZWOJU REGIONALNEGO

Projekt realizowany w ramach konkursu dotacji orga nizowanego przez Ministerstwo Rozwoju Regionalnego, współfinansowanego ze środków Unii Europejskiej w ramach Programu Operacyjnego Pomoc Techniczna

Projekt został zrealizowany w ramach konkursu dotacji organizowanego przez Ministerstwo Rozwoju Regionalnego, współfinansowanego ze środków Unii Europejskiej w ramach Programu Operacyjnego Pomoc Techniczna

Copyrights © by Karol Olejniczak, 2009

Rozdział 2.3 © by Adam Płoszaj, 2009

Aneks 2 © by Kai Bohme, Richard Boyle, Martin Ferry, Lukas Wortmann, 2009

Warszawa, 2009

STRESZCZENIE

Celem badania było zidentyfikowanie praktyk z zakresu ewaluacji efektów programów rozwoju regionalnego. Dostarczając inspiracji metodologicznych badanie miało pomóc we wzmocnieniu krajowego potencjału ewaluacyjnego.

Badanie oparto na dwóch głównych elementach. Pierwszym była analiza porównawcza metodologii wykorzystanych w 100 ewaluacjach efektów unijnych programów rozwoju regionalnego. Raporty pochodziły z 4 krajów o wysokiej kulturze ewaluacji (Szwecja, Wielka Brytania, Irlandia, Niemcy) oraz z Polski. Elementem drugim był kompleksowy przegląd literatury światowej oraz wizyty studyjne w dwóch ośrodkach akademickich U.S.A. specjalizujących się w ewaluacji.

Analiza raportów z 100 ewaluacji pokazała, że praktyka ewaluacji Funduszy Strukturalnych dostarcza bardzo ograniczonych inspiracji. *Gros* ze zbadanych ewaluacji efektów programów UE, okazała się mieć bardzo ubogi zakres i metodologię. Ich trzy podstawowe słabości to: 1.brak spójnej, osadzonej w literaturze tematu całościowej koncepcji badawczej; 2. ograniczanie "skrzynki" metod do kilku podstawowych elementów; 3. brak analizy relacji przyczynowej - efektów interwencji wychodzących poza prosty opis zaplanowanych produktów, ku rzeczywistym zmianom wywołanych programem, w tym zmianom relacji.

Natomiast przegląd literatury światowej i wizyty studyjne były bardziej owocne i przyniosły dwie inspiracje metodologiczne. Pierwszą jest "ewaluacja wspierana teorią" oraz modele logiczne. Pozwalają one opisać i badać programy w formie hipotezy (*jeśli.. to..*). To zaś umożliwia jasne przedstawienie założeń programu na temat mechanizmu oczekiwanej zmiany społecznej, stworzenia solidnej podstawy koncepcyjnej badania oraz zbudowania spójnej narracji raportu końcowego. Inspiracją drugą jest "analiza sieci". Daje ona możliwość analizy, przedstawienia w formie graficznej i oceny złożonych relacji elementów połączonych ze sobą w system. Chodzi zarówno o powiązania fizyczne - przestrzenne, jak i miękkie relacje współpracy i wymiany informacji między osobami czy organizacjami.

W podsumowaniu raportu zarysowano dalsze pola działań badawczych: 1. testowanie i adaptację nowych podejść i metod do polskich warunków, w odniesieniu do różnych typów programów, w badaniach o różnej skali; 2.poszukiwanie inspiracji na gruncie europejskiej praktyki ewaluacyjnej - wśród badań z krajów UE dotyczących programów narodowych 3.śledzenie światowych debat i dyskursów na temat metodologii ewaluacji. Wiarygodne i adekwatne metodologie ewaluacji efektów umożliwią budowanie wiedzy na temat mechanizmów rozwoju regionalnego Polski i skutecznych metod ich wspierania.

SPIS TREŚCI

Wprowadzenie 6
Cel i uzasadnienie badania 6
Przedmiot badania 8
Zespół badawczy i metodologia 10

1 Praktyka ewaluacji efektów Funduszy Strukturalnych 15
1.1 Kontekst - kultury ewaluacji w Unii Europejskiej 15
1.1.1 Płaszczyzna 1: Krajowe kultury ewaluacji 16
1.1.2 Płaszczyzna 2: Kultura ewaluacji w Komisji Europejskiej 20
1.1.3 Płaszczyzna 3: Kultura ewaluacji funduszy strukturalnych w 4 krajach 26
1.2 Wyniki analizy metodologii 100 raportów z ewaluacji 30
1.2.1 Charakterystyka analizowanej populacji raportów 30
1.2.2 Tworzenie koncepcji i metodologii badań 33
1.2.3 Metody zbierania danych 36
1.2.4 Metody analizy i oceny 41
1.2.5 Orientacja wyników i rekomendacji z badań 44
1.3 Wnioski z analizy 46

2 Poszukiwanie inspiracji poza polem Funduszy Strukturalnych 50
2.1 Kontekst - wojna paradygmatów 51
2.2 "Ewaluacja wspierana teorią" i modele logiczne - pomoc w definiowaniu zmiany 55
2.2.1 Geneza i definicja "ewaluacji wspieranej teorią" 55
2.2.2 Geneza i definicje modeli logicznych 58
2.2.3 Przykłady zastosowań 61
2.2.4 Możliwości zastosowania w ewaluacjach rozwoju regionalnego 67
2.3 Analiza sieci - pomoc w ocenie dynamiki współzależności i relacji 69
2.3.1 Geneza i definicja 69
2.3.2 Przykłady zastosowań 71
2.3.3 Możliwości zastosowania w ewaluacjach rozwoju regionalnego 77

3 Podsumowanie 80

Bibliografia 86

Aneksy 95

Spis tabel

Tabela 1: Zakres ewaluacji efektów 10
Tabela 2: Krajowe kultury ewaluacji - ranking 21 państw 17
Tabela 3: Krajowa kultura ewaluacji - próba oceny Polski 19
Tabela 4. Unijna hierarchia celów - efektów - wskaźników 25
Tabela 5. Praktyka ewaluacji w polu funduszy strukturalnych - podsumowanie czterech krajów 27
Tabela 6: Przykłady łączenia typów czasowych ewaluacji 33
Tabela 7. Przykład wykorzystania modelu w koncepcji badania 36
Tabela 8. Przykład użyteczności baz danych monitoringowych 36
Tabela 9. Ciekawe przykłady metod zbierania danych 39
Tabela 10. Ciekawe przykłady metod analizy i oceny 43
Tabela 11. Pozytywizm vs. konstruktywizm 52
Tabela 12. Teorie programu - przykład Narodowego Planu Rozwoju Irlandii 1994-1999 57
Tabela 13. Teorie programu - przykład Narodowego Planu Rozwoju Irlandii 1994-1999 58
Tabela 14. Przykład teorii Programu Zarządzania Energią Odnawianą 61
Tabela 15. Podstawowa literatura z zakresu analizy sieci 70

Spis rycin

Rycina 1. Dwa wymiary efektów interwencji 9
Rycina 2. Procedura selekcji raportów - przykład Polski 12
Rycina 3. Bloki analityczne bazy danych 13
Rycina 4. Schemat logiczny interwencji UE 24
Rycina 5. Schemat logiczny interwencji a kryteria ewaluacyjne - podejście DG-Regio 24
Rycina 6. Liczba badań według okresów programowania FS 31
Rycina 7. Daty podejmowania badań ewaluacyjnych 31
Rycina 8. Struktura typów analizowanych ewaluacji 32
Rycina 9. Źródła wykorzystane przy tworzeniu metodologii 34
Rycina 10. Narzędzia użyte na etapie strukturalizacji 35
Rycina 11. Metody i źródła użyte na etapie zbierania danych 38
Rycina 12. Inne metody analizy i oceny 41
Rycina 12. Metody użyte do analizy i oceny 42
Rycina 14. Przedmiot prowadzonych badań 44
Rycina 15. Dwa wymiary efektów interwencji 45
Rycina 16. Liczba raportów wg odsetka wniosków poświęconych efektom 46
Rycina 17. Tradycyjny model logiczny 60
Rycina 18. Model logiczny Fundacji Kellogg'a 60
Rycina 19. Teoria programu a model logiczny 60
Rycina 20. Przykład teorii programu Zarządzania Energią Odnawianą 62
Rycina 21. Rekonstruowanie teorii programu w ewaluacji efektów 66
Rycina 22. Przykład analizy sieci - program CADSES 76

WPROWADZENIE

Cel i uzasadnienie badania

Głównym celem badania, którego wyniki prezentuje niniejsze opracowanie, było **zidentyfikowanie i analiza ciekawych praktyk - tak europejskich jak i światowych - z zakresu ewaluacji efektów programów rozwoju regionalnego**. Badanie miało więc charakter eksploracyjny i stawiało trzy pytania szczegółowe:

- Jaka jest praktyka ewaluacji programów rozwoju regionalnego (współfinansowanych z funduszy UE) w Polsce i w wybranych krajach europejskich?
- Jakie inspiracje można znaleźć poza Unią Europejską i polem funduszy strukturalnych?
- Które z tych praktyk warte są zaadaptowania i propagowania w Polsce?

Przesłanki do uznania tematu metodologii ewaluacji efektów za ważny i warty podjęcia płyną zarówno z obecnej praktyki polskiego sektora publicznego jak i najnowszych badań empirycznych. Ewaluacja jest powszechnie akceptowanym i wymaganym narzędziem zwiększania efektywności i skuteczności interwencji funduszy strukturalnych (Bachtler, 2000; Rozporządzenie Rady (WE), 2006). Ewaluacje procesów zarządzania, identyfikując i usuwając blokady proceduralne, poprawiając procedury i zbliżając interesariuszy pomagają niewątpliwie technicznej absorpcji funduszy UE.

Jednak prócz bieżących, operacyjnych kwestii ważna jest refleksja strategiczna - umieszczająca skuteczność danej interwencji w kontekście szerszych teorii, rozwoju trwałego i długookresowej konkurencyjności danego układu przestrzennego. Zdolność polskiej administracji do autorefleksji i weryfikowania obranych strategii rozwoju jest kluczem do prawdziwie efektywnego, prorozwojowego wykorzystania funduszy UE (tzw. absorpcji strukturalnej).[1] Takiej refleksji służą właśnie ewaluacje efektów dostarczające naukowych dowodów i oceny rezultatów oraz oddziaływania interwencji publicznych. Są to badania ex post, ale również badania bieżące zaawansowanych interwencji, prowadzone pod koniec ich cyklu wdrażania

[1] Jak pokazuje literatura tematu, w dłuższej perspektywie to nie sama sprawność wydawania pieniędzy, ale zmiany osiągnięte dzięki funduszom przekładają się na rozwój. Sama pomoc – poprawnie wydana, ale źle ukierunkowana może być czynnikiem blokującym rozwój (Hausner, 2007).

(European Commission, 1995f; 2007a; Olejniczak, 2008, s.38). Ich liczba w najbliższych latach znacznie wzrośnie ponieważ zbliża się czas podsumowań okresu 2004-06, jak i efektów funduszy przedakcesyjnych.

Jak dotąd Polska ma jednak niewielkie doświadczenie w tego typu ewaluacjach. Wynika to z trzech przyczyn. Po pierwsze, do połowy lat 90-tych w Polsce nie było kultury zarządzania opartego na wynikach (z którym to paradygmatem zarządzania ewaluacja jest nierozerwalnie związana). Wprowadzane jest ono dopiero teraz, za sprawą wymogów i procedur UE (Olejniczak, 2007, s.393). Po drugie, tradycja ewaluacji jest stosunkowo świeża - pojawiła się dopiero w latach 1990-tych a rozwijana jest stopniowo, główne w związku z zewnętrznymi środkami pomocowymi (Kierzkowski, 2002). Wreszcie po trzecie, badania ex post dotychczas nie dotyczyły polskich programów. Ewaluacje funduszy pomocowych były zlecane bezpośrednio przez Komisję Europejską i wykonywane zwykle przez zagraniczne firmy. W przypadku badań funduszy strukturalnych koncentrowano się na bieżącym zarządzaniu. Poza tym pierwszy okres planowania dopiero dobiegł końca – zaś opóźnienia w projektach sprawiają, że całościowe analizy efektów programów będą prowadzone dopiero w najbliższych latach.

Pierwsze badania empiryczne wykorzystania ewaluacji w Polsce pokazują, że zarówno zlecający jak i wykonawcy badań mają poważne problemy z oszacowaniem efektów interwencji, a także z budowaniem wiedzy strategicznej w oparciu o ewaluacje ex post (Ferry, Olejniczak, 2008). Co więcej, jak zauważają krajowi eksperci (Szlachta, 2006; 2007), dotychczasowe metodologie ewaluacji rzadko wychodzą poza wytyczne Komisji Europejskiej - mają więc charakter techniczny i odtwórczy.

Główne pytanie adresowane przez niniejszy raport - *„W jaki sposób prowadzić ewaluacje efektów programów rozwoju regionalnego?”* – umieszcza go więc w centrum praktycznej debaty nad dostarczeniem wniosków i wskazówek do skutecznego, prorozwojowego wykorzystania środków Unii Europejskiej.

Mam nadzieję, że zrealizowane badanie pomoże w sprostaniu choć części nakreślonych powyżej wyzwań i da bazę inspiracji, źródeł oraz dalszych kierunków rozwoju metodologii ewaluacji w Polsce.

Przedmiot badania

Przedmiotem przeprowadzonego badania była metodologia badań ewaluacyjnych podsumowujących efekty programów rozwoju regionalnego. W tym stwierdzeniu wyjaśnienia i sprecyzowania wymagają cztery elementy.

Po pierwsze chodzi o termin **"metodologia"**. W niniejszym raporcie za "metodologię" uznano zarówno:

- obraną całościową koncepcję badania - podejście badawcze (tzw. *evaluation approach*)[2] jak i...
- składające się na nią konkretne, pojedyncze narzędzia i metody badawcze (np. ankiety, wywiady, przegląd dokumentów) zastosowane w danym badaniu.

Drugie wyjaśnienie dotyczy terminu **"podsumowujące badania ewaluacyjne "** (ang. *summative evaluations*). Obejmuje on zarówno badania zakończonych już programów lub też zakończonych części programów (projektów, grup projektów, działań) będących w zaawansowanym stadium wdrażania. Odnosząc to do typologii czasowej ewaluacji,[3] przyjąłem że w polu mojego zainteresowania leżą zarówno ewaluacje ex post jak i ewaluacje bieżące programów (w tym średniookresowe i ich uaktualnienia), które przekroczyły "półmetek" wdrażania.

Trzecią kwestią jest sprecyzowanie terminu **"efekty".** Efekty to stopień i skala zmian do której program dążył i którą w rzeczywistości osiągnął. Definicję efektów należy rozpatrywać w dwóch wymiarach:

- w wymiarze czasowym, który obejmuje zarówno rezultaty (bezpośredni skutek interwencji dla jej odbiorców) jak i oddziaływanie (zmianę długookresową, o charakterze strukturalnym);[4]

---

[2] Mark i inni (2000, s.75) nazywają to "*inquiry modes*" tłumacząc, że jest to punkt pośredni łączący ogólne cele badania z bardzo technicznymi metodami służącymi analizie danej interwencji. Natomiast Stufflebeam & Shinkfield (2007) przyjmując szerszą interpretację stosują termin "approach" wymiennie z "medelami ewaluacji" (ang. *evaluation models*) identyfikując jednocześnie ich 26 typów (wg. celów badania, pytań na których się koncentrują oraz zestawów metod badawczych). W niniejszej analizie przychylam się do interpretacji Marka i jego współautorów.

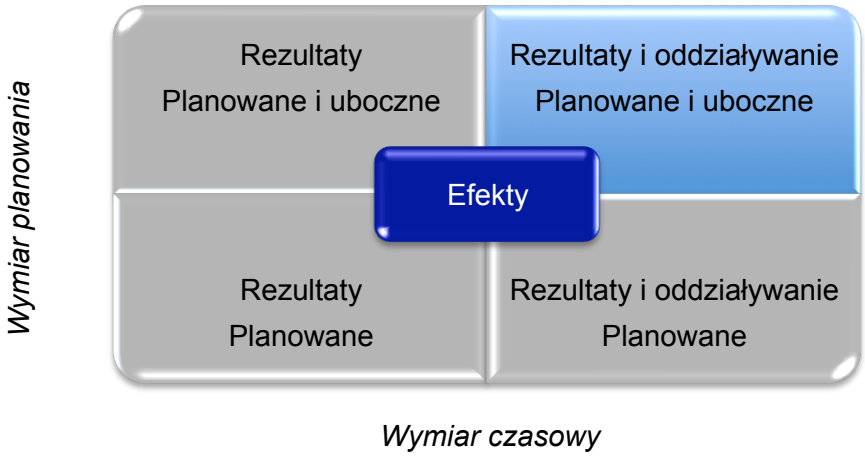
[3] Podstawowe typy czasowe ewaluacji w przypadku programów funduszy strukturalnych to: ewaluacje ex ante, końcowe (ex post) oraz bieżące - w tym średniookresowe (mid-term) i uaktualnienie badań średniookresowych (mid-term update) (szerzej patrz: Olejniczak, 2007, s.27-28)

[4] Szczegółową definicję tych terminów można znaleźć w: (European Commission, 1999 vol. 6; 2006b), zaś ich polską interpretację w: (Olejniczak, 2004, s.299)

- w wymiarze planowania - uwzględniającym zarówno efekty zaplanowane przez osoby projektujące i wdrażające daną interwencję jak i efekty tzw. uboczne (pozytywne lub negatywne), które pojawiły się choć nie były przewidziane w oryginalnym planie interwencji.

Warto zwrócić uwagę, że te dwa wymiary mogą dać cztery różne interpretacje - czasem mocno zawężające. Możliwe opcje przedstawiono na rycinie 1.

Rycina 1. Dwa wymiary efektów interwencji



Źródło: opracowanie własne

W niniejszej analizie przyjąłem pełną interpretację definicji efektów - rozumianych jako planowane i uboczne rezultaty oraz oddziaływanie danej interwencji (prawe, górne pole na rycinie 1).

Wreszcie sprawą ostatnią jest delimitacja pola **"rozwoju regionalnego"**. Na potrzeby tego projektu założono, że programami rozwoju regionalnego są wszystkie programy o charakterze przestrzennym, współfinansowane ze środków Europejskiego Funduszu Rozwoju Regionalnego. Ten drugi element definicji (finansowanie ERDF) nie zawsze był stosowany. W analizę włączono programy INTERREG (w latach 1991-2006 będące Inicjatywą Wspólnotową) jak również - co oczywiste - programy polskie sprzed momentu akcesji (finansowane przez PHARE).

Na zakończenie wyjaśnienia kwestii związanych z przedmiotem badania warto również sprecyzować oczekiwania stawiane przed ewaluacjami efektów. Zgodnie z literaturą tematu (Olejniczak, 2008, s.177-178; Stern, 2005, s.25; Stufflebeam, Shinkfield, 2007, s.25) jak i wytycznymi prawnymi UE (Rozporządzenie Rady (WE), 2006 art.47-49) ewaluacje podsumowujące mają spełniać dwie podstawowe funkcje -

rozliczenia (a przez to wzmocnienia odpowiedzialności) oraz uczenia się. Obydwie te funkcje mają charakter strategiczny. Praktyczne wskazania z tego płynące podsumowano w poniższej tabeli.

Tabela 1: Zakres ewaluacji efektów

Pole eksploracji	Orientacja wniosków
I. Badanie stopnia w jakim zrealizowano zaplanowane cele danej interwencji	Wnioski rozliczające menedżerów z wykonania planu
II. Identyfikowanie rzeczywistych efektów interwencji i analiza stopnia pożądanej zmiany społecznej, do której program doprowadził	Wnioski wyjaśniające relację przyczynowo-skutkową i pokazujące użyteczność programu dla społeczeństwa
III. Analiza mechanizmu sukcesu/niepowodzenia interwencji w tym: • rolli jaką odegrały czynniki kontekstowe (zewnętrzne) • roli jaką odegrał sposób dostarczenia interwencji (wdrażania) • roli założeń programu nt. mechanizmów pożądanej zmiany (teorii i wcześniejszej praktyki na której zbudowano założenie relacji przyczynowo-skutkowej)	Wnioski wyjaśniające mechanizm skutecznej zmiany społecznej w polu adresowanym przez dany program oraz pokazujące kierunki przyszłych skutecznych strategii.

Źródło: opracowanie własne

Warto tu podkreślić za Mark et. al. (2000, s.78), że pole drugie wychodzi poza prosty opis zaobserwowanych efektów, gdyż dotyczy przyczynowości. Aby przypisać zrealizowanemu programowi zmiany zaobserwowane w danym układzie przestrzennym trzeba najpierw wykluczyć alternatywne wyjaśnienia tych zmian i znaleźć dowody na relację przyczynowo-skutkową między zmianą a działaniami programu. Takie ujęcie jest prawdziwym wyzwaniem w przypadku analizy programów wielosektorowych, o wielowarstwowej strukturze celów i dużej liczbie działań.

Zespół badawczy i metodologia

Zespół badawczy składał się z krajowych i międzynarodowych pracowników oraz współpracowników EUROREG - Uniwersytetu Warszawskiego. W skład krajowej części zespołu weszli: mgr Adam Płoszaj, mgr Alicja Weremiuk, mgr Katarzyna

Wojnar, dr Dominika Wojtowicz, dr Michał Wolański, oraz Monika Kur. Współpracownikami zagranicznymi byli: dr Kai Böhme ze SWECO EURORFUTURES - Sztokholm, dr Richard Boyle z Institute of Public Administration (IPA) - Dublin, dr Martin Ferry z European Policies Research Centre (EPRC) - University of Strathclyde - Glasgow oraz Lukas Wortmann z Metis GmbH - Wiedeń. Zespołem kierował autor niniejszego raportu - dr Karol Olejniczak.

Całościowy plan badawczy przyjęty w niniejszym badaniu miał charakter przekrojowy (ang. *cross-sectional*). W badaniu wykorzystano takie metody jak: analizę jakościową i ilościową dokumentów (raportów z ewaluacji), przegląd literatury oraz konsultacje eksperckie. Metody zróżnicowano na każdym z etapów badania dostosowując je do potrzeb każdego z kroków badawczych. Przedstawiamy je poniżej, wraz z wykorzystanymi źródłami, według kolejności 6 zrealizowanych kroków badawczych.

Pierwszym krokiem badawczym była analiza dotychczasowej praktyki UE. Podstawową metodą była analiza literatury tematu poświęconej kulturom ewaluacji w Europie. Głównymi źródłami były: książki i raporty (Bachtler, 2000; 2001; Bachtler et al., 2000; Furubo et al., 2002; Toulemonde, 2000), seria poradników metodologicznych UE (edycje MEANS 1995, 1999 oraz Evalsed 2003, 2006 i 2007), kwerenda periodyków *Regional Studies* i *Evaluation* z ostatnich 10-lat, a także przeglądy praktyk ewaluacyjnych w polu funduszy strukturalnych prowadzone przez EPRC (www.eprc.strath.ac.uk). Na tej podstawie wybrano 4 kraje o wysokiej kulturze ewaluacji: Irlandię, Niemcy, Szwecję oraz Wielką Brytanię.

W kroku drugim przeanalizowano dogłębnie - według jednolitej struktury pytań problemowych - kulturę ewaluacji w polu Funduszy Strukturalnych w tych 4 wybranych krajach. Metodą była analiza źródeł oraz ocena ekspercka przeprowadzona przez ekspertów działających w tych 4 krajach.

W kroku trzecim wybrano, dla każdego z 4 krajów oraz dla Polski, po 20 raportów ewaluacji efektów programów rozwoju regionalnego, współfinansowanych ze środków UE. Raporty wybierano według następujących kryteriów:

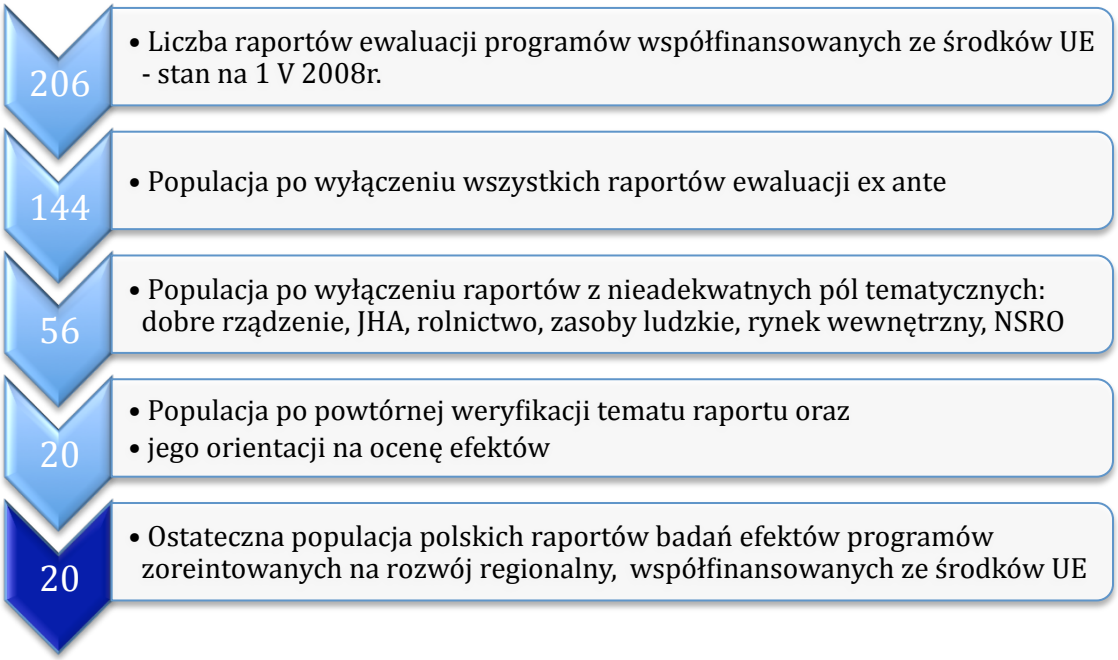
1. Program miał dotyczyć polityki regionalnej tzn. być współfinansowany z ERDF, wdrażany na poziomie regionalnym, lub też wyraźnie zorientowany na aspekty przestrzenne;
2. Badanie miało dotyczyć zakończonej interwencji (ex post) lub w przypadku badań bieżących (*on-going,* średniookresowych) programu zaawansowanego we wdrażaniu;

3. W oficjalnych celach badania miała być wymieniona analiza i ocena efektów danego programu;
4. Raport miał posiadać wyraźny rozdział (podrozdział) metodologiczny.

Zakładano, że do ostatecznej analizy z tak zebranej populacji raportów zostanie wybranych po 20 raportów z każdego kraju, o najbardziej rozwiniętych rozdziałach metodologicznych (czyli z największym prawdopodobieństwem ciekawych, rozwiniętych inspiracji).

W praktyce, po zastosowaniu pierwszych trzech kryteriów, okazało się że próba 20 raportów z poszczególnych krajów praktycznie wyczerpuje całość populacji badań efektów programów rozwoju regionalnego. Tak więc grupa 100 raportów z 5 krajów stanowi praktycznie całość badań ewaluacyjnych w polu efektów polityki regionalnej UE prowadzonych na poziomie administracji krajowych (sic!). Jako przykład procedury selekcji przedstawiamy Polskę (ryc. poniżej).

Rycina 2. Procedura selekcji raportów - przykład Polski



Źródło: opracowanie własne

Krokiem czwartym badania było stworzenie ram analitycznych do usystematyzowanego przeglądu i oceny metodologii każdego ze 100 badań ewaluacyjnych. Kategorie merytoryczne bazy stworzono w oparciu o następujące publikacje: (European Commission, 1995f; 1997; 1999; 2007a; GHK Consultants, 2006; Howell, Yemane, 2006; Mathison, 2005; Shadish et al., 2001; Tavistock Institute et al., 2003). Stworzono 7 bloków badawczych (patrz rycina 3).

Szczegółowa struktura pytań analitycznych bazy została przedstawiona w Aneksie 1 niniejszego opracowania.

Od strony technicznej zdecydowano się na oparcie bazy danych na platformie FileMaker Pro 9 (http://www.filemaker.com/products/fmp/) i zorganizowanie jej w formie 7 tabel, obejmujących dla każdego rekordu (tzn. analizowanego raportu) w sumie 100 kolumn z pytaniami - kategoriami analitycznymi.

Rycina 3. Bloki analityczne bazy danych

Źródło: baza danych projektu

Krokiem piątym badania była porównawcza analiza rekordów bazy. Objęła ona proste zestawienia statystyczne zrobione dla całości badanej populacji.

W ostatnim, szóstym kroku - poszukiwaniu szerszych inspiracji - oparto się na całościowym przeglądzie najnowszej literatury pola ewaluacji programów publicznych.[5] Do zebrania materiałów wykorzystano wizytę studyjną oraz konferencje.

[5] Oprócz pozycji książkowych, głównymi źródłami, które uwzględniono w przeglądzie były periodyki poświęcone ewaluacji takie jak: American Journal of Evaluation, Evaluation & Program Planning, Evaluation Review, Evaluation Practice, New Directions for Evaluation, Canadian Journal of Evaluation, Regional Studies oraz kwartalnik Evaluation (prowadzony przez Europejskie Towarzystwo Ewaluacyjne).

Miesięczna wizyta studyjna w USA objęła:

- kwerendy biblioteczne przeprowadzone w dwóch ośrodkach akademickich - Indiana University, Bloomington, Indiana oraz w School of Public Policy and Public Administration, The George Washington University, Washington D.C;
- udział w seminarium metodologicznym "Program Evaluation" prowadzonym w The Trachtenberg School of Public Policy and Public Administration, G.Washington University;
- wygłoszenie wykładu dla pracowników naukowych The Trachtenberg School of Public Policy and Public Administration na temat wykorzystania ewaluacji w UE oraz dyskusja na temat podobieństw i różnic amerykańskiej i europejskiej praktyce ewaluacji;
- wywiad Prof. Davidem Reingoldem vice-dyrektorem School of Public and Environmental Affairs na temat ilościowych podejść do ewaluacji efektów oraz przykładów ewaluacji programów federalnych;
- konsultację merytoryczną z Prof. Embry M.Howell z The Urban Institute - Washington D.C. na temat metod analizy porównawczej planów badań ewaluacyjnych;
- 2-tygodniowe konsultacje z Prof. Katrin Newcomer - dyrektor The Trachtenberg School of Public Policy and Public Administration, na temat podejść do ewaluacji kompleksowych programów federalnych.

Konferencje, na których zbierano materiały metodologiczne (oraz dyskutowano wstępne wyniki badania) odbyły się w Warszawie (Ogólnopolska Konferencja Ewaluacyjna) oraz Lizbonie (Międzynarodowa Konferencja European Evaluation Society).

1 PRAKTYKA EWALUACJI EFEKTÓW FUNDUSZY STRUKTURALNYCH

Rozdział składa się z trzech części. W pierwszej przedstawiono kontekst kulturowo-instytucjonalny w jakim powstawała i obecnie rozwija się praktyka ewaluacji funduszy strukturalnych. W części drugiej - będącej rdzeniem badania - zaprezentowano wyniki analizy metodologii 100 badań ewaluacyjnych efektów funduszy strukturalnych. W części trzeciej - wnioskach, podsumowano syntetycznie spostrzeżenia z badania oraz przedstawiono próbę wyjaśnienia zaskakująco słabego stanu rzeczy.

1.1 Kontekst - kultury ewaluacji w Unii Europejskiej

Wyrastająca z kultury ewaluacji praktyka badań ewaluacyjnych w Unii Europejskiej jest zwykle analizowana na trzech płaszczyznach (Olejniczak, 2005, s.42-54):

1. Kultury i praktyki badań odnoszących się do programów krajowych;
2. Kultury ewaluacyjnej i praktyki promowanej na poziomie instytucji unijnych - przede wszystkim Komisji Europejskiej;
3. Kultury i praktyki ewaluacji programów europejskich (współfinansowanych z funduszy strukturalnych) na poziomie krajów członkowskich.

Płaszczyzny te pozostają we wzajemnych relacjach. Wpływy dwóch pierwszych (krajowej i unijnej) zlewają się na polu programów UE. Płaszczyzna pierwsza - kultura krajowa warunkuje (przynajmniej teoretycznie) praktyki funduszy strukturalnych. W krajach o silnej kulturze ewaluacji rozwinięta i ugruntowana praktyka krajowa przenika do relatywnie nowego pola programów unijnych. Z drugiej strony w krajach o niskiej kulturze ewaluacji (braku doświadczeń administracji krajowej w zarządzaniu opartym na wynikach) fundusze strukturalne i standardy promowane przez Komisję (płaszczyzna 2) wymuszają praktykę badań ewaluacyjnych w polu programów UE, która to z kolei powoli rozlewa się na pierwszą płaszczyznę - programy *stricte* krajowe. Wraz ze zwiększaniem skali interwencji funduszy strukturalnych, procesy te są coraz silniejsze. Przykładem takiego mechanizmu jest Irlandia, a - od niedawna powoli staje się nim również Polska.[6]

[6] Dowodami na rozlewanie się ewaluacji na krajową kulturę administracji są między innymi: eksponowanie, przez MRR i Zespół Strategicznych Doradców Premiera, ewaluacji jako głównego narzędzia oceny krajowych polityk rozwoju, uwzględnienie ewaluacji jako elementu

Krótką charakterystykę każdej z trzech płaszczyzn przedstawiono w kolejnych podpunktach.

1.1.1 Płaszczyzna 1: Krajowe kultury ewaluacji

Głównym - i jak dotąd jedynym opracowaniem prezentującym kultury ewaluacji w ujęciu komparatywnym, w skali międzynarodowej jest Atlas ewaluacji z 2002 (Furubo et al., 2002). Niestety atlas nie uwzględnia krajów postkomunistycznych - Europy Środkowowschodniej. Pozostaje jednak wartościowym punktem wyjścia dla naszych rozważań, gdyż pokazuje część krajów europejskich (tych o najdłuższej tradycji ewaluacji) w kontekście międzynarodowym (obejmuje 21 krajów, z czego 12 to kraje europejskie, a 10 to kraje Unii Europejskiej). Kultura ewaluacji poszczególnych krajów analizowana i oceniona została w 9 aspektach. Wyniki punktacji przedstawia poniższa tabela.

reformy budżetowania zadaniowego przygotowywanego przez Ministerstwo Finansów, najnowsze zapisy w Ustawie o prowadzeniu polityk rozwoju w której ewaluacja ex ante staje się wymogiem dla wszystkich programów publicznych (art. 15.6) oraz bazujące na wynikach badań ewaluacyjnych z lat 2004-06 *Policy Recommendation Reports* przygotowywane przez MRR dla Rady Ministrów (Kozak et al., 2009, s.65-66).

Tabela 2: Krajowe kultury ewaluacji - ranking 21 państw

	I	II	III	IV	V	VI	VII	VIII	IX	Razem
Stany Zjednoczone	2	2	2	2	2	2	2	2	2	**18**
Kanada	2	2	2	2	2	1	2	2	2	**17**
Australia	2	2	2	2	1	1	2	2	2	**16**
Szwecja	2	2	2	1	2	1	2	2	2	**16**
Holandia	2	2	2	1	2	1	2	2	1	**15**
Wielka Brytania	2	2	2	2	1	1	2	1	2	**15**
Niemcy	2	2	1	2	1	1	2	1	1	**13**
Dania	2	2	2	1	1	0	2	1	1	**12**
Korea	1	1	2	2	2	0	2	1	1	**12**
Norwegia	2	1	1	1	2	1	1	2	1	**12**
Francja	2	1	1	2	2	1	1	1	0	**11**
Finlandia	2	1	1	1	1	1	1	1	1	**10**
Szwajcaria	1	1	2	2	0	0	2	0	0	**8**
N. Zelandia	1	0	1	2	0	0	1	1	1	**7**
Irlandia	1	1	1	0	1	0	1	1	1	**7**
Włochy	1	1	1	2	0	0	1	1	0	**7**
Chiny	1	1	0	0	2	0	1	0	1	**6**
Hiszpania	1	0	1	2	1	0	0	0	0	**5**
Zimbabwe	1	1	0	0	1	0	0	1	0	4
Japonia	1	0	0	1	1	0	0	0	0	3

Źródło: (Furubo et al., 2002, s.10)

Legenda: I) Ewaluacja stosowana w wielu obszarach polityk publicznych II) Wysoka podaż ewaluatorów w różnych dyscyplinach III) Ogólnokrajowy dyskurs ewaluacyjny IV) Organizacje zawodowe. V) Stopień instytucjonalizacji – rząd VI) Stopień instytucjonalizacji – parlament VII) Pluralizm instytucji lub ewaluatorów, wykonujących ewaluacje w każdym z obszarów polityki. VIII) Ewaluacja obecna w Najwyższej Izbie Kontroli. IX) Stosunek liczby ewaluacji efektów do badań produktów i procesów.

Na czele rankingu są Stany Zjednoczone i Kanada, kraje w których w latach 1950-tych narodziła się współczesna ewaluacja programów publicznych. Z krajów europejskich najwyżej uplasowały się państwa, których administracje stosunkowo wcześnie zaadaptowały zarządzanie oparte na wynikach: Szwecja, Holandia, Wielka Brytania oraz Niemcy. Warto zwrócić uwagę na stosunkowo niską pozycję Nowej Zelandii. Okazuje się, że radykalna odmiana *New Public Management* nie idzie w parze z całościową kulturą ewaluacji, gdyż zawęża praktykę badań do funkcji menedżerskich (bieżącej pomocy we wdrażaniu programów).

Analiza rozwoju i obecnego stanu warsztatu metodologicznego ewaluacji w poszczególnych krajach wykracza daleko poza ramy niniejszego opracowania. Warto jednak w tym miejscu zwrócić uwagę na fakt, że w większości krajów (poza Wielką Brytanią) ewaluacja wyrosła z audytu (traktowano ją jako uzupełnienie - bardziej miękkie i szersze - analizy efektywności). Tak więc w odróżnieniu od Wielkiej Brytanii gdzie ewaluacja ma szerokie korzenie badań społecznych, w krajach takich jak Szwecja, Holandia, czy Niemcy ewaluacja przejęła zarówno filozofię jak i oprzyrządowanie audytu - silnie mechanistyczne, nakierowane na rozliczenie menedżerów publicznych. To "dziedzictwo audytu" jest także wyraźne w European Evaluation Society - powstało ono stosunkowo późno, w porównaniu do swego amerykańskiego odpowiednika, a jego działalność była na początku silnie inspirowana przez audytorów holenderskich.

Warto zastanowić się nad możliwą pozycją Polski w tym rankingu. W tabeli trzeciej przedstawiono ocenę (wraz z uzasadnieniem) polskiej kultury ewaluacji według 9 kryteriów międzynarodowego atlasu ewaluacji. Główną słabością wydaje się ograniczenie praktyk ewaluacyjnych - w tym jej instytucjonalizacja - tylko do pola funduszy UE i związana z tym "monokultura". Porównując wyniki tabeli 3 z wynikami 2, należy również pamiętać, o tym, że obydwa zestawienia dzieli 8 lat. Nasz dystans do państw środka rankingu mógł się więc powiększyć.

Tabela 3: Krajowa kultura ewaluacji - próba oceny Polski

	I	II	III	IV	V	VI	VII	VIII	IX	Razem
Polska	0	1	1	1	1	0	1	0	0	**5**

źródło: opracowanie własne

Uzasadnienie

Kryterium	Uzasadnienie
I) Ewaluacja stosowana w wielu obszarach polityk publicznych	Tylko w polu funduszy strukturalnych (rozumianego jako jedno, choć wielosektorowe, pole); "niszowa" w polu edukacji - w odniesieniu do projektów a nie całych programów czy polityk publicznych
II) Wysoka podaż ewaluatorów w różnych dyscyplinach	Nieliczna grupa ekspertów. Kilka kierunków studiów podyplomowych - zorientowane głównie na FS. Popyt znacząco przerasta podaż
III) Ogólnokrajowy dyskurs ewaluacyjny	Powoli się rozpoczyna i wychodzi poza pole FS
IV) Organizacje zawodowe	Tylko jedna, skupiająca głównie środowisko ewaluatorów z pola edukacji
V) Stopień instytucjonalizacji – rząd	Coraz silniej widoczna na poziomie operacyjnym, lecz tylko w departamentach związanych z FS
VI) Stopień instytucjonalizacji – parlament	Zupełnie nieobecna
VII) Pluralizm instytucji lub ewaluatorów, wykonujących ewaluacje w każdym z obszarów polityki	Dominacja instytucji związanych z wdrażaniem FS. Powoli powstaje sieć regionalna - również jednak związana z polem programów UE
VIII) Ewaluacja obecna w Najwyższej Izbie Kontroli.	VIII. Całkowicie nieobecna. Pierwsze pomysły "audytu efektywnościowego" podejmowane w MF
IX) Udział ewaluacji efektów w ewaluacjach produktów i procesów	Wyraźna dominacja ewaluacji procesów i produktów (proporcja 10 do 1). Wynika z działania ewaluacji tylko w polu FS, w którym dotychczasowy etap cyklu programów uniemożliwiał analizę efektów.

Źródło: analiza własna z wykorzystaniem źródeł: (Daniluk, 2008; Dobrzyńska, 2008; Kierzkowski, 2002; Krajowa Jednostka Oceny, 2005, 2008; Olejniczak et al., 2008b; Szlachta, 2006; Zawicki, 2007)

Zawężając perspektywę praktyki ewaluacji wyłącznie do pola polityk regionalnych w państwach członkowskich UE wyróżnić można 4 grupy krajów (Bachtler, 2001, s.49-51; Olejniczak, 2008, s.79-80; Toulemonde, 2000):

1. państwa o długiej tradycji krajowej (sięgającej lat 1970-tych) i wysokiej kulturze ewaluacji (Wielka Brytania, Holandia, Niemcy, Francja, Szwecja, Dania);
2. państwa bez tradycji ewaluacyjnych ale o rozwiniętej kulturze ewaluacji zaszczepionej przez Fundusze Strukturalne (Irlandia, Belgia, Austria, Włochy - część północna);
3. państwa bez tradycji i o wciąż niskiej kulturze ewaluacji (Grecja, Hiszpania, Portugalia)
4. 10 nowych krajów członkowskich - bez tradycji ewaluacyjnych i wciąż wahających się pomiędzy drugą a trzecią grupą (Polska, Czechy, Węgry, etc.).

W literaturze tematu podkreślane są dwie kwestie (Bachtler, Wren, 2006, s.149-150; Olejniczak, 2005). Po pierwsze wyraźny jest podział na kraje północy i południa, który pokrywa się z odmiennymi kulturami administracji publicznej; widać silną zależność między kulturą administracji publicznej opartej zarządzaniu przez cele i budżetowaniu zadaniowym a wysoką kulturą ewaluacji. Ewaluacja jest po prostu jednym z elementów nowoczesnego zarządzania opartego na wynikach (ang. *performance-management*).

Po drugie do Irlandii, krajów południa Europy i nowych państw członkowskich ewaluacja przyszła z zewnątrz - jako element *know-how* związanego z obsługą środków pomocowych USAid, Banku Światowego i Unii Europejskiej. Głównym czynnikiem promowania zmiany były wymogi funduszy przedakcesyjnych, a potem strukturalnych UE. Krótką dyskusję na temat unijnego *know-how*, przedstawiono w kolejnym podpunkcie.

1.1.2 Płaszczyzna 2: Kultura ewaluacji w Komisji Europejskiej

Ewaluacja na poziomie instytucji UE pojawiła się w połowie lat 1980-tych wraz ze znaczącym wzrostem budżetu Wspólnot Europejskich. Szczegółowe omówienie genezy i kolejnych etapów rozwoju praktyki ewaluacyjnej w UE można znaleźć w Bachtler & Taylor (1999). W tym miejscu przywołujemy tylko trzy momenty przełomowe:

- pod koniec lat 1980-tych budżetowa reforma J.Delorsa (I pakiet Delorsa) spowodowała znaczące zwiększenie budżetu UE i wydatków w polu funduszy strukturalnych; ewaluacja miała być narzędziem zdyscyplinowania (głównie

państw beneficjentów pomocy) efektywności zarządzania tak znaczącymi środkami;

- w połowie lat 1990-tych, gdy końca dobiegał 1 okres programowania polityki regionalnej UE zaczęto poszukiwać narzędzia informacji o jej efektach; w konsekwencji - na nowy okres programowania wyeksponowano rolę ewaluacji w zarządzaniu funduszami strukturalnymi;
- koniec lat 1990-tych - kryzys Komisji Santera i wywołana nim reforma unijnego zarządzania (*Sound and Efficient Management 2000*) - ewaluacja znalazła się w pakiecie kluczowych narzędzi zarządzania opartego na wynikach, została także wpisana w Rozporządzenie budżetowe UE jako wymóg prawny dla wszystkich działań mających wpływ na budżet UE

Ta ewolucja pokazuje wyraźnie rozlewanie się ewaluacji z pola funduszy strukturalnych na wszystkie inicjatywy podejmowane przez Komisję Europejską.

Z perspektywy tematu niniejszego badania szerszego omówienia wymaga natomiast kwestia podejść metodologicznych rozwijających się na poziomie UE i promowanych przez Komisję Europejską (unijny *know-how* ewaluacji).

Historycznie, metodologie badań ewaluacyjnych rozwijały się w trzech, relatywnie odizolowanych od siebie ośrodkach w Komisji Europejskiej.

Pierwszym ośrodkiem była DG Research, która stosowała ewaluację już od połowy lat 1980-tych, głównie do oceny ex ante jak i ex post wspieranych projektów innowacyjnych i grantów badawczych. Metodologia badań tych, zwykle naukowych projektów, opierała się na tzw. *peer reviews*, czyli na dyskusji i ocenie podejmowanej przez panele naukowców - specjalistów danej dziedziny. DG Research czerpała swoje inspiracje metodologiczne ze świata nauki - procedur recenzowania stosowanych w międzynarodowych periodykach naukowych, czy w konkursach grantowych. Innowacją było to, iż DG Research zaczęła stosować panele ekspertów do przewidywania efektów projektów, ich stopnia innowacyjności. W momencie gdy skala inwestycji zwiększyła się (poziom programów) próbą zobiektyzowania wyników było triangulowanie metod jakościowych (różne panele ekspertów, wywiady z jak najszerszą, zróżnicowaną grupą interesariuszy). Dobrym przykładem takiego podejścia w praktyce są ostatnie ewaluacje Programów Ramowych UE z lat 1999-2004 (Arnold et al., 2008).

Ośrodkiem drugim była część dyrekcji External Aid - biuro EuropeAid odpowiedzialne za projekty i programy pomocowe skierowane do krajów spoza UE - regionów Afryki,

Bałkanów, byłych republik radzieckich ale także - aż do połowy lat 1990-tych - krajów Europy Środkowowschodniej (program PHARE). Dyrekcja ta czerpała silne inspiracje z metod wdrażania projektów pomocowych Banku Światowego i USAid. Elementem tej "skrzynki narzędziowej" była i ewaluacja. Podejście EuropeAid miało orientację projektową i opierało się na filozofii "Zarządzaniu Cyklem Projektu" (ang. *Project Cycle Management*). Kluczowym narzędziem były tu matryce logiczne (tzw. LOGFRAME) wymagane między innymi od wszystkich beneficjentów PHARE (European Commission, 2001; 2004c). Metodologia ta dobrze sprawdzała się na poziomie projektowym ale z czasem zaczęto ją kopiować na poziomie całych programów. Tu jednak zaczęły się problemy - matryce logiczne - przejrzyste dla pojedynczych projektów, stawały się wyjątkowo skomplikowane i trudne do ogarnięcia dla multisektorowych, wielocelowych interwencji pomocowych. Co więcej, podstawowe założenia wdrażania cyklu projektu (ewaluacja po zakończeniu projektu jako podstawa budowy kolejnej strategii projektowej) nie przystają do rzeczywistości wdrażania programów. Nowe programy rozpoczynane w trakcie zamykania poprzednich, ewaluacje ex post robione dla starych programów w połowie trwania nowych.

Wreszcie trzecim ośrodkiem - najbardziej widocznym (gdyż związanym z największym polem interwencji) była DG-Regio. Dyskusję metodologiczną prowadziła od 1993 roku, w ramach cyklu MEANS (Evaluation Methods for Actions of Structural Nature) - serii konferencji i seminariów służących wymianie doświadczeń i wypracowywaniu wspólnych standardów dla praktyk ewaluacji funduszy strukturalnych w różnych krajach członkowskich. Na tej podstawie wydano w 1995 siedem tematycznych zeszytów - poradników (European European Commission, 1995a; 1995b; 1995c; 1995d; 1995e; 1995f; 1995g), zaś 4 lata później ujednolicony 6 tomowy poradnik ewaluacji MEANS (European Commission, 1999 vol.1-6). W 2003 i 2007 doczekał się on kolejnych dwóch edycji (pod nazwą Evalsed), tym razem w formie stron internetowych (European Commission, 2007a; GHK Consultants, 2006; Tavistock Institute et al., 2003). MEANS i Evalsed są jak dotąd najbardziej kompleksowym wydawnictwem metodologicznym nt. ewaluacji - nie tylko w polu FS ale w całej Komisji Europejskiej. Zawierają zarówno krótką podbudowę teoretyczną ewaluacji (geneza, funkcje, cele, kryteria ewaluacyjne), analizę specyfiki badań poszczególnych pól interwencji i sektorów, jak i syntetyczny przegląd metod.

Oprócz tych publikacji DG-regio wydawała też zeszyty robocze (Working Papers) na lata 2000-06 (European European Commission, 2000b; 2000a; 2002b; 2002a; 2004b) i 2007-13 (European Commission, 2006a; 2006c; 2007b; 2007c) zawierające

pewne minimum wytycznych nt. planowania, monitoringu jak i sposobów ewaluacji programów UE. Należy jednak podkreślić, że te publikacje DG-Regio, jak i innych dyrekcji są poradnikami, a nie wydawnictwami *stricte* metodologicznymi. Pomyślane były jako pomoc przede wszystkim dla zlecających ewaluację, a nie wyczerpujące źródło metodologii badań.

W ostatnich 5 latach, patrząc z perspektywy pola funduszy strukturalnych, można zaobserwować tendencje do zlewania się i przenikania odmiennych podejść wypracowanych przez trzy dyrekcje Komisji.

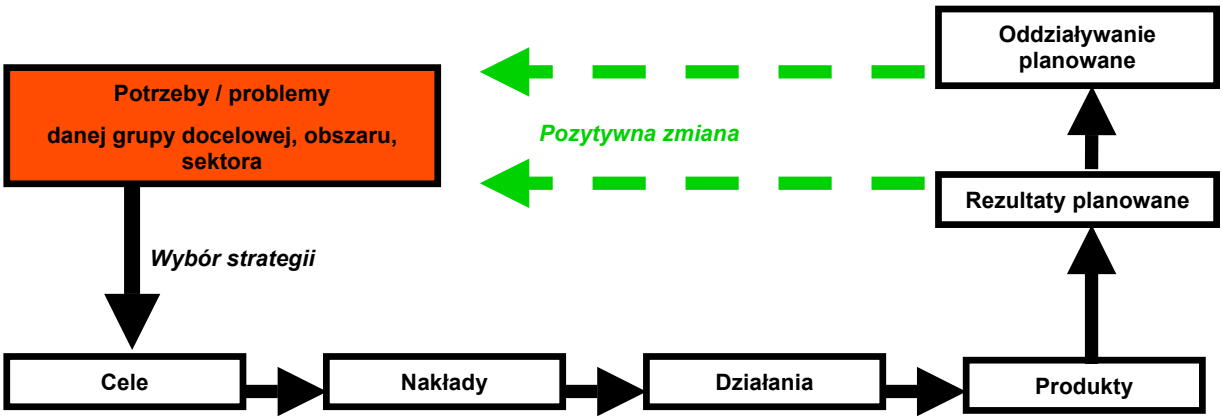
Po pierwsze, część doświadczeń EuropeAid (szczególnie wykorzystanie matryc logicznych) zostało przeniesionych do pola funduszy strukturalnych. Stało się to dzięki nowym państwom członkowskim, wykorzystujących swą wiedzę z PHARE do zarządzania nowymi programami UE.

Po drugie, w polu funduszy coraz częściej znajdują zastosowanie metody z pola badań i rozwoju. Wynika to z faktu, że w głównym nurcie FS pojawia się coraz więcej programów nakierowanych na wspieranie innowacji, badań i rozwoju. Aby móc poddać je ewaluacji sięga się coraz częściej do *peer reviews* i innych metod wypracowywanych przez DG Badania i Rozwój.

Wreszcie wymianie metodologii sprzyja także fakt, że od 1999 koordynację nad działaniami ewaluacyjnymi w Komisji przejął jeden ośrodek - DG-Budżet. Jej strona internetowa, roczne podsumowania badań ewaluacyjnych na poziomie Komisji (np. European Commission, 2003) oraz krótkie poradniki (European Commission, 2004a) są dobrym, pierwszym źródłem informacji o aktualnych praktykach ewaluacji w Komisji.

Rdzeniem podejścia metodologicznego prezentowanego przez Komisję Europejską w polu funduszy strukturalnych jest określony ciąg logiczny interwencji (patrz poniżej) oraz powiązane z nim hierarchie celów-efektów-wskaźników (patrz: tabela).

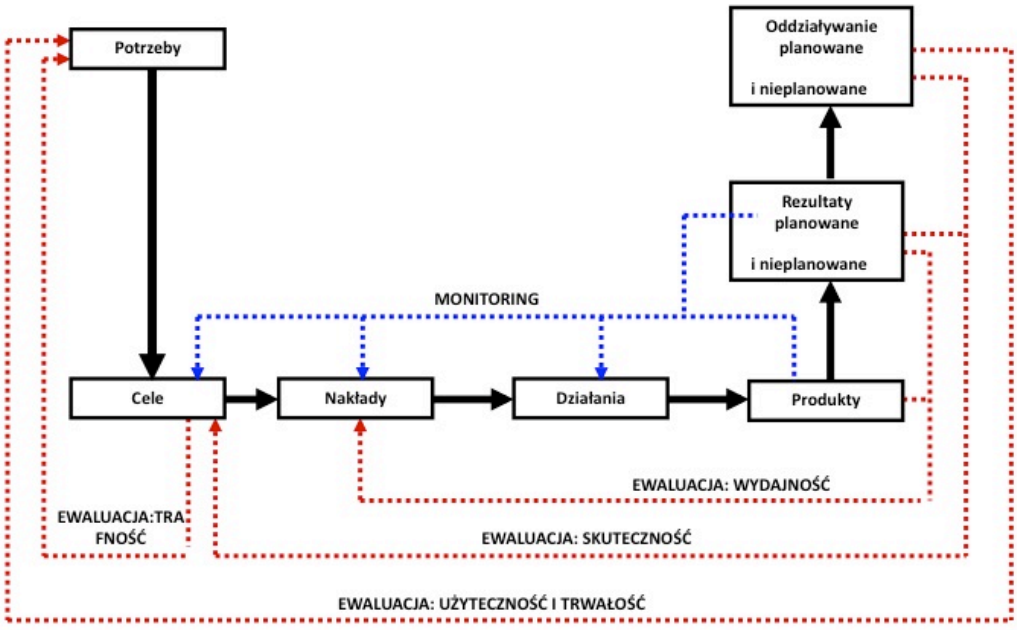
Rycina 4. Schemat logiczny interwencji UE



Źródło: (European Commission, 1999; 2000a; 2001)

Ten ciąg logiczny promowany przez DG-Regio, jest w zasadzie identyczny z wcześniejszymi pracami EuropeAid. Jedyną różnicą jest iż DG-Regio podchodzi dużo bardziej sztywno do kryteriów ewaluacyjnych - podając 5 kluczowych (trafność, wydajność, skuteczność, użyteczność, trwałość) i wiążąc je z poszczególnymi elementami ciągu logicznego. Podejście to przejęła i promuje także DG-Budżet.

Rycina 5. Schemat logiczny interwencji a kryteria ewaluacyjne - podejście DG-Regio



Źródło: (European Commission, 1999, s.71; 2000a, s.9)

Tabela 4. Unijna hierarchia celów - efektów - wskaźników

Hierarchia celów	Hierarchia efektów	Hierarchia wskaźników
Cel globalny, ogólny **(ang. *Global objective*)** Cel Programu, do którego aplikuje dana interwencja - priorytet, do którego realizacja jej będzie się przyczyniać	**Oddziaływanie** **(ang. *Impact*)** Długookresowe, strukturalne skutki interwencji działające 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: (European Commission, 1999, 2000a, 2001) za (Olejniczak, 2004, s.299)

Analizując podejście promowane przez Komisję (szczególnie w zeszytach roboczych DG-Regio) warto zaznaczyć, że jest to dość schematyczne, wskaźnikowe i menedżerskie podejście. Tworzone "pomocniki ewaluacyjne" mają charakter ogólny,

kierowany do zlecających ewaluację i, co należy raz jeszcze podkreślić, nie mogą zastępować fachowej literatury tematu.

Obecnie Komisja - w polu polityki spójności i funduszach strukturalnych - wyraźnie obrała nową strategię. Stara się przesunąć odpowiedzialność za ewaluacje efektów na aktorów krajowych. Wynika to po części z ograniczonych zasobów samej Komisji (po rozszerzeniu UE o kolejne 17 krajów nawet tzw. "miękka koordynacja" staje się wyzwaniem), a po części z szerszej strategii nowego zarządzania europejskiego opartego na otwartej metodzie koordynacji.[7]

1.1.3 Płaszczyzna 3: Kultura ewaluacji funduszy strukturalnych w 4 krajach

Z perspektywy celu niniejszego badania - poszukiwania inspiracji racjonalnym jest skoncentrowanie uwagi na państwach o najbardziej rozwiniętej kulturze ewaluacji na płaszczyźnie krajowej, gdyż daje to relatywnie wyższe prawdopodobieństwo zaobserwowania ciekawych, niesztampowych metodologii, które przeniknęły do funduszy strukturalnych z długoletniej praktyki krajowej. Oddzielnym, acz równie obiecującym źródłem inspiracji wydają się kraje bez wcześniejszych doświadczeń, które w krótkim czasie zbudowały silną kulturę ewaluacji w oparciu o fundusze UE (Austria, Irlandia). Wydaje się, że Polska ma podobną sytuację i perspektywy.

W związku z powyższym do badania wybrano cztery kraje o wysokiej kulturze ewaluacji: Niemcy, Szwecję i Wielką Brytanię - jako przykłady krajów o długiej tradycji i silnej kulturze ewaluacji oraz Irlandię - przykład kraju bez tradycji ewaluacyjnej ale z dobrą praktyką ewaluacji funduszy strukturalnych.

Ostatnia syntetyczna analiza krajowych praktyk ewaluacyjnych w obszarze funduszy strukturalnych pochodzi z 2000 roku (EPRC, Fraser of Allander Institute, 2000, s.51-54). W związku z tym, poniżej przedstawiono krótką, aktualną charakterystykę praktyki Funduszy Strukturalnych (FS) w czterech wybranych krajach. Została ona przygotowana przez krajowych korespondentów. Pełne teksty ekspertyz narodowych zamieszczono w Aneksie 2.

[7] ang. *Open method of coordination* oznacza, że wyznaczamy wspólny cel, zaś sposoby dojścia do niego pozostawiamy w gestii krajów i aktorów społecznych.

Tabela 5. Praktyka ewaluacji w polu funduszy strukturalnych - podsumowanie czterech krajów

Kraj	Kontekst i Perspektywy	Popyt	Podaż	Produkty	Wykorzystanie
Irlandia	W l.70 i 80tych relatywnie dobrze rozwinięty system ewaluacji projektów ale brak ewaluacji programów Poziom programów pojawił się dopiero z FS Drugi impuls dla ewaluacji - boom gospodarczy przełomu XX-XXIw. i reformy zarządzania, jak dotąd o mieszanych wynikach	Główny popyt tworzony przez NDP/CSF Evaluation Unit w ramach MF, koordynuje on wszystkie plany ewaluacyjne Wydziały ewaluacji działają dla każdego z PO; Część z nich ma charakter niezależnych komórek i prowadzi badania własne	Wydziały Ewaluacji ministerstw są też często inst. realizującymi ewaluacje własnymi zasobami Rynek ewaluatorów dość wąski, oparty na relatywnie małych, wyspecjalizowanych w badaniach ekonom. firmach, inst. badawczy tylko jeden - wyspecjalizowany w modelach ekonom.	Badania wydziałów ewaluacji zwykle dogłębne kwestie związane z danym Działaniem PO, prezentujące opinii publicznej wyniki; Raporty zewnętrznych ewalutorów zwykle szersze, łączące kilka działań, priorytetów, dostarczające dyskusji KM dany PO	Wyniki raportów mid-term i ex post wpłynęły na strategiczną orientację niektórych działań (by np. uniknąć biegu jałowego w MSP, zatrzymano część subsydiów); niektóre wywołały też publiczną debatę Zwykle jednak bieżące pomagały menedżerom Ex ante standardowo pomogły w planowaniu
Niemcy	Ewaluacja obecna w regionalnej polityce krajowej od l.1970; b. dobrze rozwinięta lecz słabo połączona z FS W polu FS brak silnej "tożsamości" programów UE - badania tych programów, jak i dyskusja wokół nich, głównie w wąskim gronie instytucji	Główna instytucje zlecające - ministerstwa Landów, zwykle departementy wdrażające FS; Brak koordynacji na poziomie Landów Studia zlecane zwykle ad hoc Konsekwentne budowanie doświadczeń w zlecaniu badań i ich kontroli	Duża i aktywna wspólnota ewaluacyjna Początkowo znaczna reprezentacja inst. badawczych, obecnie silna dominacja firm konsultingowych Dużo firm, duża konkurencja, mała fluktuacja, ciężko	Przewaga studiów wymaganych regulacjami FS, sporadyczne studia efektów Raporty zwykle zawierające użyteczną dla menedżerów, wiedzę operacyjną Jakość i przystępność raportów oceniana wysoko, choć zaznacza się, że	Używane do raportowania i weryfikowania efektów - wyniki dyskutowane jednak wewnętrznie, rzadko publicznie Najbardziej użyteczne - bieżące - zorientowane na procesy i rekomendacje nt wdrażania

Kraj	Kontekst i Perspektywy	Popyt	Podaż	Produkty	Wykorzystanie
	wdrażających Nic nie wskazuje na to, by ten trend się zmienił w nowej perspektywie		"wejść" na rynek, ewaluatorzy z długim często 10-20 letnim stażem Pole konsultacji i wymiany doświadczeń - DeGEval - Tow. Ewaluacyjne	metodologie praktycznie się nie zmieniły od lat (bazowanie na danych z monitoringu, wywiadach) Raporty prezentowane w internecie, nie ma scentralizowanej bazy, mechanizmu wymiany info, brak publikacji naukowych	Bardzo ograniczony wpływ na decyzje polityczne
Szwecja	Długa tradycja ewaluacji prog publ. oparta ekspertach i komisjach badawczych skupiających polityków, ekspertów, grupy interesu Ewaluacja FS weszła więc w tę tradycję, zmuszając jedynie pewnej modyfikacji terminologii (typy ewal.)	Mimo iż system zarządzania zmieniał się kluczowi aktorzy pozostają ci sami: Ministerstwo Przedsiębiorczości, Energii i Komunikacji oraz NUTEK - agencja rozwoju gospodarczego i regionalnego Zlecane są ewaluacje programów, tematyczne i bieżące - na 3 poziomach, organizacji programów, ich wdrażania (w ujęciu	Dość ograniczony, gdyż wymogi i organizacja kontraktów wymagają od firm dużej sieci współpracowników; Dominują prywatne firmy konsultingowe, uniwersytety są zwykle podwykonawcami	Liczba badań tematycznych znacząco rośnie zwykle badania horyzontalne prezentujące wnioski w przekroju przez wszystkie prog szwedzkich FS, z wnioskami zorientowanymi na konkretne polityki (zatrudnienia, MSP) Raporty powszechnie dostępne, choć reorganizacja stron www sprawia, że znikają; brak centralnej bazy z kolejnych okresów	Uzywane głownie przez inst. wdrażające FS, członków Komitetów Monitroujących Ich głowną funkcją jest raportowanie i dostarczanie uzasadnienia dla programów; ta refleksyjna funckaj słabo widoczna w praktyce widać głównie funkcję bieżącego zarządzania

Kraj	Kontekst i Perspektywy	Popyt	Podaż	Produkty	Wykorzystanie
		porównawczym między reg) oraz kluczowych projektów		programowania	
Wielka Brytania	Tradycja ewaluacji sięgająca połowy lat 1970-tych i reform menedżeryzmu Ewaluacja silnie zintegrowanym elementem polityk krajowych - wymagana dla wszystkich nowych inicjatyw publ. Pole ewaluacji FS postrzegane jako coś odrębnego od praktyk krajowych - zupełnie od nich wyizolowane	Zlecanie ewaluacji w przypadku Anglii - scentralizowane i koordynowane dla 12 regionów przez MA programów angielskich, w Szkocji i Walii zlecane przez biura regionalne - choć z pewnymi wytycznymi - popularne użycie poradników ewaluacji Powszechne stosowanie strategii ewaluacji, plany przewidują szereg ewaluacji bieżących jak i mid-termy w 2010r.; są też badania ad hoc jak i tzw. *rolling studies* (ciągłe -kilkuletnie kontrakty)	Wszystkie ewaluacje wykonywane przez niezależnych konsultantów Duży dynamiczny rynek, silną pozycję mają międzynarodowe firmy, choć jest też dużo drobnych firm konsultingowych; obecne są także instytuty Rosnący popyt wciąż stwarza perspektywy wejścia na rynek	Dużo raportów zwykle wpisujących się w przygotowywane przez rządy regionalne strategie ewaluacji Silna orientacja na kwestie bieżące, procesy wdrażania, ostatnio widoczne zorientowanie na badania innowacji, R&D, przedsiębiorczości jak i próby ewaluacji oddziaływania miękkich programów Coraz więcej ewaluacji projektów - szczególnie projektów strategicznych	Raporty używane zwykle przez menedżerów prog. i beneficjentów do rozwiązywania bieżących problemów ex ante i MTE zwykle do planowania nowych prog Częste studia CAV wykorzystywane do budowania dowodów nt wpływu/braku wpływu FS Wyniki badań często zasilają także dyskusje na forum Komitetów Monitorujących dane programy

Źródło: wyciąg z ekspertyz dr Richarda Boyle (przegląd praktyk irlandzkich), Lukas'a Wortmann'a (przegląd praktyk niemieckich), dr Kai'a Böhme (praktyki szwedzkie) oraz dr Martina Ferry (przegląd praktyk brytyjskich). Pełne teksty ekspertyz znajdują się w Aneksie 2.

1.2 Wyniki analizy metodologii 100 raportów z ewaluacji

Celem przeprowadzonej analizy metodologii raportów ewaluacji było po pierwsze scharakteryzowanie tendencji w metodologiach ewaluacji efektów programów Funduszy Strukturalnych (FS), zaś po drugie zidentyfikowanie i opisanie ciekawych praktyk.

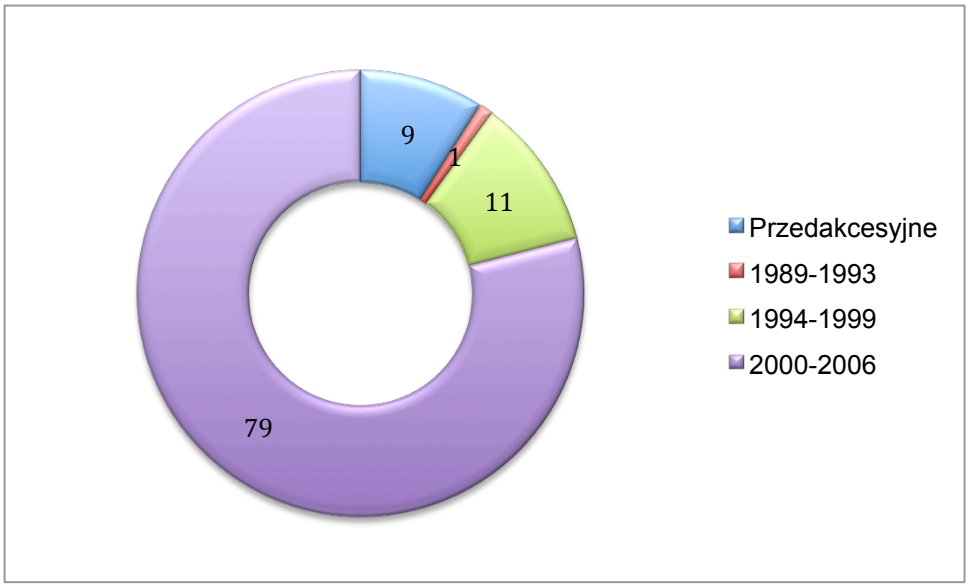
Tak jak przedstawiono w rozdziale poświęconym metodologii niniejszego badania, analizą objęto populację 100 raportów z 5 krajów (po 20 raportów z każdego kraju). Ta liczba stanowi praktycznie całą populację ewaluacji efektów programów regionalnych współfinansowanych przez UE w rzeczonych krajach. Wyniki poniższej analizy są więc reprezentatywne i mogą być traktowane jako całościowy obraz sytuacji w polu programów regionalnych współfinansowanych z UE, w tych 5 krajach. Zaś z racji tego, że wybrane państwa należą do krajów o najwyższej kulturze ewaluacji w Europie, poczynione obserwacje dają przesłanki do formułowania szerszych spostrzeżeń na temat całej Unii Europejskiej.

1.2.1 Charakterystyka analizowanej populacji raportów

Analizowana populacja wyniosła 100 raportów z 5 krajów: Irlandii, Niemiec, Szwecji, Wielkiej Brytanii (jako potencjalnych źródeł dobrych praktyk) oraz Polski.

W początkowych założeniach badawczych poszukiwano ewaluacji ze wszystkich kolejnych okresów budżetowych funduszy strukturalnych (lata 1989-93, 1994-99, 2000-2006). Dominują jednak raporty odnoszące się do programów z lat 2000-2006 (patrz: rycina 6). Wynika to po części z omawianego już w poprzednim rozdziale stopniowego ugruntowywania się praktyki ewaluacji w kolejnych okresach programowania. Główną przyczyną są jednak wymogi Rozporządzenia z lat 2000-06, które nakładały na stare państwa członkowskie przeprowadzenie ewaluacji średniookresowych (*mid-term*) oraz ich uaktualnień (*mid-term update*) dla wszystkich programów funduszy strukturalnych (Council of the European Union, 1999 art.42).

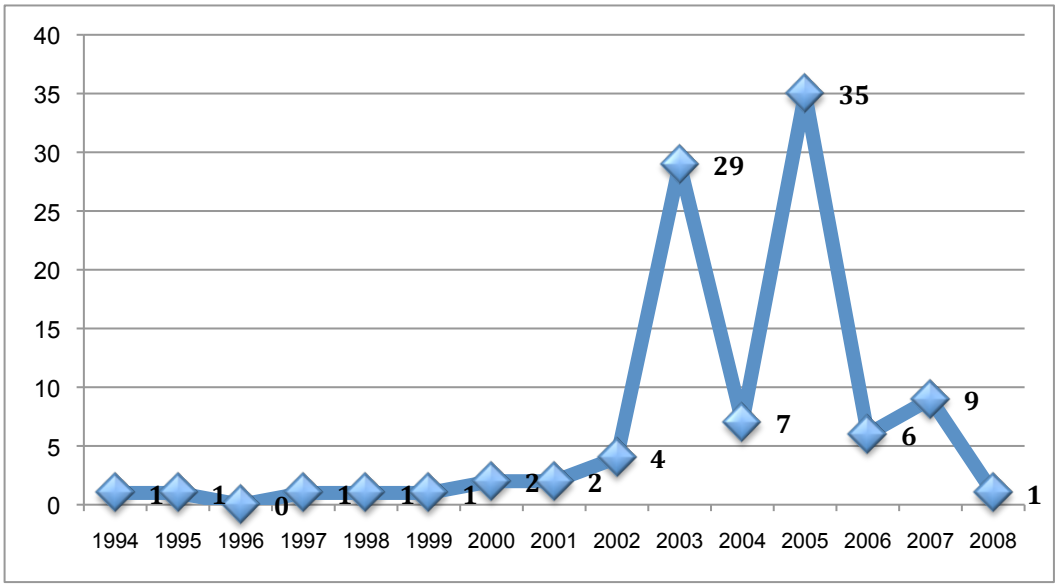
Rycina 6. Liczba badań według okresów programowania FS



Źródło: opracowanie własne na podstawie bazy danych projektu

Te "fale" ewaluacji średniookresowych napędzane wymogiem Rozporządzenia widać wyraźnie na poniższej rycinie. Najwyższe natężenie badań przypada na lata 2003 (aż 26 z 29 badań to *mid-term'y*) i 2005 (32 z 35 to uaktualnienia *mid-term*).

Rycina 7. Daty podejmowania badań ewaluacyjnych

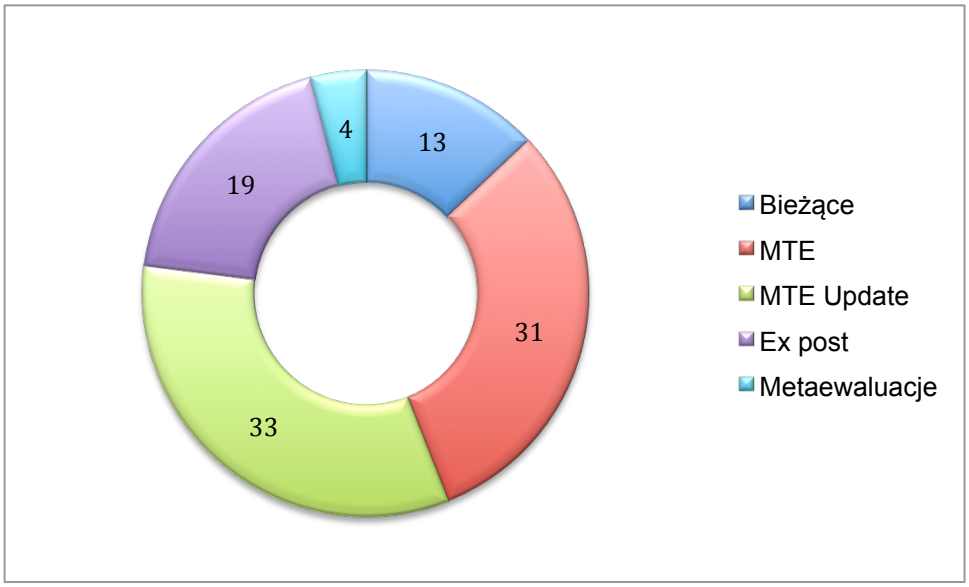


Źródło: opracowanie własne na podstawie bazy danych projektu

Tę dominację potwierdza także wyraźnie proste zestawienie typów czasowych (ryc. 8) - ponad 60% całej badanej populacji stanowią ewaluacje średniookresowe i ich

uaktualnienia. Różnica między liczbą ewaluacji średniookresowych a ich uaktualnień wynika z wyłączenia kilku raportów w procesie selekcji. Badania te nie dotyczyły one efektów a jedynie bieżących procesów.[8]

Rycina 8. Struktura typów analizowanych ewaluacji



Źródło: opracowanie własne na podstawie bazy danych projektu

Tylko 19 badanych przypadków było ewaluacjami ex post - a więc badaniami przeprowadzonymi po formalnym zakończeniu wdrażania danej interwencji (zwykle w 2 lata po). Aż 11 z nich jest ewaluacjami polskimi. Wynika to z faktu, że zgodnie z zapisami rozporządzeń funduszy strukturalnych odpowiedzialność za badania ex post spoczywa w pierwszym rzędzie na Komisji Europejskiej - DG-Regio. Natomiast w przypadku programów pomocowych zewnętrznych (pierwsze edycje PHARE oraz pomoc przedakcesyjna) odpowiedzialność za badania ex post była dużo częściej przesuwana przed EuropeAid na administracje krajów wdrażających daną interwencję. Należy jednak zaznaczyć, że w polu funduszy strukturalnych to zdjęcie odpowiedzialności za analizę ex post z administracji krajowych nie oznacza rezygnacji z analiz efektów interwencji. Możliwość bieżących badań zaawansowanych programów (decyzje na temat podjęcia badań *ad hoc* leżały w wyłącznej gestii instytucji wdrażających dany program), jak również uaktualnienie ewaluacji średniookresowych stwarzało szerokie pole praktyki oceny efektów. W

[8] Były to przypadki badań programów o bardzo późnym starcie (wydatkowanie środków w 2003 roku na poziomie 0%) - nie mogło być więc mowy o analizie jakichkolwiek efektów. Takie analizy pojawiły się dopiero przy uaktualnieniu mid-term, w 2005 roku.

przypadku mid-term i mid-term update ocena efektów była nawet wymogiem narzuconym wytycznymi Komisji.

Oczywiście wszystkie badania ex psot stawały przed problemem przeniesienia wniosków ze swoich analiz na nowe edycje programów. problem ten wynikał z omawianego już rozmijania się ewaluacji końcowych z planowaniem kolejnych edycji tej samej interwencji. W tym kontekście, w grupie ewaluacji ex post ciekawymi przypadkami są raporty z badania programu PEACE (Irlandia Północna) oraz PHARE CBC Polska-Niemcy, w których starano się połączyć ewaluację ex post z badaniem kolejnych edycji programów (patrz ramka).

Tabela 6: Przykłady łączenia typów czasowych ewaluacji

Celem programu PEACE było wparcie procesu pokojowego w Irlandii Północnej poprzez rozwój regionalny promowany prze lokalne partenrstwa projektowe. Instytucja wdrażająca starając się zapewnić ciągłość koncepcji pomiędzy edycjami programu połaczyła, w ramach jednego kontraktu ewalucję ex post programu PEACE I i ewaluację mid-term programu PEACE II.
Instytucja Wdrażająca program PHARE CBC 1999-2001 Polska-Niemcy postanowiła zapewnić transfer doświadczeń z granicy zachodniej do przygotowywanego w tym czasie Programu Sąsiedztwa Polska-Białoruś-Ukraina. W zleceniu na ewaluację ex post zawarto zastrzeżenie, by wnioski z badanai PHARe-CBC zawierały dobre praktyki i sugestie dla przyszłych beneficjenetów Programu wschodniego.

Źródło: opracowanie własne na podstawie bazy danych projektu i raportów z badań

Przechodząc do charakterystyki ewaluand (programów będących przedmiotem badanych raportów) prócz dominującej grupy programów regionalnych o charakterze multisektorowym (prawie połowa raportów), wyróżniają się też podgrupy raportów dotyczące programów o bardziej sprofilowanym polu interwencji: MSP & innowacji (20), współpracy transgranicznej (12), B&R (10) oraz transportu (10).

Wreszcie jeśli chodzi o finansową skalę interwencji, to raporty dotyczyły programów o średniej wartości 300 mln EUR. Grupa programów dużych (ponad 1 mld EUR budżetu) liczyła 28 przypadków (z czego aż 11 to interwencje z budżetem ponad 4 mld EUR).

1.2.2 Tworzenie koncepcji i metodologii badań

W tym punkcie analizy skoncentrowano się na koncepcji analizowanych ewaluacji. Zbadano z jakich inspiracji (literatury, metodologii, wcześniejszych badań) skorzystali

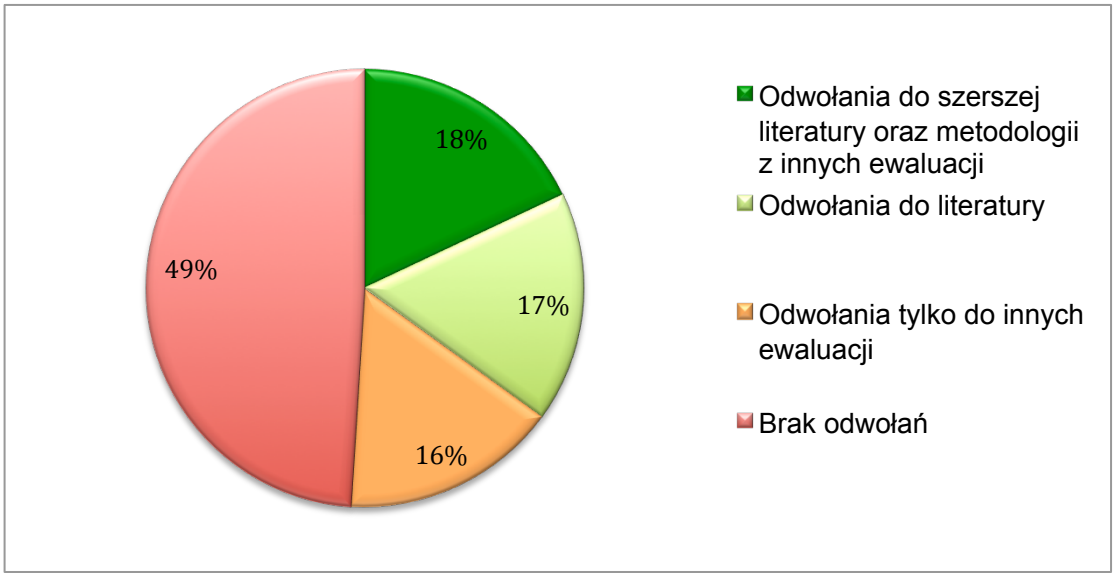
ewaluatorzy i po jakie narzędzia sięgnęli by wspomóc kluczowy etap badania - strukturalizację.

Źródła inspiracji przedstawia rycina 9. Wyniki są dużym rozczarowaniem. Prawie połowa analizowanych badań nie miała żadnych udokumentowanych podstaw całościowej koncepcji badawczej (sic!). Kolejne 15% bazowało wyłącznie na pomysłach z wcześniejszych, analogicznych badań.

Tylko niewielki odsetek badań (18%) był budowany na szerszych podstawach metodologicznych i koncepcyjnych - zarówno teoretycznych (podręczniki, poradniki) jak i praktycznych (inne badania). *Gros* z tych badań pochodzi ze Szwecji (7) i - co pocieszające - Polski (również 7).

Przyglądając się szerszej grupie badań, które odwoływały się do literatury metodycznej (35% badanej populacji) stwierdzić można, że w prawie połowie przypadków (44%) były to wyłącznie takie źródła jak: MEANS, Evalsed oraz - dużo częściej - zeszyty robocze Komisji (*Working Papers*). Tych ostatnich nie można jednak traktować jako poważnego "źródła metodologicznego", gdyż są to zwykłe wytyczne administracyjne określające minimum pytań i metod. Podsumowując tę obserwację trzeba stwierdzić, że inspiracje metodologiczne były wyjątkowo słabe.

Rycina 9. Źródła wykorzystane przy tworzeniu metodologii

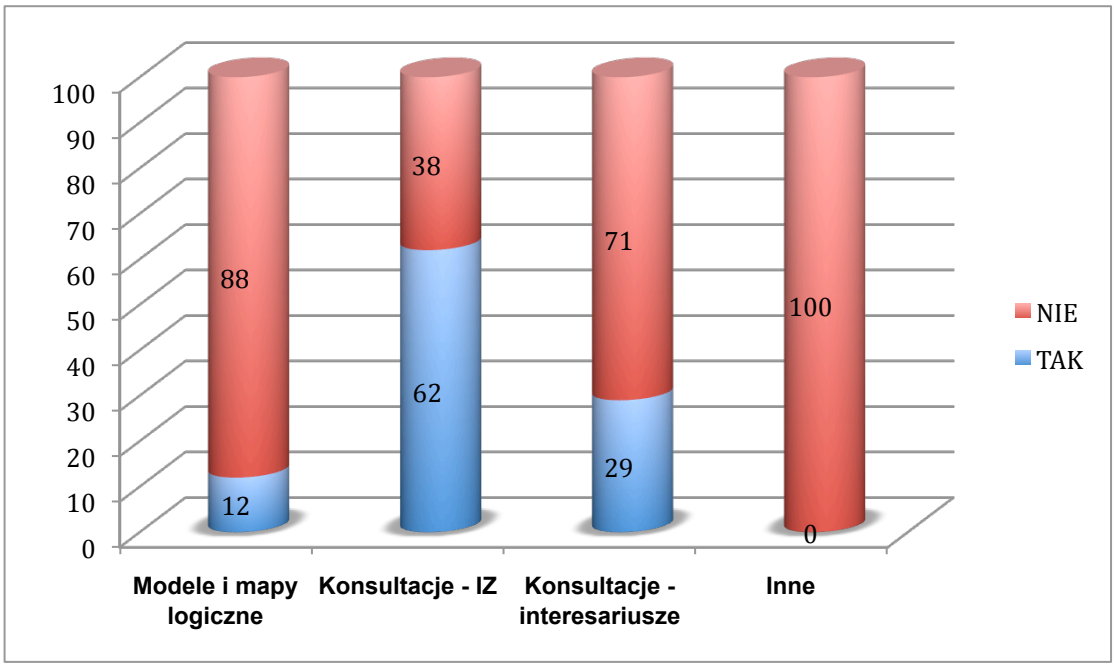


Źródło: opracowanie własne na podstawie bazy danych projektu

Kwestie narzędzi użytych w fazie strukturalizacji badania prezentuje to rycina 10. Trzeba stwierdzić że przytłaczająca większość ewaluacji ograniczyła się do standardu "porozmawiania ze zlecającym", a czasem również innymi

interesariuszami - Grupą Sterującą Ewaluacją, kluczowymi instytucjami programu, interesariuszami z poziomu Komisji Europejskiej. Uderzający jest całkowity brak innowacyjności. W analizowanych raportach nie pojawił się żaden przypadek metody strukturalizacji spoza wymienionych trzech opcji (a właściwie dwóch - bo opcje dwa i trzy różnią się tylko rozmówcą).

Rycina 10. Narzędzia użyte na etapie strukturalizacji



Źródło: opracowanie własne na podstawie bazy danych projektu

Warto natomiast zwrócić uwagę na nieliczną grupę przypadków, w których użyto modeli i map logicznych (narzędzie to jest szerzej omawiane w rozdziale 2.2 niniejszego raportu). Ewaluatorzy opierali się zwykle na standardowym modelu Komisji,[9] adaptowali go jednak do specyfiki programu (przykład brytyjski - patrz tabela i rycina poniżej), a czasem dokonywali nawet znaczących przekształceń. Modele pomagały zwykle uporządkować hierarchię celów i efektów programu, rozgraniczyć pomiędzy produktami a efektami odnosząc je do wskaźników (było to szczególnie pomocne gdyż ich interpretacja była często zniekształcana w trakcie wdrażania przez beneficjentów) i wreszcie połączyć koncepcję analizy z kryteriami

[9] Chodzi o dyskutowany już model: potrzeby-cele-nakłady-działania-produkty-rezultaty-oddziaływanie i powiązane z nim kryteria ewaluacyjne (porównaj: European European Commission, 2000a, s.9)

oceny. W dwóch przypadkach szwedzkich model pokazywał także ciąg przyczynowo-skutkowy interwencji i posłużył jako "szkielet" ustrukturalizowania całego badania (wszystkich jego pól analizy) a potem oceny (przykład ten omawiany jest w dalszej części opracowania).

Tabela 7. Przykład wykorzystania modelu w koncepcji badania

W raporcie Średniookresowym Programu Wschodnia Walia [204] oparto się na standardowym modelu Komisji Europejskiej (nakłady - działania - produkty). Wprowadzono jednak pewną modyfikację. Wyróżniono aspekt odgórny i oddolny. Do tego pierwszego przypisano oddziaływanie, oceniane według kryterium trwałości efektów i mierzone zmianami w trendach regionalnych i kontekście programu. Do tego drugiego przypisano rezultaty oceniane według ich użyteczności dla beneficjentów ostatecznych, a mierzone stopniem spełenienia oczekiwań i zaspokojenia potrzeb. Obydwa te wymiary składały się na Wartość Dodaną Wspólnoty.

Źródło: (CSES, 2003, s.13-14)

1.2.3 Metody zbierania danych

W tym punkcie analizowane były źródła i metody zbierania danych wykorzystane w 100 ewaluacjach efektów. Zestawienie przedstawia rycina 11.

Z zestawienia wynika, że najczęściej opierano się na danych wtórnych - przetwarzaniu baz danych z monitoringu programu, ogólnodostępnych krajowych i regionalnych statystykach (kategoria: inne dane wtórne), a także na przeglądzie zapisów i dokumentów programowych, wniosków o dofinansowanie, wniosków aplikacyjnych beneficjentów, raportów rocznych, etc. (inne dane administracyjne).

W tym kontekście kluczowym wydaje się jakość baz danych z monitoringu programów.

Tabela 8. Przykład użyteczności baz danych monitoringowych

W irlandzko-angielskim programie PEACE I & II stworzenie i prowadzenie bazy monitoringowej zostało zlecone przez Instytucję Zarządzajacą prywatnej firmie. Baza została zaprojektowana według podstawowych segmentów modelu logicznego interwencji. Miała 5 segmentów. Pierwszy (nakłady) zawierała informacje o funduszach programu, jak i dodatkowych źródłach finansowania. Moduł drugi miał wskaxniki procesu - liczbę partnerstw, wybieraną orientację projektu, info o równych szansach (w kontkeście proporcji irlandzko-byrtyjskich). Moduł trzeci - działań i pdoruktów zawierał dane o projektach. W module czwartym umieszczano dane nt rezultatów lokalny wzrost zarobków, miejsc pracy, rewitalizacja, a także info o stopniu społecznej inkluzji - integrację grup docelowych, rozwój

lokalnych wspólnot. Wreszcie ostatni moduł oddziaływania programu zawierał dane nt wzmacniania efektu współpracy (np. kolejne projekty) jak i dowodów promocji pojednania. Tak ustrukturalizowana baza umożliwiła ewaluatorom kompleksową analizę stopnia realiacji celów programu, jak również postawienie pytań o zasadność niektórych działań.

Źródło: opracowanie własne na podstawie bazy danych projektu i raportów z badań (Annex C2)

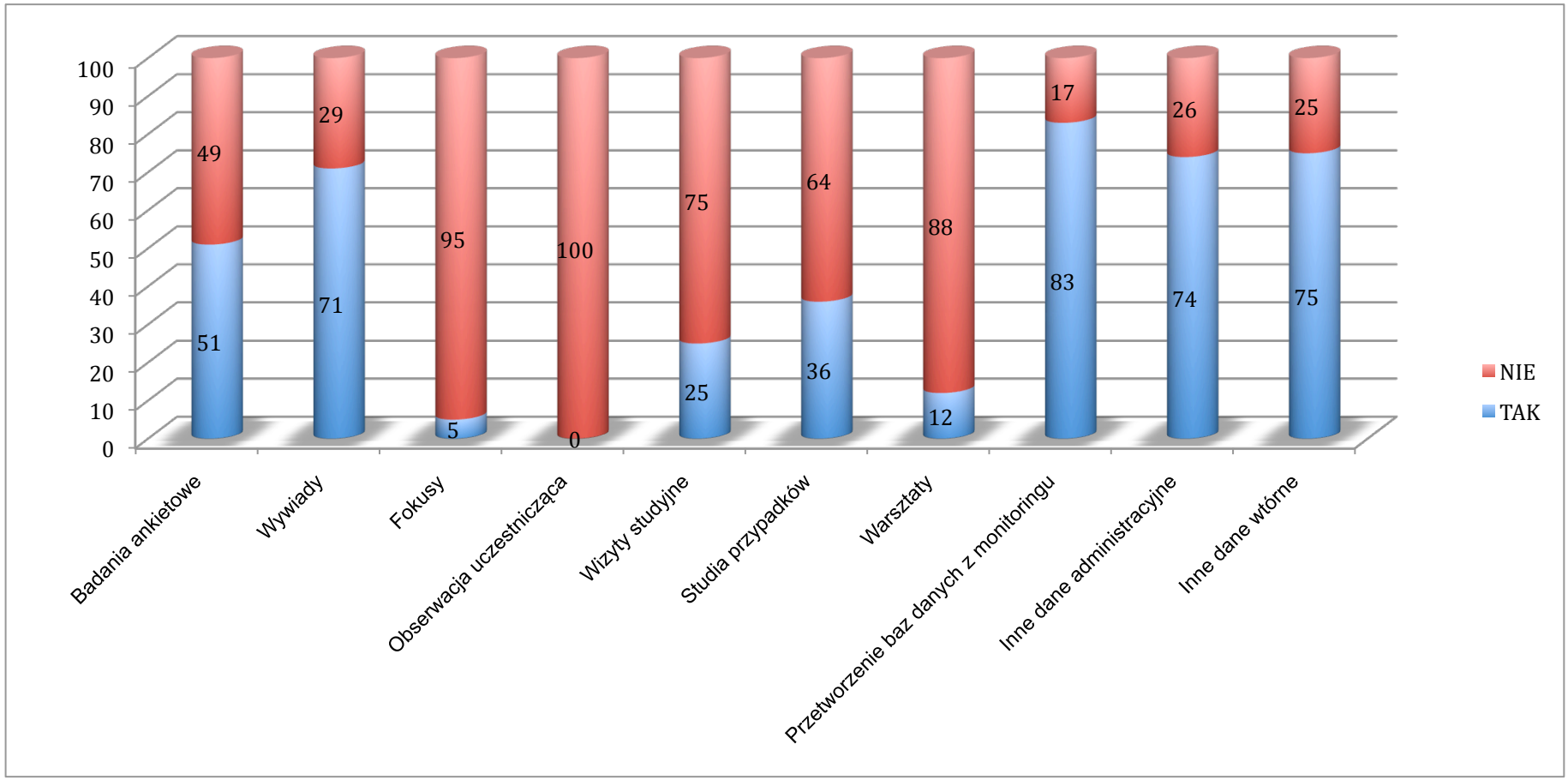
W przypadku danych pierwotnych przeważały wywiady oraz badania ankietowe. Te pierwsze prowadzono najczęściej z kluczowymi osobami w systemie wdrażania danego programu, wybranymi menedżerami projektów lub beneficjentami (zwykle jako element studiów przypadków), koordynatorami partnerstw (programy transgraniczne). Ankietowanymi natomiast byli zwykle bezpośredni beneficjenci projektów (tzw. beneficjenci końcowi), instytucje aplikujące o fundusze, których wnioski zostały odrzucone w procesie selekcji, a w kilku przypadkach także potencjalni beneficjenci, którzy nie wzięli udziału w konkursach aplikacyjnych. Badania beneficjentów ostatecznych (osób, społeczności, firm, które odczuwały skutki projektów) zdarzały się niezmiernie rzadko.

Pojawiło się także prawie 40 sytuacji korzystania ze studiów przypadków jako metody zbierania (a potem i analizy) danych. Były one zwykle łączone z wizytami studyjnymi projektów, choć w dwóch badaniach wizyty terenowe były sposobem zidentyfikowania projektów do studiów przypadków. Przedmiotami studiów przypadków były zwykle projekty, zdarzały się jednak badania partnerstw beneficjentów (np. instytucji czy sieci instytucji, które wygrywały po kilka projektów), a jednym przypadku przedmiotem analizy były nie projekty lecz wybrane układy przestrzenne (gminy) objęte programem. Średnia liczba studiów przypadków wahała się między 5 a 10. Choć było też 6 ewaluacji z ponad 20 studiami. Absolutnym ewenementem była polska ewaluacja programu RAPID, w której zrealizowano 55 studiów przypadków. Miały one jednak charakter raczej opisowych informacji o projekcie niż dogłębnych studiów. Ten przypadek pokazuje zależność w której większa liczba case'ów wiązała się z ograniczeniem i spłyceniem metod eksploracji np. ograniczono się do wywiadów telefonicznych i krótkiego przeglądu dokumentów.

Stosunkowo rzadką metodą były grupy fokusowe. W dwóch z pięciu przypadków przybrały one formę dyskusji na forum grup roboczych zorganizowanych według priorytetów programu.

Zaobserwowane, omawiane powyżej ciekawe praktyki z pola zbierania danych zaprezentowano syntetycznie w tabeli 9.

Rycina 11. Metody i źródła użyte na etapie zbierania danych



Źródło: opracowanie własne na podstawie bazy danych projektu

Tabela 9. Ciekawe przykłady metod zbierania danych

Program	Metoda / narzędzie	Opis
Ewaluacja ex post Programu PHARE-CBC 1994-99 Polska-Niemcy Ewaluacja ex post Programu PHARE-CBC 1999-2001 Polska-Niemcy	ankieta szkolna kwerenda prasy lokalnej	Aby ocenić wpływ programu PHARE-CBC na świadomość mieszkańców pogranicza przeprowadzono ankietę szkolną w tzw. miastach podzielonych (np. Zgorzelec - Goerlitz). Nauczyciele w wybranych szkołach podstawowych rozprowadzili wśród dzieci ankiety - te zaś przekazały je rodzicom. Zwrot ankiet był bardzo wysoki. Ich wyniki natomiast b.zblizone do alternatywnego badania reprezentatywnej grupy polaków z pogranicza. W przypadku drugiego badania, prócz ankiet szkolnych przeprowadzono kwerendę polskiej i niemieckiej prasy lokalnej - aby zorientować się w jakim kontekście i kiedy (jeśli w ogóle) pojawiają się wzmianki o programie PHARE.
South Yorkshire Objective 1 Programme Mid Term Evaluation [209]	Badania społeczne Warsztaty	Na potrzeby ewaluacji przeprowadzono badanie ankietowe 160 interesariuszy programu w podregionie - NGOsów, instytucji edukacyjnych, władz lokalnych, stowarzyszeń biznesu. Pytania dotyczyły wskazania i oceny ewentualnych widocznych efektów programu. Warsztaty zorganizowanow trzech głównych polach. Pierwszym były warsztaty z trzema strategicznymi, tematycznymi grupami programu. Celem było - bazuąc na pierwszych wnioskach - zidentyfikować kluczowe pola korekt programu. Drugim były spotkania z każdym z Lokalnych Strategicznych Partnerstw (wdrażających projekty) gdzie dyskutowanoa implikacje programu dla danych obszarów lokalnych. Wreszcie ostatnia fala warsztatów dotyczyła kwestii horyzontalnych.

Program	Metoda / narzędzie	Opis
		Dyskutowali eskperci wspierani przez przedstawicielie danych regionów.
PHARE SSG 2003 - Komponent regionalny [008]	Wywiady	W ramach badania przeprowadzono między innymi wywiady z grupą wielokrotnych beneficjentów programu, po to by zrozumieć mechanizm ich sukcesu ale również motywy stojące za konsekwentnym udziałem w programie
Szwedzka ewaluacja bieżąca pt. Fundusze strukturalne UE – narzędzie rozwoju regionalnego [302]	Studia przypadków	Badanie jest przykładem ciekawego, nietypowego wyboru studiów przypadków. Zostały one wybrane na podstawie analizy danych statystycznych. Jako dwa studia wybrano gminy o podobnych uwarunkowaniach rozwojowych, ale różnych efektach programów strukturalnych UE. W ramach case'ów zrealizowano 17 pogłębionych wywiadów.

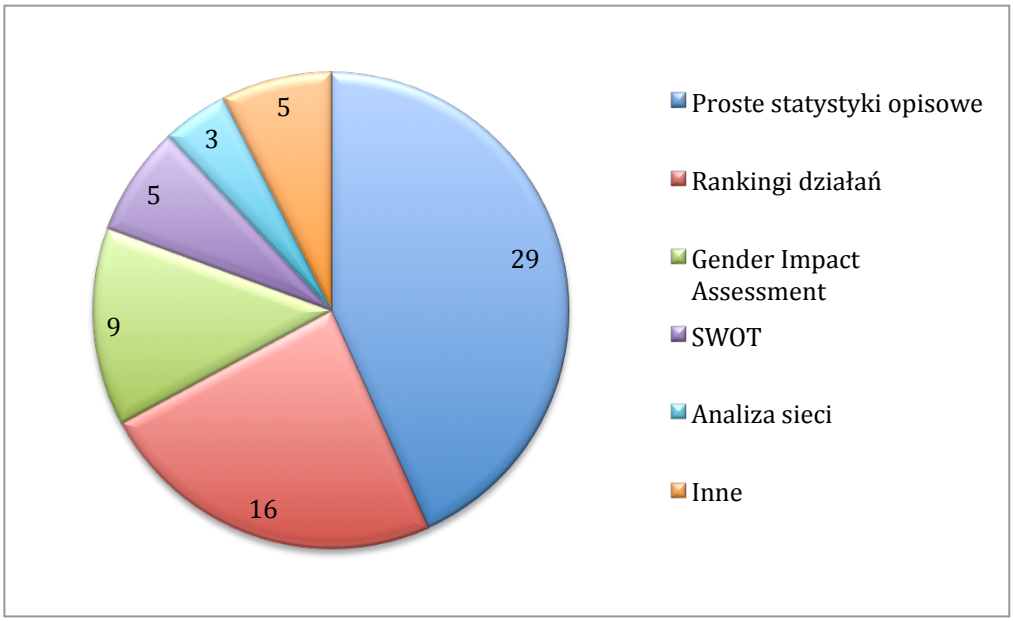
Uwaga: W nawiasach kwadratowych [ ] podano nr porządkowy danej ewaluacji w bazie danych projektu

Źródło: opracowanie własne na podstawie bazy danych projektu oraz raportów z badań

1.2.4 Metody analizy i oceny

W kolejnej części opracowania podjęto temat użycia metod w ostatnim etapie badań ewaluacyjnych - analizie i ocenie. Użycie popularnych metod przedstawiono na rycinie 13 (patrz: następna strona). Jedyną wybijającą się grupą są raporty szwedzkie używające analizy wielokryteriowej i analizy nakładów-produktów. Choć trzeba tu zaznaczyć, że część badanych raportów za analizę nakładów-produktów podawała zwyczajne zestawienie kosztów projektów i ich produktów, co ma niewiele wspólnego oryginalnym rozumieniem tej metody. Zaskakującym faktem jest iż mimo, że analizy dotyczyły zjawisk przestrzennych to po narzędzia analizy przestrzennej sięgano sporadycznie. Brak GIS można wytłumaczyć kosztami, barierą technologiczną, czy czasochłonnością wprowadzania danych natomiast uderzający jest brak zwykłych map! Mapy, jeśli już się pojawiały, były tylko tłem - ilustracją wprowadzającą do analizy. Wyjaśnienia wymaga co znalazło się w licznej grupie "inne". Jej strukturę przedstawiliśmy poniżej.

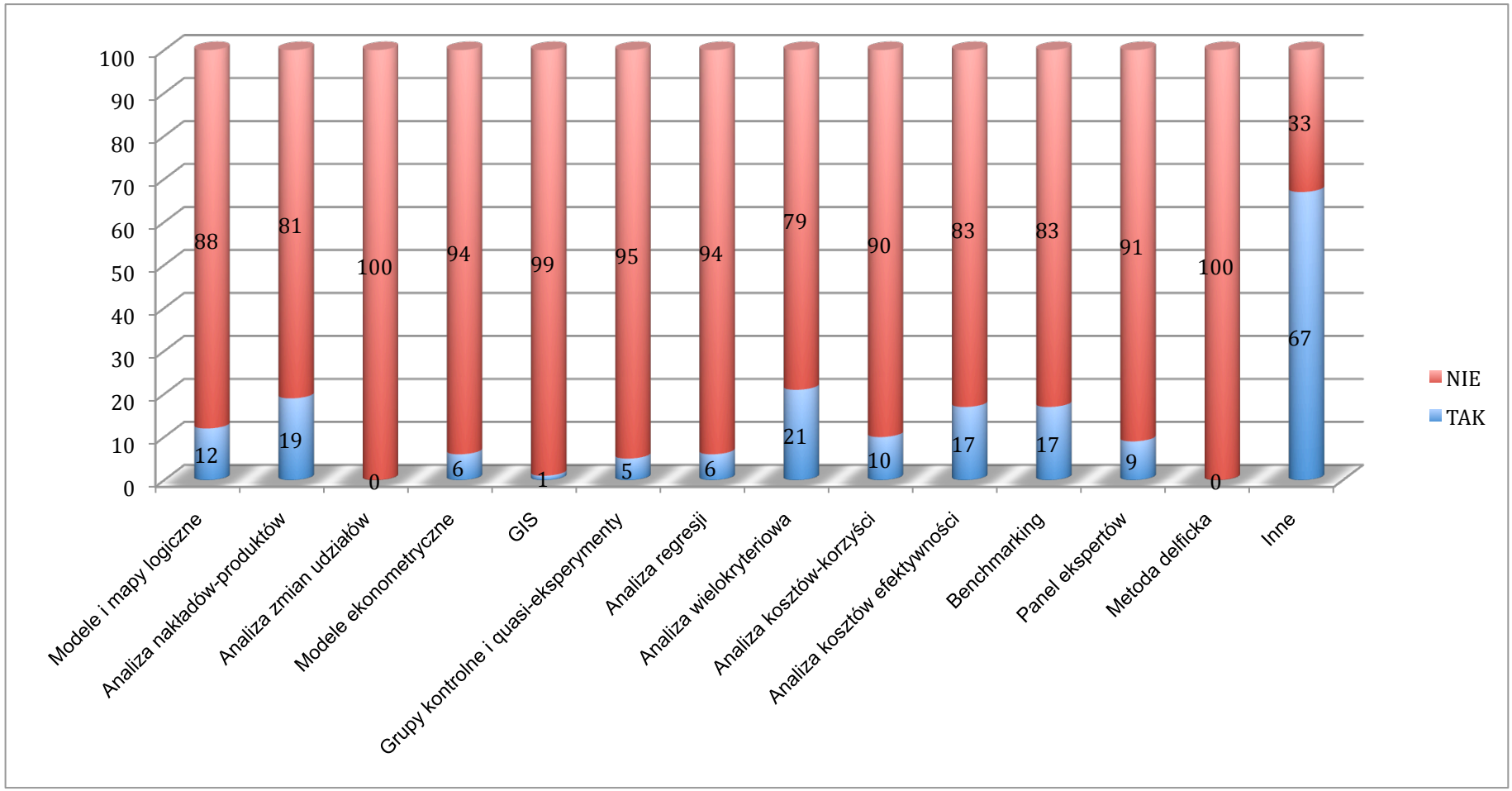
Rycina 12. Inne metody analizy i oceny



Źródło: opracowanie własne na podstawie bazy danych projektu

Jak widać *gros* badań oparło się na zwykłej opisowej statystyce trendów społeczno-ekonomicznych w polu - sektorze działania programu ale bez powiązania ich z efektami programu, zestawienia wydatków finansowych i produktów. Co wyjątkowo rozczarowuje to fakt, że prób badania relacji przyczynowych było dosłownie kilka, zaskakujących rozwiązań metodologicznych również niewiele. Ich podsumowanie przedstawiono w tabeli 10.

Rycina 13. Metody użyte do analizy i oceny



Źródło: opracowanie własne na podstawie bazy danych projektu

Tabela 10. Ciekawe przykłady metod analizy i oceny

Program	Metoda / narzędzie	Opis
Irlandia - The Programme for Research in third level Instititions (PRTLI) - Impact Assessment [118]	Analiza cytowań i ocena peer review	Przeprowadzono analizy bibliometryczne - intensywność cytowań publikacji wydanych przez pracowników afiliowanych do danych instytucji - beneficjentów programu. Ponadto międzynarodowi uczestnicy komitetu oceniali po 10 publikacji pod kątem świadomości istnienia danego centrum badawczego w świecie.
Szwecja - Fundusze strukturalne, przedsiębiorczość i kondycja przedsiębiorstw. Doświadczenia z programów funduszy strukturalnych 2000-2006 [319]	Typologie	Dla firm biorących udział w programie stworzono typologię przedsiębiorstw wg. faz ich funkcjonowania (przetrwanie, dynamiczny rozwój, dojrzałość). Do tego ułożono typologię idealnych programów pomocy publicznej dla każdej z faz (odpowiednio: doradztwo, finansowanie, działania sieciowe, etc.), a następnie w oparciu o dane statystcyzne klasifikowane poszczególne firmy na grupy. Tym samym sprawdzono przystawalność programu do rzeczyistych potrzeb.
Szwecja - Fundusze strukturalne UE – narzędzie rozwoju regionalnego [302]	Porównania	Zastosowano metodę *difference-in-difference*; porównano gminy, gdzie realizowano projekty z FS UE i te, gdzie ich nie realizowano; zbadano szeregi czasowe danych statystycznych jak też przeprowadzono panele dyskusyjne
Irlandia - Evaluation of Agency Supports for R&D Performed in the Business Sector [116]	Zestawienia	Zweryfikowano listę 300 przedsiębiorstw sklasyfikowanych jako "top w B+R" (150 irlandzkich i 150 z własnością zagraniczną) pod kątem ich uczestnictwa w badanych programach wspierających B+R.

Uwaga: W nawiasach kwadratowych [ ] podano nr porządkowy danej ewaluacji w bazie danych projektu

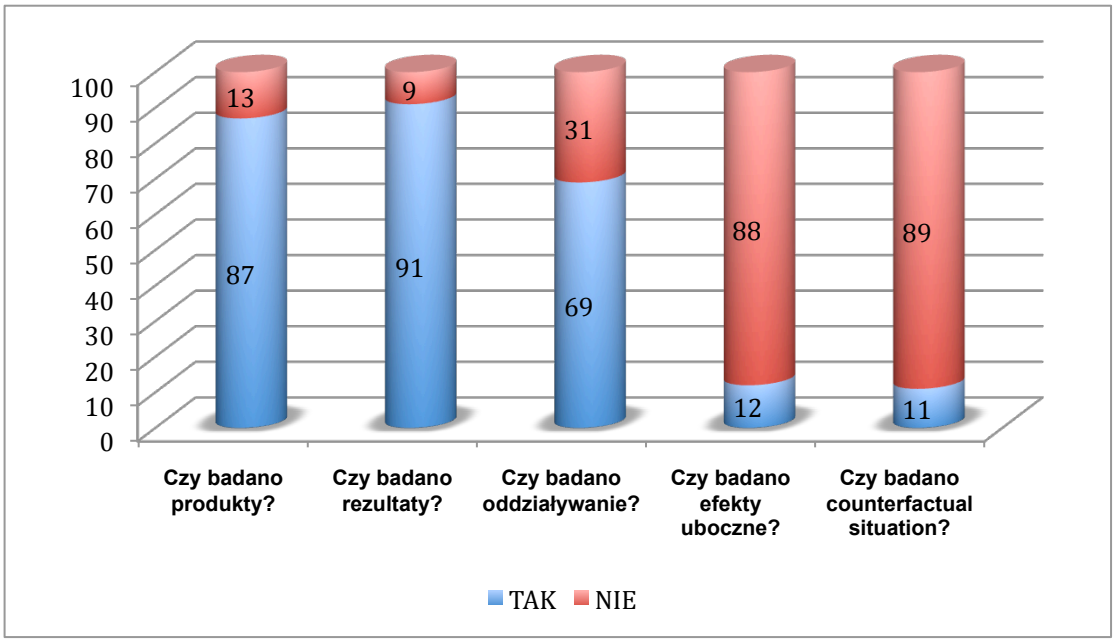
Źródło: opracowanie własne na podstawie bazy danych projektu oraz raportów z badań

1.2.5 Orientacja wyników i rekomendacji z badań

Ostatnim elementem analizy raportów ewaluacji efektów było całościowe spojrzenie na to co w tych badaniach zostało poddane analizie i ocenie oraz na co zostały ukierunkowane ich wnioski i rekomendacje.

Pierwsze z zestawień prezentuje przedmiot analizy według nazewnictwa użytego w poszczególnych raportach.

Rycina 14. Przedmiot prowadzonych badań



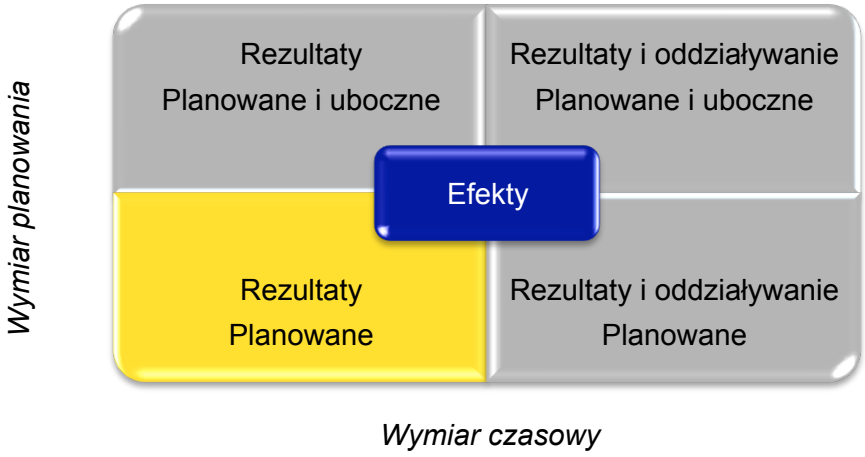
Źródło: opracowanie własne na podstawie bazy danych projektu

Interesująca jest przewaga badań rezultatów nad produktami. Po części wynika to z tego, że w kilku przypadkach analiza rozpoczynała się pod poziomu rezultatów (produkty programu były po prostu podane). Jednak ta przewaga rezultatów (jak i duża liczba analiz oddziaływania) jest pozorna. W rzeczywistości to co nazywane było przez ewaluatorów rezultatami lub oddziaływaniem programu, zwykle - po uważnej lekturze i zestawieniu z właściwymi definicjami okazywało się produktami programu a w najlepszym przypadku krótkookresowymi rezultatami. Wynikało to głównie z tego, że jako produkt traktowano liczbę projektów. Analiza produktów była więc zwykłym zestawieniem statystyk zrealizowanych projektów, podczas gdy ich produkty (np. liczba przeszkolonych osób, zbudowana infrastruktura) były interpretowane już jako rezultat. Przy takim zabiegu *gros* studiów skoncentrowało się

na opisowej statystyce unikając kwestii rzeczywistych efektów. Potwierdzają to sporadyczne przypadki analizy sytuacji bez interwencji (oparte zwykle na prostej ekstrapolacji trendów np. case szwedzki - 302) oraz ewentualnych efektów ubocznych.

Odnosząc niniejszą analizę orientacji raportów do dwóch wymiarów efektów przedstawionych we wstępie do tego opracowania, trzeba stwierdzić, że analizowane studia mieszczą się w lewym dolnym rogu schematu. Zaś jeśliby zastosować ostrzejszą definicję rezultatów i oddziaływania - studia te znalazłyby się w ogóle poza schematem - to znaczy nie byłyby studiami efektów.

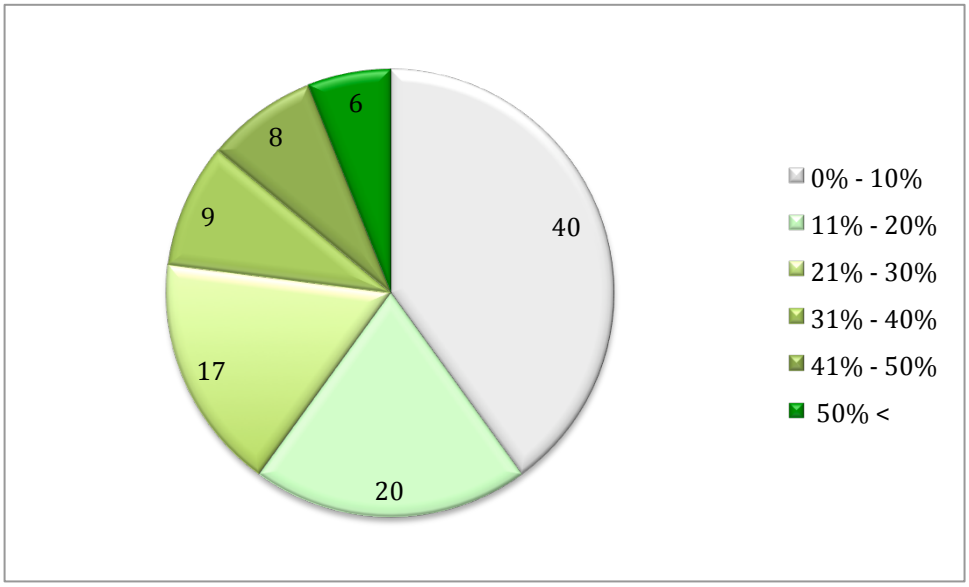
Rycina 15. Dwa wymiary efektów interwencji



Źródło: opracowanie własne

Dodatkowym źródłem weryfikacji tego fenomenu jest przeprowadzona analiza orientacji wniosków i rekomendacji. Kolejna rycina prezentuje liczbę raportów według odsetka ich wniosków, poświęconych efektom. Tylko 6 raportów poświęciło ponad 50% swoich wniosków i rekomendacji kwestii efektów. Przytłaczająca większość odwoływała się do efektów tylko w co dziesiątym wniosku. Wykres rażąco przedstawia dominację kwestii procesowych i bieżących nad kwestiami strategicznymi. W rzeczywistości były to ewaluacje zorientowane na procesy.

Rycina 16. Liczba raportów wg odsetka wniosków poświęconych efektom



Źródło: opracowanie własne

Wreszcie ostatnią obserwacją jest fakt, że tylko w czterech przypadkach (sic!) ewaluatorzy odwołali się do teorii i szerszych koncepcji rozwoju. Obserwacja ta potwierdza wyjątkowo techniczny charakter przeprowadzonych ewaluacji.

1.3 Wnioski z analizy

Podsumowując obserwacje z przeprowadzonego przeglądu 100 badań efektów Funduszy Strukturalnych można postawić następujące trzy wnioski.

Po pierwsze, odnosząc analizowane badania do definicji metodologii przedstawionej we wstępie (metodologia = koncepcja badania + zestaw narzędzi i metod), trzeba stwierdzić, że przytłaczająca większość z nich pozbawiona była metodologii rozumianej w sposób całościowy! Ewaluacje te cechowały się zaskakująco niskim poziomem koncepcyjnym - praktycznie brakiem spójnej, przemyślanej koncepcji badawczej (podejścia badawczego) osadzonej w literaturze tematu. Metodologia ograniczała się zwykle do doraźnego, sztampowego zestawu metod i narzędzi. Co więcej, poza kilkoma przypadkami (zwykle ewaluacji opartych na modelach) brakowało jakichkolwiek szerszych podejść badawczych jak i głębszych wyjaśnień - uzasadnienia wyborów metodologicznych.

Po drugie proponowany zestaw metod i narzędzi właściwie ograniczał się do kilku podstawowych elementów - badań ankietowych beneficjentów (zwykle końcowych,

nie ostatecznych), wywiadów z zarządzającymi, ogólnej statystyki, *desk research* i danych z monitoringu. Metody analizy były praktycznie ograniczone do prostych statystyk opisowych. Co więcej ich opis był zwykle bardzo oględny.

Słabość metodologii, a w najlepszym przypadku, pomijanie przez autorów jej jasnego uzasadnienia i wyjaśnienia są wyjątkowo niepokojące, gdyż podważają wiarygodność tych badań. Tymczasem zarówno teoria jak i badania empiryczne jasno wskazują, że metodologia ewaluacji (jej przejrzystość, spójność i jakość) jest głównym czynnikiem warunkującym jakość i wiarygodność wyników i rekomendacji, a co za tym idzie, ich wykorzystanie w sektorze publicznym (Cousins, Leithwood, 1986; Ferry, Olejniczak, 2008; Weiss, Bucuvalas, 1980).

Te dwa pierwsze punkty znajdują swoje odzwierciedlenie w kwestii trzeciej - ograniczonych wnioskach z analizowanych badań. Można spytać retorycznie - cóż można powiedzieć o efektach skomplikowanych interwencji publicznych bez spójnej strategii badawczej i jasno uzasadnionych metod? No cóż, niewiele. Potwierdza to uderzająca wręcz orientacja badań na kwestie bieżące, procesy, postępy w wydawaniu pieniędzy. Jeśli już pojawiają się informacje o produktach czy rezultatach to mają charakter opisowy. Przy ocenie ewaluatorzy zwykle prezentują stopień realizacji celów programu, praktycznie nigdy ocenę tego, czy uzyskano pożądaną zmianę społeczną.

Gdzie można szukać przyczyn tej słabości metodologicznej ewaluacji funduszy strukturalnych zorientowanych (przynajmniej formalnie) na ocenę efektów? Moim zdaniem wyjaśnienia można upatrywać w pięciu kwestiach.

Pierwszą jest szersza polityka i formalny podział odpowiedzialności. Zgodnie z zapisami rozporządzeń nt. Funduszy Strukturalnych ewaluacja ex post (a więc całościowa refleksja nad efektami FS) leży w kompetencji Komisji Europejskiej. W praktyce administracje krajowe rzadko interesują się prawdziwymi efektami programów UE, gdyż po pierwsze nie są do tego prawnie zobowiązane, a po drugie uwaga krajowych polityków i opinii publicznej skierowana jest na sukces w szybkim wydawaniu zewnętrznych środków unijnych. Miernikiem sukcesu według którego rozliczana jest administracja jest poziom absorpcji, a nie jej strukturalny charakter. W rezultacie transfer krajowych praktyk ewaluacji do pola funduszy jest minimalny. Zlecane studia koncentrują się więc na tym co potrzebne urzędnikom krajowym - doradzaniu w usprawnieniach procedur i ocenie płynności wydatkowania pieniędzy.

Drugim wyjaśnieniem jest strukturalny błąd zakresu ewaluacji średniookresowych i ich uaktualnień (które stanowiły większość analizowanej populacji raportów). Wymagany przez Komisję zakres zadań obejmował zarówno ocenę bieżących postępów i pomoc w usprawnieniu procesów wdrażania jak i ocenę efektów. Tymczasem, jak dowodzą studia empiryczne (Eser, Nussmueller, 2006; Olejniczak, 2008) te dwie funkcje - operacyjna i strategiczna są nie do pogodzenia w jednym badaniu. W sytuacji gdy trzeba było wbierać ewaluatorzy koncentrowali się na doradzaniu w doraźnych procesach gdyż te były ważniejsze z perspektywy Zlecających. Co więcej, były też dużo prostsze i mniej politycznie delikatne niż ocena efektów.

Za trzecią przyczynę można uznać zastosowanie bardzo zawężającej definicji ewaluacji efektów - zorientowanej na programowe cele (tzw. *objective-oriented approach*). Praktycznie wszystkie badania (poza kilkoma, które analizowały efekty uboczne) sprawdzały stopień realizacji zapisów programowych (czy zrealizowano zapisany cel), bez wchodzenia w trudną analizę relacji między pożądanymi zmianami społecznymi a rzeczywistymi efektami działań. Taka szersza refleksja wymagałaby zresztą spójnej, opartej na teoriach zmiany, koncepcji badawczej - a tej w większości badań zabrakło.

Słabość metodologii (rozumianej całościowo - jako koncepcja oraz metody) i uderzający wręcz brak zakorzenienia w szerszej literaturze, można tłumaczyć czwartym czynnikiem - dominacją podejścia konsultingowego nad podejściem bardziej naukowym. Ewaluatorzy - pochodzący zwykle z firm doradczych (tylko w 20% analizowanych przypadków pojawiały się instytucje naukowe) przyjmowali zwykle rolę konsultantów, bo była ona łatwiejsza i nie wymaga specyficznej wiedzy odnoszącej się do danego pola. Tymczasem przy ocenie efektów wymagana jest rola nauczycieli-mentorów - oparta na dogłębnej znajomości tak teoretycznej jak i praktycznej danego pola interwencji (Ferry, Olejniczak, 2008, s.74-75).

Dodatkowym, ostatnim wyjaśnieniem bardzo ograniczonych metodologii badań może być brak literatury poświęconej metodologiom funduszy strukturalnych. Zaskakujące iż na rynku europejskim - w ofercie renomowanych wydawnictw - nie ma praktycznie publikacji naukowych poświęconych metodologii badań funduszy strukturalnych. Co więcej w periodykach branżowych artykuły o ewaluacji programów UE (metodologii tych badań) są sporadyczne.[10] Oczywiście może tu działać mechanizm sprzężenia

[10] W periodyku *Evaluation* w ostatnich 10 latach pojawiły się tylko 3 artykuły związane z funduszami strukturalnymi. Kwestie pieniędzy UE w kontekście rozwoju regionalnego

zwrotnego - ponieważ nie robi się badań o bardziej naukowej orientacji więc nie powstają później publikacje z tego zakresu, zaś brak publikacji utwierdza konsultingowe, proste podejście do metodologii.

Podsumowując można stwierdzić, że analizowane badania nie tworzą wiarygodnej bazy do budowania szerszej wiedzy o efektach funduszy strukturalnych, a co za tym idzie - mechanizmach pożądanej zmiany społeczno-gospodarczej. Okazały się prostym, biurokratycznym ćwiczeniem z opisu wydawania funduszy. Nie można ich niestety potraktować jako źródła inspiracji dla Polski. Pocieszającym może być, jedynie obserwacja, że na tym - słabym tle polskie badania wypadały relatywnie dobrze. Nie zmienia to jednak faktu, że inspiracji trzeba szukać gdzie indziej.

dyskutowane są za to często w kwartalniku *Regional Studies*. Niemniej samej ewaluacji Funduszy Strukturalnych *sensu stricte* poświęcony był jak dotąd jeden numer specjalny (2006 40[20]). Patrząc na tematykę wystąpień ostatniej (Lizona, X 2008) konferencji największego branżowego stowarzyszenia (European Evaluation Scoeity) tematyka funduszy strukturalnych jest na marginesie bieżących debat.

2 POSZUKIWANIE INSPIRACJI POZA POLEM FUNDUSZY STRUKTURALNYCH

Obserwacje z przeprowadzonego w poprzednim rozdziale przeglądu praktyk ewaluacyjnych w polu funduszy strukturalnych prowadzą do wniosku, iż jest to bardzo ograniczone źródło inspiracji dla Polski. Gdzie w takim razie poszukiwać inspiracji i czego miałyby one dotyczyć?

Racjonalnym polem poszukiwania inspiracji wydają się bogate praktyki amerykańskie. USA nie mają co prawda odpowiednika europejskich funduszy strukturalnych - szerokiej interwencji skierowanej na ponadsektorową politykę regionalną. Niemniej praktyce tamtejszego sektora publicznego nie są obce wieloletnie, multisektorowe, posiadające złożoną strukturę celów programy interwencji publicznych. Oczywiście ta odmienność wymaga ostrożności, testowania i adaptowania rozwiązań na polskim gruncie.

Analizując słabości dotychczasowych praktyk ewaluacji programów regionalnych funduszy strukturalnych trzeba stwierdzić, że kluczowym wydają się trzy wyzwania.

Pierwszym jest znalezienie podejść, czy narzędzi ułatwiających konstruowanie spójnych, naukowo wiarygodnych podstaw metodologicznych, rozumianych całościowo - jako koncepcja badawcza oraz zestaw metod i narzędzi.

Drugim wyzwaniem jest znalezienie sposobów powiązania analiz efektów konkretnych programów UE z szerszymi teoriami i praktyką rozwoju. Tym samym chodzi o wyjście poza prosty opis produktów i mechaniczną ocenę celów programu, na rzecz analizy relacji przyczynowych i prób wpisania ich w szerszy kontekst mechanizmów rozwoju społeczno-gospodarczego.

Wreszcie trzecim polem poszukiwań, są sposoby analizy i oceny wpływu na kapitał relacyjny uczestników programów UE. Patrząc na strukturę analizowanej w poprzednim rozdziale populacji raportów widać wyraźnie silną grupę ewaluacji programów współpracy terytorialnej oraz programów wspierania innowacyjności. W obydwu tych przypadkach jednym z głównych oczekiwanych efektów jest wzmocnienie współpracy i relacji między uczestnikami programów - postrzegane jako "miękkie" i trudno mierzalne.

Opierając się na przekrojowym przeglądzie literatury światowej i wywiadach z amerykańskimi ekspertami, jako dwie główne dwie inspiracje zidentyfikowaliśmy

"ewaluacje wspierane teorią" (wraz z modelami logicznymi) oraz analizę sieci. Naszym zdaniem - co staramy się dowieść w kolejnych podrozdziałach - mogą one pomóc w sprostaniu zarysowanym wyzwaniom. Obydwie zaprezentowane poniżej inspiracje mają także ważną cechę - ujmują kwestie analizy efektów w sposób całościowy, wprowadzają systemowe, ustrukturalizowane myślenie oparte na mocnych podstawach badawczych. Tym samym mogą w znaczący sposób przyczynić się do pogłębienia i poszerzenia analiz ewaluacyjnych, wzbogacenia ich o relację z teorią, przy jednoczesnym zakorzenieniu w wiarygodnej empirii.

Przed przejściem do ich prezentacji zdecydowaliśmy się jednak na przedstawienie szerszego kontekstu - wciąż aktualnego konfliktu paradygmatów. Wyznacza on ramy dyskursu na temat wiarygodności dowodów z badań ewaluacyjnych i sposobów oceny programów publicznych.

2.1 Kontekst - wojna paradygmatów

Ocena efektów programów publicznych stała się w latach 1980-tych areną wielkiej debaty metodologiczno-filozoficznej, starcia między dwoma paradygmatami naukowymi: klasycznym pozytywizmem - opartym na założeniu, że obiektywną wiedzę można uzyskać poprzez obserwację, a konstruktywizmem - zakładającym brak obiektywnej wiedzy a jedynie możliwość poznawania odmiennych punktów widzenia, wartości i interakcji (Tavistock Institute et al., 2003).

"Wojna" wynikała z dwóch, biegunowo odmiennych wizji filozofii definiowania, postrzegania i poznania świata. Podsumowano je w poniższej tabeli.

Tabela 11. Pozytywizm vs. konstruktywizm

	Pozytywizm	Konstruktywizm
Ontologia	Rzeczywistość jest dana i jest bytem zewnętrznym	Rzeczywistość tworzą aktorzy społeczni
Epistemologia	Potrafię się zdystansować – analizować obiektywnie i chłodno	To kim jestem warunkuje to w jaki sposób postrzegam świat
Aksjologia	Mogę oceniać obiektywnie, dystansując się od własnych wartości	Moje oceny są warunkowane moimi przekonaniami
Generalizacja i podejście do analizy	Podstawa to dedukcja, mogę generalizować	Podstawa to indukcja, nie można generalizować
Podejście logiczne	Jestem w stanie określić co jest przyczyną a co skutkiem	Nie mogę odróżnić przyczyny od skutku, mogę tylko określić relacje elementów

Źródło: (Christie, 2008)

Te dwa odmienne paradygmaty mają bardzo praktyczne i silne przełożenie na orientację i sposób prowadzenia badań ewaluacyjnych. Mianowicie warunkują kwestie centralne dla ewaluacji - to jak definiujemy JAKOŚĆ PROGRAMU oraz WIARYGODNE DOWODY skuteczności jego działań - efektów.

Dla pozytywistów jakość programu jest definiowana według obiektywnych kryteriów które program spełnił, a które zweryfikowano w procesie pomiaru (ang. *Quality as measured*). Wiarygodnymi dowodami są te, które osiągnięto w procesie badawczym analogicznym dla badań nauk przyrodniczych - powtarzalnym, opartym na podejściu ilościowym oraz eksperymentalnym planie badawczym (a jeśli jest to niemożliwe - quasi-eksperymentalnym). Tzw. "złotym standardem" są losowe grupy kontrolne (ang. *randomised control trials* - RCT).

Z kolei dla konstruktywistów jakość programu jest definiowana odczuciami i postrzeganiem jego interesariuszy (ang *Quality as experienced*). Wiarygodnymi dowodami sukcesu są więc jakościowe strategie badawcze, badania obejmujące jak największą liczbę opinii jak najszerszej grupy aktorów, ich różne punkty widzenia. Sam ewaluator jest w tym wypadku nawet nie tyle oceniającym co "dyrygentem" zmiany społecznej.

W latach 1990-tych "wojna paradygmatów" przycichła w literaturze tematu, by w ostatnich latach powrócić jednak ze zdwojoną siłą i to w polu praktyki ewaluacyjnej. Burzę wywołała inicjatywa Departamentu Edukacji USA ustanawiająca - jako standard naukowych metod ewaluacji programów edukacyjnych - kontrolne grupy losowe (RCT). Administracja Busha uznała to podejście za "złoty standard" dla wszystkich pól tematycznych programów federalnych. Wkrótce zaś zaczęła promować w ramach PART (Program Assessment Rating Tool) - punktując wysoko te programy publiczne, których efekty mierzone były quasi-eksperymentami, zaś obniżając punktację programom z ewaluacjami opartymi na jakościowych strategiach badawczych. Wywołało to gorącą dyskusję na forum American Evaluation Association. Konflikt dotarł także do Europy i miał swój wyraz w liście otwartym European Evaluation Society protestującym przeciw narzucaniu jednego standardu wszystkim rodzajom badań ewaluacyjnych (2007).

Próbą praktycznego rozwiązania tego palącego konfliktu było podejście realistyczne (ang. *realist approach*) (Henry et al., 1998; Pawson, Tilley, 1997) i zbliżona do niego "ewaluacja zorientowana na użycie" (Patton, 1997; 2008). Wydaje się, że tą ścieżką podążają również przedstawiciele *"theory driven evaluation"* - Chen, Donaldson czy Rossi. Jakość programu w tym wypadku jest definiowana na podstawie rzeczywiście osiągniętych efektów zakładanej zmiany społecznej ale z uwzględnieniem wyjaśnienia jej mechanizmu. Wiarygodne dowody sukcesu to te, które zostały zbudowane na przejrzystym ciągu logicznym i poparte szerszymi teoriami, innymi badaniami oraz - na ile to możliwe - badaniami własnymi.

Przy konfrontacji paradygmatów "Realizm" podejmuje jednak wyraźną krytykę pozytywizmu. Zawraca uwagę, iż pozytywizm nie może być stosowany w analizach systemów społecznych, bo w odróżnieniu od nauk przyrodniczych nie można kontrolować wszystkich czynników, ani też powtórzyć eksperymentu w takich samych warunkach; a także iż metoda quasi-eksperymentów nie wyjaśnia dlaczego pojawiły się, lub nie, efekty (a informacja o przyczynach i mechanizmie sukcesu jest równie ważna dla polityków i opinii publicznej jak sama ocena programu) (Pawson, 2006). Z drugiej strony realiści dowodzą, iż konstruktywizm relatywizując i znosząc osąd staje się bezużyteczny przy ocenie efektów, gdyż nie dostarcza decydentom informacji potrzebnych do planowania i kierowania programami publicznymi.

Najnowszym głosem w tej dyskusji jest książka pod redakcją Donaldsona, Christie & Mark'a (2008). Zbiera ona i waży poglądy oraz argumenty przedstawicieli różnych paradygmatów. W konkluzji wydaje się jednak zbliżać do podejścia realistycznego. Donaldson (2008, s.248) stwierdza, że to co uważamy za "wiarygodne dowody" zależy od 5 kwestii: pytania będącego przedmiotem zainteresowania zlecających, kontekstu, teorii ewaluacji która została przyjęta jako kompas dla badania oraz ograniczeń praktycznych (czasu i zasobów).

Dyskusje z lat 1980-tych na temat paradygmatów praktycznie ominęły pole funduszy strukturalnych, gdyż unijna ewaluacja była wtedy w stadium embrionalnym. W 1 edycji poradnika Evalsed (Tavistock Institute et al., 2003) zauważano, że ewaluacja europejska czerpie ze wszystkich trzech paradygmatów (pozytywizmu, konstruktywizmu i realizmu). Nie odniesiono jednak tej obserwacji do pola funduszy strukturalnych. Dopiero Bachtler i Wren (2006, s.146-148) zidentyfikowali ślady trzech paradygmatów w różnych typach badań ewaluacyjnych, wiążąc je z funkcjami jakie spełniały te badania w systemie europejskim. Należy jednak stwierdzić, że wpływ tych paradygmatów był w dużej mierze nieuświadomiony zaś debaty metodologiczne przez lata koncentrowały się na technicznych kwestiach (założenia danego modelu, reprezentatywność ankiet, poprawność danych), a nie fundamentach filozoficznych.

Podejmując obecnie chęć pogłębienia poważnej refleksji nad efektami funduszy strukturalnych trzeba być świadomym napięć które ze sobą niosą te trzy paradygmaty, a także ich zalet i ograniczeń. Mają one bezpośredni, bardzo praktyczny wpływ na metodologię badawczą, a co za tym idzie, na wiarygodność wniosków z badań i ostateczną ocenę efektów programów. Prezentowane w dalszej części inspiracje dopuszczają pewien eklektyzm w doborze paradygmatów. Niemniej bliżej im do realizmu, choć np. "ewaluacje wspierane teorią" mogą sięgać po quasi-eksperymentalne plany badawcze, oparte przecież na paradygmacie pozytywistycznym.

2.2 "Ewaluacja wspierana teorią" i modele logiczne - pomoc w definiowaniu zmiany

2.2.1 Geneza i definicja "ewaluacji wspieranej teorią"

"Theory-driven evaluation" (TDE), nazywana również *"Theory-based evaluation"*, pojawiła się w połowie lat 1980-tych jako reakcja na rosnące rozczarowanie ewaluacją zawężoną do biurokratycznej analizy nakładów - produktów (Chen, Rossi, 1983; Bickman, 1987). Główny orędownik i propagator tego podejścia - Huey-Tsych Chen - w swojej przełomowej książce *"Theory-Driven Evaluations"* zauważył, iż ówczesne ewaluacje powszechnie pomijały odwołania do jakichkolwiek teorii, czy szerszych koncepcji przy formułowaniu elementów programu, jego racjonalności i relacji przyczynowych (Chen, 1994, s.17-18).[11]

Według Chen'a taka dominacja ateoretycznej, wąskiej praktyki ogranicza użyteczność ewaluacji. Proste zestawienia nakładów-produktów nic nie mówią o mechanizmach przyczynowych, zupełnie ignorują efekty uboczne - wychodzące poza oficjalne cele, jak i kontekst organizacyjno-polityczny. Co więcej, podejmowanie decyzji politycznych na podstawie ich wyników niesie ze sobą zwiększone niebezpieczeństwo tzw. błędów pierwszego i drugiego typu (type I & II error)[12], gdyż takie badania nie wyjaśniają "dlaczego" nastąpiła porażka - czy zawiniła koncepcja programu, czy może mało skuteczne wdrażanie, czy kierunek programu był poprawny ale skala zbyt mała, czy też o wszystkim zadecydował kontekst.

Jako alternatywę Chen zaproponował więc podejście, które strukturalizuje całościowe myślenie o programie (w tym pytania ewaluacyjne i plan badawczy) w formie swoistej teorii interwencji publicznej. Należy podkreślić, że TDE nie jest

[11] Na poparcie swojej tezy Chen przywołuje interesujące badania próby 175 ewaluacji publikowanych w różnych periodykach (Lipsey et al., 1985). Wnioski są zbieżne z wynikami pierwszej części niniejszego raportu. Amerykańskie ewaluacje z lat 1980-tych, podobnie jak obecne ewaluacje funduszy strukturalnych są ateoretyczne i pozbawione głębszej struktury analitycznej.

[12] Terminy te zapożyczone zostały ze statystyki. Type I error to sytuacja w której, ewaluator uznaje, że program przyniósł efekty, gdy w rzeczywistości zmiana była wynikiem wyłącznie np. procesów kontekstowych (sam program nie miał żadnego pozytywnego wpływu). Type II error to sytuacja w której badacze uznają, że program nie miał wpływu podczas, gdy wpływ się pojawił ale był np. przesunięty w czasie, w przestrzeni (Daponte, 2008). Możemy sobie łatwo wyobrazić, że decyzje strategiczne co do kontynuacji lub zamknięcia programu podejmowane na podstawie takich przesłanek będą w obydwu przypadkach społecznie szkodliwe. Na przykład w przypadku 1 marnowane byłyby środki publiczne na program, który nie działa. W przypadku 2 zamknięto by program, który działał i przynosił efekty.

metodą lecz czymś szerszym - sposobem myślenia o ewaluacji (Frechtling, 2007, s.5). Wartością tego podejścia jest to, że zasadniczo jest ono neutralne metodologicznie (Donaldson, 2007, s.11). Tym samym pozostaje poza filozoficznymi sporami - opisywaną w poprzednim podrozdziale wojną pozytywistów z konstruktywistów. Może stanowić organizacyjną podstawę zarówno dla badań ilościowych, qasi-eksperymentalnych, jak i metod jakościowych, czy nawet mieszanych planów badawczych (tzw. *mix-method desings*).

Dzięki tej otwartości TDE rozwija się dynamicznie od połowy lat 1990-tych, stając się obecnie jednym z głównych podejść do ewaluacji. Znajduje też wielu kontynuatorów, którzy udoskonalają to podejście oraz wzbogacają je o nowe interpretacje specyficzne dla pól tematycznych (porównaj: wydanie specjalne *New Directions for Evaluation:* Petrosino et al., 2000). Prócz takich autorów jak Chen (2004) i Donaldson (2007), do TDE odwołują się również amerykańscy klasycy ewaluacji - C.H. Weiss (1997), M.Q.Patton (2008), P.H.Rossi (Rossi et al., 1999), jak i europejscy promotorzy ewaluacji realistycznej (Pawson, 2002).

Theory-driven evaluation to *"całościowa lub kontekstowa ocena programu oparta koncepcyjnie na teorii interwencji"* (Chen, 2005b, s.415). Jej celem jest "*nie tylko informowanie o realizacji i wartości programu ale również o tym w jaki sposób i dlaczego program osiągnął dane rezultaty"*. Ujmując to nieco inaczej - "ewaluacja wspierana teorią" pokazuje nie tylko jakie są rzeczywiste efekty programu ale wyjaśnia również "dlaczego". TDE jest więc przydatna zarówno do oceny efektów (funkcja rozliczeniowa - wzmacniania odpowiedzialności) jak i budowania szerszej wiedzy o mechanizmach interwencji publicznych (poznawcza funkcja ewaluacji).

Jak wskazuje sama nazwa - podejście to przykłada szczególne znaczenie do roli "teorii" w badaniu ewaluacyjnym. Paradoksalnie jednak - jak zauważa Weiss (Weiss, 2004) - ewaluatorzy nie wypracowali dotąd ujednoliconego rozumienia "teorii programu". Patton (2008, s.339) wskazuje, że to czym nazywana jest teoria programu zależy zarówno od kontekstu badania, skali danej interwencji jak i tradycji naukowej, z której pochodzi dany autor.

Podejmując próbę uporządkowania aparatu pojęciowego można zakłożyć, że każdy program składa się z trzech rodzajów teorii, które należy zidentyfikować, opisać a następnie ocenić (Chen, 2005; Donaldson, 2007; Pawson, 2009). Są to:

1. **Teorie bazowe** - a więc wiedza, będące inspiracjami dla programu szersze teorie społeczne, wcześniejsze doświadczenia, wyobrażenia, intuicyjne przeświadczenia osób lub grup interesariuszy, które tworzyły program;

2. **Teoria zmiany** (ang. *theory of change*) - nazywana także teorią koncepcyjną (Chen, 2004, s.257), lub - chyba najtrafniej - "teorią oddziaływania programu" (Donaldson, 2007, s.25) - to przyjęte przez twórców danego programu konkretne założenie, co do przyczyn które mają doprowadzić do osiągnięcia pożądanego celu (rozwiązania danego problemu społecznego); jest to określenie mechanizmu przyczynowo-skutkowego, najczęściej zapisywane w formie szeregu stwierdzeń "jeśli... to..";
1. **Teoria wdrażania** (ang. *process theory*) - nazywana też teorią wdrażania programu (Donaldson, 2007) - to strategia zamiany nakładów na oczekiwane produkty, wskazanie jakie konkretne działania i ich sekwencja są potrzebne, by to osiągnąć (Chen, 2005a).

Przykład trzech rodzajów teorii składającej się na "teorię programu" przedstawiono w poniższej ramce.

Tabela 12. Teorie programu - przykład Narodowego Planu Rozwoju Irlandii 1994-1999

Palącym problemem Irlandii na początku lat 1990-tych był wyjątkowo niski poziom rozwoju gospodarczego. Podczas tworzenia Narodowego Planu Rozwoju na lata 1989-93 (NPR), główną *inspiracją teoretyczną* **(teoria bazowa)** były teorie nowej ekonomii mówiące o kluczowej roli czynnika endogennego - kapitału społecznego i zasobów ludzkich w trwałym rozwoju współczesnych gospodarek.

Kierując się tą teorią społeczno-ekonomiczną, Irlandczycy sformułowali następującą ***teorię zmiany*** dla swojego NPR: *"jeśli będziemy inwestować w podniesienie kapitału ludzkiego to stworzymy trwałe podstawy dla rozwoju i nasza gospodarka zacznie się rozwijać"*.

To założenie przetłumaczono na konkretną ***teorię działania*** obejmującą działania (szkolenia, inwestycje w infrastrukturę edukacyjną, wzmacnianie aktywności zawodowej kobiet, etc.), procedury oraz system instytucji wdrażających Program Operacyjny Rozwoju Zasobów Ludzkich.

Tak więc na teorię irlandzkiego programu Rozwoju Zasobów Ludzkich złożyły się więc trzy elementy: inspiracje teoretyczne, konkretna teoria zmiany oraz teoria działania.

Źródło: opracowanie własne

Zgodnie z założeniami "ewaluacji wspieranej teorią" każdy program jest traktowany jako hipoteza (*jeśli... to...*), jako praktyczny test dla szerszej teorii rozwoju społeczno-gospodarczego. Całą sekwencję "teorii programu" można zapisać następującym stwierdzeniem:

Opierając się na własnym doświadczeniu oraz ustaleniach społeczno-ekonomicznej teorii Z [teoria bazowa] jeśli zrealizujemy X to dzięki temu osiągniemy Y [teoria zmiany]. Najlepszym sposobem przejścia od X do Y będzie sekwencja działań A, B i E [teoria działania].

Zastosowanie tego "szablonu" przedstawiono w poniższej tabeli, na przykładzie projektu Regionalnej Strategii Innowacji.

Tabela 13. Teorie programu - przykład Narodowego Planu Rozwoju Irlandii 1994-1999

Problemem w polskim regionie X jest jego niska konkurencyjność wynikająca - jak ustalono w diagnozie - z niskiej innowacyjności lokalnych firm. Opierając się na wcześniejszych doświadczeniach greckich i włoskich regionów biorących udział w unijnym projektach Regionalnych Systemów Innowacji jak i na teoriach gron oraz RSI (Cooke, 2004; Porter, 1998) [szeroka teoria] uznano, że zbudowanie RSI w tym regionie podniesie innowacyjność firm, co z kolei wpłynie pozytywnie na konkurencyjność regionu [teoria zmiany]. Zdecydowano, że stworzenie skutecznego RSI wymaga następującej sekwencji działań [teoria działania]: 1.analizy potencjału innowacyjnego przeprowadzonej przez wszystkich głównych interesariuszy w regionie, 2.wypracowania wspólnej strategii i określenia sektorów priorytetowych 3.zorganizowania platform i zasad komunikacji między kluczowymi uczestnikami systemu tj. firmami - władzami regionalnymi - instytucjami naukowymi.

Źródło: opracowanie własne

2.2.2 Geneza i definicje modeli logicznych

Podczas gdy TDE wyrosła z potrzeb świata nauki jako potrzeba rozszerzenia analiz ewaluacyjnych, modele logiczne przyszły ze świata praktyków jako potrzeba ustrukturalizowania coraz bardziej kompleksowych interwencji publicznych, tak by można było komunikować się w sposób zrozumiały z ich interesariuszami. Obydwa te nurty spotkały się ze sobą.

Modele logiczne to po prostu ilustracja w jaki sposób działa dany program (Kellogg Foundation, 2004). Knowlton i Philips (2008, s.5) stwierdzają, że są to *"narzędzia opisujące relacje pomiędzy elementami niezbędnymi we wdrażaniu programu. Modele logiczne określają zarówno to co jest planowane (co robimy) i to co oczekiwane (co uzyskamy). Są mapą drogową prowadzącą do określonego punktu (efektu)"*.

Jako początki modeli logicznych wskazuje się pracę Suchmana z lat 1960-tych, w której rekomendował przedstawianie interwencji jako łańcucha celów - bezpośrednich, pośrednich i ostatecznych - którym przyporządkowane byłyby

konkretne, następujące po sobie działania (Patton, 2008, s.340; Rogers, 2005, s.232).

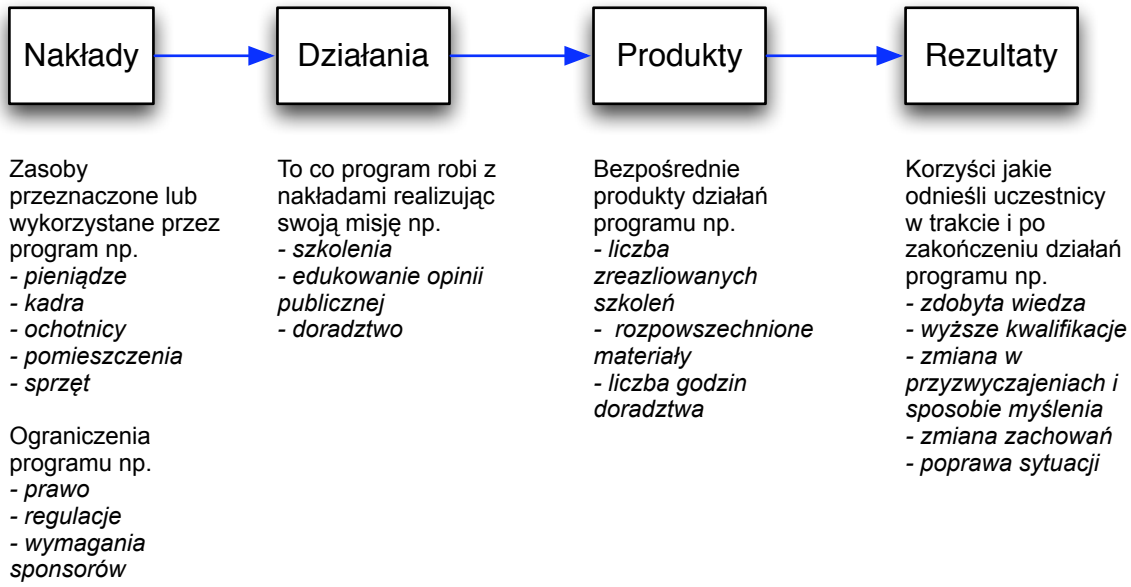
W pierwszych modelach pokazywano prostą drogę od nakładów do efektów (nakłady - działania - produkty - efekty) (United Way, 1996) (ryc. 17). Jednak w najbardziej popularnej interpretacji - zaprezentowanej przez Fundację Kellogg'a (ryc. 18) modele logiczne miały już wydzielone dwa rodzaje efektów, a także wyraźnie wyodrębnione dwa bloki - wdrożeniowy (*co robimy? - what?*) i koncepcyjny (*dlaczego coś robimy? - why?*). Bloki te odpowiadają podziałowi na "teorię działania" i "teorię zmiany" w TDE.

Ciąg nakładów - efektów pozostał standardem modeli logicznych. Zdarzają się jego modyfikacje związane z uszczegóławianiem rodzajów efektów (wyróżniane są efekty krótkoterminowe, średnioterminowe oraz długoterminowe), pokazywaniem związku wynikowego między problemami a strategią (porównaj schemat logiki interwencji UE), czy zaznaczaniem miejsca klienta-odbiorcy programu w łańcuchu logicznym (produkty - klienci - króto i średniookresowe rezultaty - oddziaływanie - (Mc Laughlin, Jordan, 2004, s.9).

Interpretacja Komisji Europejskiej stosowana do interwencji UE, spłaszczyła wyróżnienie między "teorią zmiany" (oddziaływania) a "teorią działania". Tymczasem najnowsza literatura światowa wiąże modele logiczne ściślej z TDE i rozgranicza je na dwa typy: modele oddziaływania - ilustrujące "teorię zmiany", oraz modele wdrażania - ilustrujące "teorię działania" programu (Bamberger et al., 2006, s.175-179; Knowlton, Phillips, 2008). Omawiamy to w kolejnym punkcie.

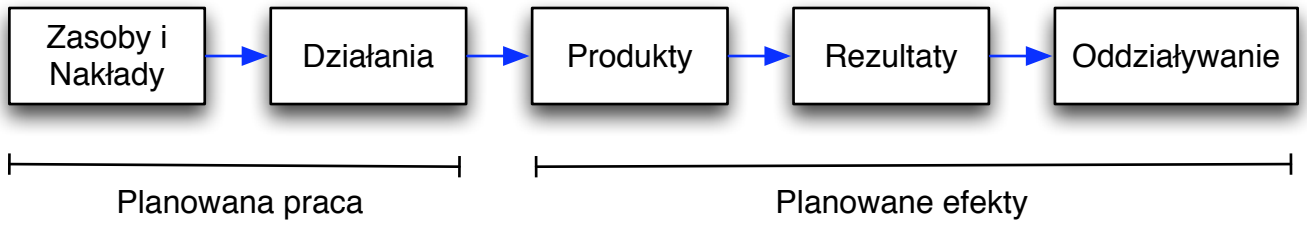
Rycina 17. Tradycyjny model logiczny

Źródło: United Way of America (1996)



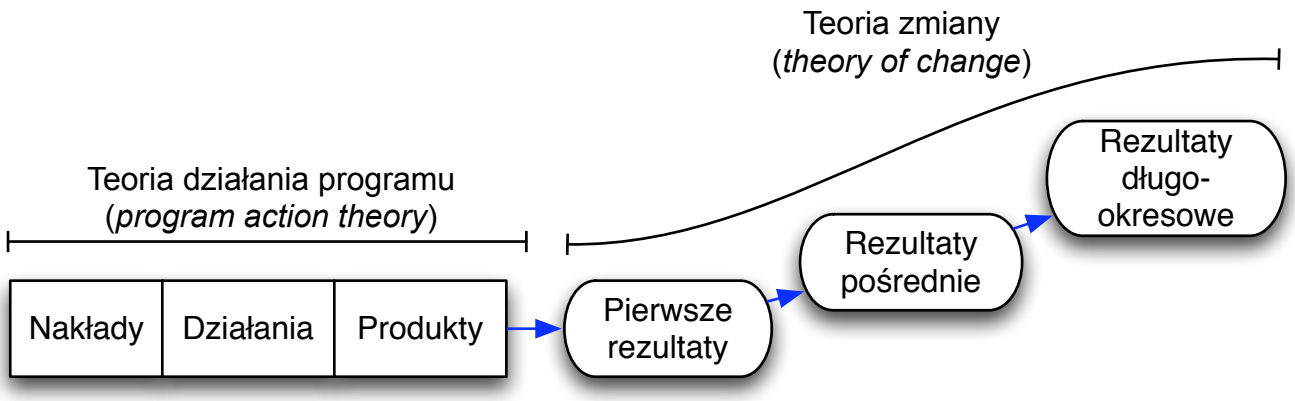
Rycina 18. Model logiczny Fundacji Kellogg'a

Źródło: Kellogg Foundation (2004)



Rycina 19. Teoria programu a model logiczny

Źródło: Donaldson (2007) s.25



2.2.3 Przykłady zastosowań

Zanim przejdziemy do zaprezentowania przykładów zastosowań *Theory Driven Evaluation* (TDE) i modeli logicznych należy raz jeszcze wyjaśnić praktyczną relację między nimi. Otóż TDE jest podejściem ewaluacyjnym (ang. *evaluation approach*) - a więc szerszą filozofią myślenia o badaniu i strategią jego strukturalizowania, podczas gdy modele logiczne są technicznym narzędziem jej obrazowania.

Współcześnie teorie programu i modele logiczne są ściśle ze sobą zrośnięte. Modele logiczne w sposób graficzny mogą prezentować "szersze teorie" i "teorię zmiany" będące rdzeniem koncepcyjnym danej interwencji (polityki, programu, czy projektu) (Frechtling, 2007, s.1), jak i samą "teorię działania" - sekwencję działań oraz ich produkty - porządkującą kwestie operacyjne. Tę zależność można przedstawić jak na rycinie 19.

W praktyce tworzona teoria i budowane modele logiczne nie mają jednego standardowego schematu - ich kształt i stopień szczegółowości zależy celu badania, pytań badawczych, skali programu. Przykład takiego zastosowania w odniesieniu do jednego z programów amerykańskich przedstawiono na ryc. 20 zaś wstępem opatrzono w poniższej tabeli.

Tabela 14. Przykład teorii Programu Zarządzania Energią Odnawianą

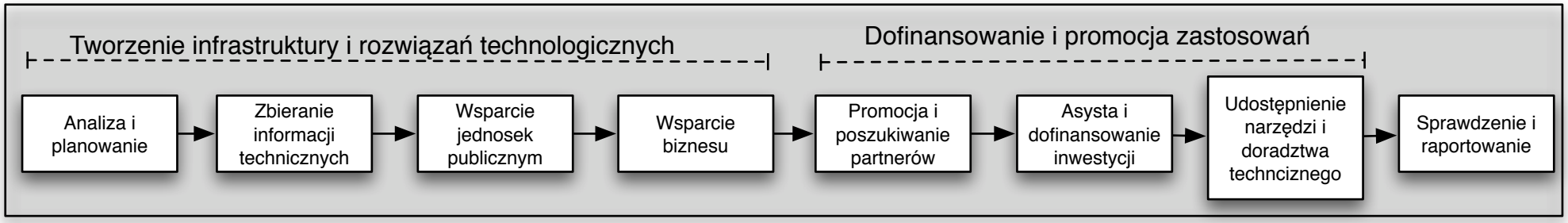
Amerykański Departament Energii, Biuro Wydajności Energetycznej i Energii Odnawianej (EERE) jest instytucją federalną i zajmuje się badaniami własnymi z zakresu "czystych" technologii energetycznych, koordynacją szeregu programów jak i promocją energooszczędnych rozwiązań technologicznych oraz usług. EERE ma swoje biura w całych Stanach Zjednoczonych - większość programów wdrażana jest na poziomie regionalnym. Analizowana interwencja (złożona z kilku pakietów działań) była nakierowana na dyfuzję innowacji i technologii związanych z energooszczędnością. Ewaluatorzy mieli zidentyfikować mechanizm przenoszenia się innowacji i adaptowania nowych praktyk w każdej z grup docelowych. Prócz prezentowanego poniżej modelu stworzono szczegółowe modele dla każdego z typu działań. Jako szerszą podstawę teoretyczną zaadaptowano teorię dyfuzji zmian społecznych E.Rogers'a. Główny model czyta się od góry ("co robimy") do dołu ("w jakim celu - co chcemy osiągnąć").

Źródło: (Knowlton, Phillips, 2008, s.110-112)

Rycina 20. Przykład teorii Programu Zarządzania Energią Odnawialną

Źródło: Knowlton & Phillips (2008)

BIURO WYDAJNOŚCI ENERGETYCZNEJ I ENERGII ODNAWIALNEJ U.S.A., WYDZIAŁ ENERGII, ZDECYDOWAŁO O ZAINWESTOWANIU W NASTĘPUJĄCE DZIAŁANIA...



Znając te działania i odpowiadające im produkty....

Opisujemy i mierzymy oczekiwaną reakcję następujacych grup docelowych:

PO TO BY POSZCZEGÓLNE GRUPY DOCELOWE...

Pracownicy naukowi	Jednostki publiczne	Jednostki produkcyjne i biznesowe	Użytkownicy indywidualni
tworzyli i rozwijali wiedzę techniczną jak i rynkowa dzięki której rozwiązania energooszczędne będą bardziej dostępne i łatwiejsze do wdrażania	zmieniały polityki publiczne, swoją strukturę oraz strategie działania, w celu usprawnienia wdrażania rozwiązań energooszczędnych	tworzyły i unowocześniały produkty, tworzyły i dostosowywały kanały dystrybucji, rozwijały energooszczędne instalacje oraz infrastrukturę wspomagającą	adoptowali, powielali i przyzwyczajali się do energooszczędnych rozwiązań jak i świadomego oszczędnego wykorzystywania źródeł energii

i pokazujemy jak te działania łączą się z oddziaływaniem

DZIĘKI CZEMU...

nastąpi redukcja zużycia energii i emisji gazów, podniesienie produktywności i globalnego bezpieczeństwa, tworzenie zrównoważonych mikrośrodowisk

Opierając się na przeglądzie literatury i case'ów amerykańskich można stwierdzić, że "ewaluacja wspierana teorią" (ang. *theory driven evaluation*) składa się z trzech etapów (Donaldson, 2007, s.20-49; Chen, 2004):

Etap 1: Zbudowanie "teorii programu" i przedstawienie jej w formie graficznej - zwykle w dwóch częściach: modelu teorii zmiany i modelu teorii działania;

Etap 2: Sformułowanie - w oparciu o stworzony model - pytań ewaluacyjnych oraz wyłonienie pytań priorytetowych;

Etap 3: Odpowiedź na pytania - a więc przeprowadzenie badania empirycznego (zebranie danych, analiza, konkluzje).

W całej procedurze kluczowym jest etap 1 - zbudowanie "teorii programu". Staje się on rdzeniem dalszej analizy - swoistą mapą drogową. Patton wyróżnia trzy podejścia do budowy teorii programu (2008, s.344-346):

- podejście dedukcyjne - budowane na teoriach pochodzących z przeglądu literatury naukowej;
- podejście indukcyjne - rozpoczynane badaniami empirycznymi, na podstawie których tworzona jest teoria programu;
- podejście zorientowane na użytkowników - budowane na podstawie dyskusji z użytkownikami programu i głównymi interesariuszami, identyfikujące i precyzujące ich założenia, którymi się kierują.

Zestawiając ich zalety i ograniczenia Patton wskazuje, iż podejście trzecie jest najbardziej efektywne. Również Donaldson (2007) - mimo iż kładący silniejszy nacisk na użycie szerszych teorii społecznych, akcentuje angażowanie interesariuszy.

W przypadku ewaluacji efektów - będącej przedmiotem niniejszego opracowania "budowanie teorii programu" nie polega na budowaniu od podstaw, lecz na rekonstrukcji i prześledzeniu zależności *post factum*. Ponadto już podczas rekonstruowania teorii programu pojawiają się pytania badawcze.

Dlatego też racjonalną wydaje się następująca procedura złożona z 7 kroków, wspierana modelem logicznym. Poniżej opisano krótko każdy z nich odniesieniem do modelu (patrz: rycina 21).

W **Kroku 1** należy jasno zdefiniować efekty programu jakich oczekują interesariusze - chodzi o kluczową zmianę społeczną jaką program miał wspierać. Patton (2008

s.336) zauważa, że w tym miejscu niezbędnym jest sprecyzowanie wszystkich pojęć użytych przy opisie efektów - szczególnie wyjaśnienie terminów wartościujących.[13] Wymaga to przeglądu dokumentów programu ale przede wszystkim dyskusji z głównymi interesariuszami.

Krok 2 wymaga nazwania strategii, która została użyta do dostarczenia (skutecznie lub nie) oczekiwanych efektów. Tu także podstawą jest przegląd dokumentów jak i konsultacje z interesariuszami, gdyż często "teoria zmiany" nie jest zapisana wprost w dokumentach programowych. Dzięki dwóm pierwszym krokom powstaje hipoteza programu "*jeśli... to...*".

W **kroku 3** natomiast ewaluatorzy starają się zidentyfikować założenia (teorie bazowe) leżące u podstaw ogólnej "teorii zmiany" programu. W tym celu muszą dokonać szerokiego przeglądu literatury, przeprowadzić wywiady z osobami, które tworzyły wizję programu. Wbrew pozorom etap ten jest bardzo czasochłonny gdyż zidentyfikowanie teorii czy inspiracji praktycznych wymaga szerokiej kwerendy bibliotecznej i pogłębionych studiów z danej dziedziny. W Polsce - z uwagi na ograniczony dostęp do literatury międzynarodowej - etap ten może być jeszcze większym wyzwaniem.

Krok 4 poświęcony jest analizie "teorii działania", a więc sposobom przekształcenia zasobów i nakładów w produkty. Analizowane są także rodzaje i skala osiągniętych produktów (w tym struktura i liczba beneficjentów). W zasadzie ta część opiera się w dużej mierze na prostej, opisowej analizie statystycznej danych z monitoringu programu. Warunkiem płynności tego etapu jakość danych z monitoringu. Jeśli dane na temat produktów wymagają uzupełnienia, wtedy badanie może się znacząco wydłużyć.

Krok 5 i 6 stawiają przed ewaluatorami największe wyzwanie. Mogą być też - w zależności od skali programu i stopnia rygoru metodologicznego - najbardziej czasochłonne. W tym miejscu ewaluatorzy starają się zrekonstruować a następnie przetestować teorię zmiany (ciąg przyczynowo-skutkowy) łączącą produkty z rezultatami (jakie produkty wywołały dane rezultaty) oraz rezultaty z oddziaływaniem (jaką szerszą, strukturalną zmianę wywołały rezultaty interwencji).

---

[13] Np. stwierdzenia typu: "zwiększenie potencjału innowacyjnego", "lepsze stanowiska pracy", "podniesienie szans rynkowych" wymagają uszczegółowienia: jaką oczekiwaną charakterystykę miałyby mieć "lepsze stanowiska pracy"? Jak definiowany jest potencjał innowacyjny i do jakiego poziomu miałby być podniesiony, by uznać go za zadowalający? Co wchodzi w zakres terminu "szanse rynkowe"? Czy tylko podstawowe umiejętności techniczne czy np. także pomoc psychologa i zbudowanie poczucia pewności siebie?

Analizując związek przyczynowo-skutkowy trzeba uwzględnić także czynniki kontekstowe - charakterystykę środowiska w jakim wdrażano interwencję (Bamberger et al., 2006, s.175-176) - klimat polityczny, gospodarczy, czynniki instytucjonalne mogą silnie zaburzyć relację przyczynowo-skutkową. Ponadto autorzy piszący o TDE zauważają, że relacja między produktem a rezultatem nie zawsze musi być bezpośrednia (szczególnie przy dużych programach). Wprowadzają więc w tym miejscu dwa ważne pojęcia: mediatorów (ang. *mediators*) oraz moderatorów (ang. *moderators*). Te pierwsze to czynniki wpływające na wdrożenie lub efekty interwencji, które mogą być jednak zmodyfikowane przez tę interwencję.[14] Te drugie to jakościowe (np. płeć, pochodzenie) lub ilościowe (poziom motywacji, doświadczenie) czynniki, które mają wpływ na kierunek i siłę relacji między przyczyną a skutkiem (Donaldson, 2007, s.27-32).

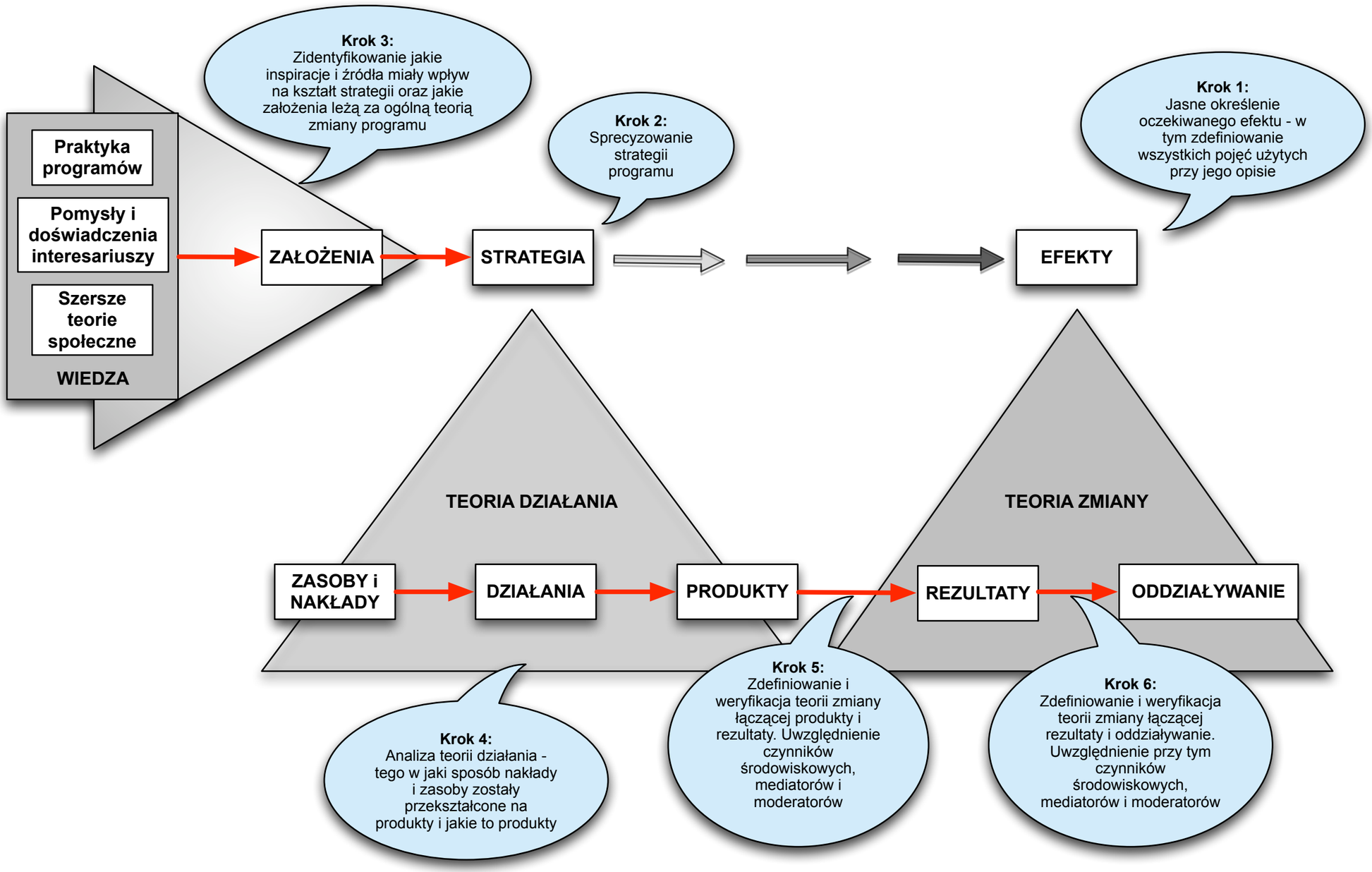
Do analizy i weryfikacji teorii zmiany Chen sugeruje metody ilościowe, z naciskiem - jeśli to możliwe - na quasi-eksperymenty (Chen, 2004). Oczywiście analiza przyczynowości wymaga oceny stopnia zmiany między sytuacją sprzed a po interwencji. Służyć temu może szereg standardowych porównawczych planów badań (szerzej patrz: Russ-Eft, Preskill, 2001, s.147-176; Reichardt, Mark, 2004). Wybór danego planu jest warunkowany skalą interwencji, dostępnością danych jak i poprzednimi badaniami danego programu. W tym miejscu warto jednak zaznaczyć - opierając się na wstępnych przeglądach doświadczeń z USA i UE, że grupy kontrolne znajdują zastosowanie głównie w polu zasobów ludzkich - gdzie beneficjentami ostatecznymi są osoby i wsparcia przedsiębiorczości - gdzie przedmiotem interwencji jest przedsiębiorstwo. Gminy, regiony są głównie tłem do analizy przestrzennej wyników badań jednostek. Racjonalna strategią wydaje się więc koncentrowanie na beneficjentach ostatecznych oraz analizie segmentów programu, by następnie połączyć je w modelu.

Wreszcie **krokiem 7**, ostatnim jest całościowa ocena efektów programu - podsumowanie skuteczności przejścia od strategii do efektów. Dokonywana jest na podstawie zebranych dowodów. W tym kroku wyjaśniany jest także mechanizm sukcesu lub niepowodzenia, a wraz z nim racjonalność teorii zmiany leżącej u podstaw programu.

[14] Bamberger (2006, s.176) podaje przykład programu mikrokredytów dla gospodyń domowych, na ich aktywizację zawodową i zakładanie przedsiębiorstw. Mediatorem jest tu postawa mężów. Jeśli będą otwarci na pracę żony poza domem i wspierający jej pomysł biznesu efekty programu będą silniejsze.

Rycina 21. Budowanie teorii programu krok po kroku

Źródło: zmodyfikowane schematy Knowlton & Phillips (2008) s.23-23 i 36, 45



2.2.4 Możliwości zastosowania w ewaluacjach rozwoju regionalnego

Wydaje się, że "ewaluacja wspierana teorią" może stać się cennym wsparciem metodologicznym ewaluacji efektów funduszy strukturalnych. Na użyteczność podejścia opartego na "teorii programu" zwrócił ostatnio uwagę w polskim kontekście Górniak (Górniak, 2007, s.12-13). W tym miejscu warto jednak uporządkować i uwypuklić argumenty, które przemawiają za zastosowaniem "ewaluacji wspieranej teorią" w polu programów rozwoju regionalnego wspieranych ze środków UE.

Po pierwsze, TDE dobrze artykułuje racjonalność programu - jego założenie na temat konkretnej zmiany społecznej (Weiss, 2004) - identyfikuje kluczowe pola i połączenia logiczne, jak i relacje zadaniowe programu poddawanego ewaluacji. W przypadku FS, programów o skomplikowanych strukturach celów, tworzonych przez kilka instytucji, taki zabieg ujęcia programu w formie hipotezy (*jeśli...to...*) ma szansę uporządkować myślenie o programie i jego funkcjach.

Po drugie, "ewaluacja wspierana teorią" strukturalizuje myślenie o programie i umożliwia stworzenie solidnej podstawy koncepcyjnej danego badania. Dzięki temu można:

- postawić jasny problem badawczy (weryfikacja programu - hipotezy)
- utrzymać dyscyplinę naukową podczas badania (TDE daje logiczną strukturę do zbierania i analizy danych)
- stworzyć jasną narrację raportu końcowego.

Tym samym wyniki badania są oparte na mocniejszych, spójnych i bardziej przejrzystych przesłankach, zaś ich prezentowanie nie ma formy opisu lecz jest spójną i logiczną opowieścią wyjaśniającą ciąg działań, efektów do których doprowadziły jak i czynników determinujących te efekty.

Po trzecie, co związane z pierwszym argumentem, TDE umożliwia zidentyfikowanie i wyartykułowanie teorii (nie zawsze uświadomionych), którymi kierowali się tworzący program interesariusze. Jak pokazują wyniki analiz rozdziału 1 niniejszej pracy, FS brakuje najwyraźniej osadzenia w szerszych teoriach społeczno-gospodarczych i bardziej naukowego podejścia. Mimo skali środków, jest to ćwiczenie jak dotąd w dużej mierze administracyjne. Nawiązanie do koncepcji i odnoszenie programów *explicite* do teorii i doświadczeń rozwoju regionalnego mogłoby podnieść zarówno jakoś badań jak i debaty publicznej opartej na ich wynikach (bardziej świadomej *evidence-based policies*).

Odnosząc się do tabeli 1 niniejszego opracowania (zakres ewaluacji efektów) można także stwierdzić, że TDE pomaga w eksploracji wszystkich trzech pól ewaluacji efektów - weryfikacji założonych celów, analizy rzeczywistych efektów jak i wyjaśnienia przyczyn osiągnięcia lub nie tychże.

Wreszcie, co kluczowe dla ewaluacji efektów - nawet w przypadku dużych programów, dla których użycie losowych grup kontrolnych jest niemożliwe - TDE pozwala do pewnego stopnia wypowiadać się na temat przyczynowości. Udowadniając, że program posuwał się według założonych sekwencji działań, zaś reakcja odbiorców była zgodna z założeniami, ewaluator może powiązać działania i produkty z efektami (Weiss, 2004).

Oczywiście ewaluacja wspierana teorią jak i modele logiczne niosą ze sobą także pewne ograniczenia i wyzwania.

Po pierwsze, mechaniczne sztywne stosowanie jednego modelu może prowadzić do zawężenia perspektywy badania, a co za tym idzie wyłączenia z analizy efektów ubocznych (nie przewidzianych w modelu). Bamberger (2006, s.173) uspokaja jednak, że teoria programu zwykle jest w stanie sprecyzować te efekty uboczne już na początku, zaś w trakcie badania uczulić ewaluatorów i menedżerów na ich ewentualne pojawianie się.

Po drugie w przypadku interwencji dużych, rozległych i mających kilka celów (a do takich niewątpliwie należą FS) pojawia się wyraźny wybór "coś za coś" pomiędzy uwzględnieniem wszystkich czynników, a utrzymaniem czytelności modelu. Co więcej w takich sytuacjach większość z narzędzi (planów badawczych) weryfikacji relacji przyczynowo-skutkowej ma bardzo ograniczone zastosowanie. Wydaje się, że można sobie radzić z tym problemem poprzez zawężanie perspektywy badania na pojedyncze elementy programu (ścieżki, działania, etc.). Jest to bardziej pracochłonne i kosztowne, bo każdy z takich segmentów jest oddzielnym badaniem - weryfikacją teorii danej części programu, ale z drugiej strony pozwala wciąż rozważać przyczynowość działań.

Wreszcie ostatnie wyzwanie przytacza Patton (2008, s.358) - możliwe zachwianie równowagi między rozważaniami teoretycznymi a praktyką może zmienić ewaluację w badania naukowe, oderwać od potrzeb konkretnych interesariuszy i ograniczyć jej utylitarny charakter. Co więcej, Scriven zauważa, że nadmierne teoretyzowanie może też spolaryzować i zniechęcić interesariuszy podważając ich własny model

teorii programu oparty nie na szerokich, wysublimowanych koncepcjach lecz zdroworozsądkowych, praktycznych przemyśleniach.

Pomimo wyszczególnionych powyżej ograniczeń "ewaluacji wspieranej teorią", należy stwierdzić, że może być ona użytecznym i przydatnym podejściem do ewaluacji efektów funduszy strukturalnych.

2.3 Analiza sieci - pomoc w ocenie dynamiki współzależności i relacji[15]

2.3.1 Geneza i definicja

Ewaluacja jako proces badawczy może i powinna wykorzystywać różne metody i techniki badawcze wypracowane na gruncie szeroko pojmowanych nauk społecznych. Jedną z nich jest analiza sieci, której poświęcony jest niniejszy podrozdział. Wykorzystywane w niej podejścia badawcze i techniki omawiane są w wielu publikacjach przede wszystkim w języku angielskim, ale również po polsku. Natomiast opracowań poświęconych zastosowaniu analizy sieci w ewaluacji jest relatywnie mało. Związane jest to zapewne z tym, że analiza sieci jest bardzo rzadko stosowana w praktyce ewaluacji. Nie znaczy to jednak, że jest mało przydatna. Wynika to raczej z tego, że analiza sieci zaczęła być wykorzystywana w ewaluacji stosunkowo niedawno czego symptomem może być to, że dopiero w 1998 r. pojawia się jako podejście metodologiczne w programie konferencji organizowanej przez American Evaluation Association (Durland, Fredericks, 2005, s.7). Ponadto: *„metod analizy sieci (…) jest bardzo dużo, ich wykorzystanie wymaga umiejętności korzystania ze specjalnych programów do analizy sieci i w konsekwencji prowadzenie badań sieciowych przez początkujących badaczy jest bardzo trudne”* (Batorski, 2008, s.188). Trzecim bardzo istotnym czynnikiem jest to, że analiza sieci nie jest metodą uniwersalną, tj. możliwą do zastosowania, w każdym badaniu ewaluacyjnym. Z drugiej strony w wielu przypadkach nie ma lepszej metody niż analiza sieci – tzn. zastosowanie innych metod da wyniki gorszej jakości.

[15] Rozdział autorstwa Adama Płoszaja - pracownika EUROREG - Uniwersytet Warszawski.

Analiza sieci (czy też sieciowa) ma długą tradycję, bogatą literaturę i szereg efektownych zastosowań (Freeman, 2004). Najobszerniejszy nurt badań sieciowych rozwijany jest w obrębie nauk społecznych jako analiza sieci społecznych (ang. Social Network Analysis: SNA – skrót ten często jest używany także w polskojęzycznych opracowaniach fachowych, por. np. (Swianiewicz et al., 2008). Jednakże nie sposób wyobrazić sobie współczesną analizę sieci bez wkładu nauk ścisłych, a przede wszystkim matematyki i teorii grafów (Wasserman, Faust, 2007, s.10-17).

Czym jest analiza sieci? Jeden z czołowych badaczy sieciowych pisze tak: *„Teoria sieci koncentruje się na relacjach między jednostkami, na wzorach interakcji. Rola charakteru jednostek jest marginalizowana, czy nawet pomijana, w nadziei na odkrycie głębszych praw. Teoretyk sieci patrzy na każdy system połączonych elementów i widzi abstrakcyjny wzór punktów połączonych liniami. To właśnie ten wzór ma znaczenie, architektura relacji, a nie tożsamość punktów. Wiele sieci niemających z sobą nic wspólnego, oglądanych z takiej wysokości, wygląda tak samo"* (Bendyk, 2004, s.257). Zatem analiza sieciowa to metodyczne badanie sieci. Z kolei *„terminem 'sieć' możemy określić każdy system złożony z elementów połączonych relacjami"* (Jonak, 2007).

Niniejszy podrozdział ma na celu pokazanie możliwości wykorzystania analizy sieciowej w praktyce ewaluacji. Z tego względu podstawy teoretyczne analizy sieci nie będą tu szczegółowo omawiane. Z jednej strony wymagałoby to dość długiego wywodu, z drugiej dostępnych jest wiele pozycji wprowadzających w tę tematykę. Większość tej literatury dostępna jest w języku angielskim, jednak są już również pozycje polskojęzyczne.

Tabela 15. Podstawowa literatura z zakresu analizy sieci

Ogólnym i jasno napisanym wprowadzeniem do analizy sieci społecznych jest artykuł Łukasza Jonaka (2007). Tekst Dominika Batorskiego (2008) jest bardziej szczegółowy oraz napisany na potrzeby odbiorców zainteresowanych ewaluacją. Edwin Bendyk (2004) badania nad sieciami przedstawia w szerszym kontekście „nowej nauki sieci". Z kolei Philip Ball (2008) w prosty sposób zapoznaje czytelnika z najnowszymi osiągnięciami szeroko pojętych badań sieciowych. W przypadku wydawnictw angielskojęzycznych podstawowym i wyczerpującym wprowadzeniem do analizy sieci społecznych jest ośmiuset stronicowe opracowanie Wassermana i Faust'a (2007). Warto również polecić dostępny on-line podręcznik Hannemana i Riddle'a *„Introduction to social network methods"* (www.faculty.ucr.edu/~hanneman/nettext)

.

2.3.2 Przykłady zastosowań

Analiza sieciowa może być zastosowana w badaniach ewaluacyjnych na różne sposoby w zależności od przedmiotu ewaluacji. Możemy analizować relacje między osobami oraz między organizacjami (instytucjami). Przedmiotem badań mogą być też powiązania między miastami, regionami, państwami lub innymi całościami które możemy wydzielić i analizować. Ponadto analiza sieciowa może być wykorzystana do ścisłego badania wszystkich fenomenów, które można skonceptualizować jako sieci. Mogą to być na przykład powiązania systemu celów strategii czy programów operacyjnych (Davis, 2005, 2007).

Przede wszystkim, co niejako oczywiste, analiza sieci jest przydatna w ewaluacji programów i projektów, których przedmiotem są *explicite* sieci - systemy połączone relacjami. Będą to wszelkiego rodzaju inicjatywy tworzenia i wspierania sieci instytucji (np. Krajowa Sieć Innowacji, czy sieć Regionalnych Ośrodków Europejskiego Funduszu Społecznego), ale także programy w których współpraca, relacje-interakcje, przepływy informacji stanowią *clu* działania (programy współpracy terytorialnej, regionalne systemy innowacji, grona - klastry).

Warto podkreślić, że można analizować sieć korzystając z innych metod niż analiza sieciowa. Takie podejście spotykane jest bardzo często (Rank, Williams, 1999) Szałaj, Ledzion, 2008). Nie oznacza to, że analiza sieci może być zastąpiona innymi metodami (np. standardowymi badaniami ilościowymi). Nie znaczy to również, że badanie sieci bez korzystania z analizy sieciowej jest niepoprawne metodologicznie. Takie podejście dostarczy nieco innej wiedzy o badanym przedmiocie. Nie pozwoli jednak na systematyczne scharakteryzowanie relacji zachodzących w danej sieci. Dlatego właśnie w takich przypadkach wykorzystanie analizy sieciowej jest szczególnie pożądane.

W kolejnych częściach niniejszego tekstu zostaną przedstawione przykłady zastosowań analizy sieciowej w badaniach ewaluacyjnych, a także w badaniach, które co prawda nie były ewaluacjami w ścisłym sensie lecz zastosowane w nich podejście może być bezpośrednio przeniesione do praktyki ewaluacji.

Sieć implementacji ZPORR w regionach (Swianiewicz et al., 2008)

Omawiane badanie jest jednym z nielicznych przypadków zastosowania analizy sieciowej do badania relacji między instytucjami w Polsce. Co prawda nie było ono

przeprowadzone jako ewaluacja, jednak jego charakter sprawia, że można je rozpatrywać jako element badania ewaluacyjnego. Przedmiotem badania była nieformalna sieć instytucji zaangażowanych w implementację Zintegrowanego Programu Rozwoju Regionalnego (ZPORR) w województwie małopolskim i dolnośląskim. W przypadku województwa małopolskiego uwzględniono 40 instytucji, a w przypadku dolnośląskiego 52. Informacje o powiązaniach między instytucjami pochodziły od przedstawicieli tych instytucji i były zbierane podczas wywiadów. Pytania zadawane poszczególnym aktorom dotyczyły częstości kontaktów w ogóle oraz kontaktów prywatnych (pozasłużbowych) pracowników danej instytucji z pracownikami innych instytucji zaangażowanych we wdrażanie ZPORR w regionie. Zebrane w ten sposób dane dotyczyły zatem deklarowanych powiązań między instytucjami z perspektywy przedstawicieli analizowanych instytucji. Analiza sieciowa była jedną z metod stosowanych w badaniu i z założenia służyła z jednej strony jako wzbogacenie badania, z drugiej jako narzędzie potwierdzające wyniki osiągnięte innymi metodami.

Wyniki badania pokazały że sieci w obu województwach są do siebie podobne. W obu regionach centralne miejsce w sieciach implementacji zajmowały instytucje formalnie odpowiedzialne za wdrażanie ZPORR: Urząd Marszałkowski i Urząd Wojewódzki. Jednakże autorzy badania wskazują na kilka istotnych różnic. Przede wszystkim dotyczy to tego, że respondenci z dolnośląskiego chętniej mówili o swoich kontaktach niż respondenci z małopolski, co przekłada się na mniejszą gęstość sieci w Małopolsce i większą na Dolnym Śląsku. Ponadto wyraźne są również różnice w miejscu stolicy województwa w sieci implementacji ZPORR. Urząd Miejski Wrocławia zajmuje centralną pozycję w sieci instytucji, natomiast magistrat krakowski pozycję raczej peryferyjną. W Małopolsce dość duże znaczenie miały organizacje pozarządowe (przede wszystkim NSZZ Solidarność), natomiast na Dolnym Śląsku ich znaczenie było znacznie mniejsze. Ponadto autorzy wymieniają jeszcze kilka innych cech różniących sieci implementacji ZPORR w badanych regionach.

Jak można powyższe analizy przełożyć na praktykę ewaluacji? Jednym z potencjalnie bardzo owocnych zastosowań może być wykorzystanie charakterystyk sieci implementacji jako zmiennych wyjaśniających przebieg i efekty ZPORR w regionach. Relacje między instytucjami zaangażowanymi we wdrażanie mogą w istotny sposób wpływać np. na poziom i tempo absorpcji oraz efektywność interwencji. Jeżeli w toku ewaluacji okazałoby się, że charakter powiązań w sieci wdrażania istotnie wpływa na przebieg interwencji to łatwo można na tej podstawie

sformułować rekomendacje dotyczące optymalnej struktury współpracy między instytucjami wdrażającymi.

Sieć instytucji zaangażowanych w realizację pilotażowego programu poprawy usług dla niepełnosprawnych (Fredericks, 2005)

Badanie zostało przeprowadzone w USA i dotyczyło przebiegu wdrażania 5-letniego programu personalizacji opieki nad niepełnosprawnymi. Celem programu była poprawa opieki (czy ogólnie usług dla niepełnosprawnych) przy jednocześnie bardziej efektywnym wykorzystaniu zasobów. Program wdrażany był w 6 instytucjach, zlokalizowanych w różnych miejscach (na wsi, w mieście, na terenach podmiejskich) i różniących się skalą działania (od 2000 do 10000 odbiorców końcowych). Zakres działalności poszczególnych instytucji jest duży - od stałej opieki przez dzienne pobyty po miękkie działania wspierające rodziny niepełnosprawnych, a także wyjazdy wakacyjne, opiekę zdrowotną, rehabilitację, kursy i szkolenia, itp. Z 6 instytucji implementujących program 4 zostały objęte ewaluacją. Informacje na temat powiązań między tymi organizacjami oraz z innymi instytucjami, z którymi współpracują zostały zebrane podczas wywiadów przeprowadzonych z dyrektorami poszczególnych instytucji. W sumie uzyskano sieć 22 instytucji zaangażowanych w implementację programu. Analiza sieci powiązań pozwoliła dopełnić i lepiej wyjaśnić wyniki badania osiągnięte za pomocą innych metod i źródeł.

Podstawowym wnioskiem płynącym z przeprowadzonych analiz było to, że kluczowi dla wdrażania programu aktorzy byli ze sobą mocno powiązani, co niewątpliwie wpływało na powodzenie programu. Jednakże okazało się, że intensywna komunikacja między instytucjami w znacznej mierze ograniczała się do przekazywania prostych, technicznych informacji (np. terminy spotkań). Nie zaobserwowano natomiast przepływu bardziej wartościowej wiedzy, tzw. „dobrych praktyk". Ewaluatorzy ustalili, że niektóre podmioty bardzo dobrze rozwiązywały pewne kwestie, które dla innych były dużym problemem. Mimo istniejących kanałów przepływu wiedzy owe „dobre praktyki" nie rozprzestrzeniały się w sieci instytucji w zgodny z oczekiwaniami sposób.

Omawiany przykład pokazuje, że badając powiązania między instytucjami, należy uwzględnić nie tylko istnienie formalnych relacji, czy też relacji deklarowanych. Często konieczna jest pogłębiona jakościowa analiza. Dopiero takie podejście umożliwia uchwycenie rzeczywistej efektywności funkcjonowania badanej sieci.

Analiza sieci jako narzędzie identyfikacji zasobów wiedzy (Brik S., 2005)

W 2004 roku w Idaho National Laboratory w Stanach Zjednoczonych postanowiono skonsolidować prowadzone dotąd niezależnie projekty badawcze związane z paliwami wodorowymi w jeden duży program „Hydrogen Initiative Program". Program miał się stać podstawą nowej strategii biznesowej laboratorium. Aby sprawnie uruchomić tę nową inicjatywę przeprowadzono ocenę kompetencji w omawianym zakresie istniejących w Laboratorium. Było to szczególnie istotne ze względu na dużą liczbę badaczy pracujących w różnych budynkach i lokalizacjach. Wszystkich pracowników (w sumie 38 osób), którzy mieli być zaangażowani w te nowa inicjatywę poproszono o wskazanie osoby, którą uważają za eksperta w każdej z 47 wyróżnionych kategorii wiedzy związanej z paliwami wodorowymi. Badani mogli wskazywać zarówno osoby z laboratorium jak i spoza niego. Następnie wyniki ankiety przeanalizowano za pomocą metod analizy sieciowej i w efekcie sporządzono 47 grafów prezentujących ekspertów w poszczególnych zakresach. Wyniki analiz zostały skonsultowane z grupą fokusową złożoną z części badaczy z laboratorium – umożliwiło to szczegółową i wiarygodną interpretację poszczególnych grafów.

Wyniki badania umożliwiły wskazanie grypy badaczy najczęściej wskazywanych jako ekspertów czyli kluczowych dla powodzenia inicjatywy. Ponadto okazało się, że w niektórych dziedzinach laboratorium dysponuje kilkoma ekspertami a w niektórych tylko jednym. Taka wiedza może mieć bardzo duże znaczenie dla zarządzania Programem. Jeśli bowiem jedyny ekspert jest np. w wieku przedemerytalnym należałoby odpowiednio wcześniej podjąć działania zmierzające do zatrudnienia bądź wyszkolenia osoby, która mogła by zająć miejsce eksperta gdy ten przejdzie już na emeryturę.

Przywoływany przykład pokazuje jak można wykorzystać analizę sieciową do mapowania kompetencji, doświadczenia i wiedzy zgromadzonych w organizacji bądź organizacjach. Podejście takie może być użyteczne np. przy badaniu inicjatyw klastrowych czy ocenie regionalnego potencjału naukowo-badawczego.

Ewaluacja sieci partnerów w projektach INTERREG IIIB CADSES (Olejniczak et al., 2008a)

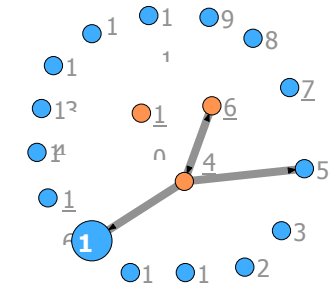
Jednym z elementów oceny udziału polskich partnerów w projektach realizowanych w ramach programu INTERREG IIIB CADSES była analiza sieci kontaktów między partnerami realizującymi poszczególne projekty. Charakterystyczną cechą

programów współpracy terytorialnej jest to, że realizowane są zazwyczaj przez wiele instytucji z różnych regionów i krajów. W założeniu ma to prowadzić do nawiązywania stałej współpracy oraz wymiany doświadczeń i dobrych praktyk. Jednakże w rzeczywistości okazuje się, że współpraca może układać się w różny sposób. Informacje na temat współpracy z partnerami w ramach projektów zebrano od przedstawicieli polskich instytucji (partnerów projektów CADSES) podczas wywiadu telefonicznego. Pytano o intensywność współpracy badanej instytucji z wszystkimi instytucjami zaangażowanymi w dany projekt na różnych etapach jego realizacji (formułowanie koncepcji projektu, realizacja projektu) oraz przed nim (czy wcześniej współpracowali) i po jego zakończeniu (współpraca planowana i współpraca realizowana). Takie podejście pozwoliło na prześledzenie relacji między partnerami projektu (i ich wpływu na jego realizację) oraz na ocenę trwałości współpracy po zakończeniu projektu. Analizy przedstawiano w formie graficznej (przykład: patrz rycina). Przeprowadzone analizy pozwoliły rozróżnić projekty ze względu na mniejszą lub większą intensywność współpracy na poszczególnych etapach oraz pokazać na ile realizacja programu przyczyniła się do nawiązania trwałej współpracy między partnerami projektów. Jednym z ważnych wniosków z badania jest to, że udział w projekcie niekoniecznie przekładał się na nawiązanie trwałej współpracy.

Omawiany przykład pokazuje w jaki sposób można wykorzystać analizę sieci w ewaluacji "miękkich" interwencji mających na celu budowanie partnerstw między instytucjami. Istotne jest także to, że zbierając informacje na temat relacji łączących poszczególne instytucje w różnych momentach realizacji projektu (programu) można w ścisły sposób scharakteryzować nie tylko stan sieci, ale również jej dynamikę.

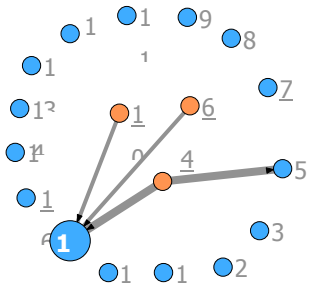
Rycina 22. Przykład analizy sieci - program CADSES

Współpraca przed projektem



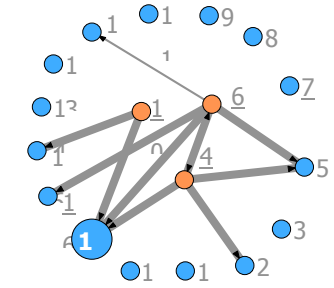
współpraca przed projektem

Współpraca na etapie aplikowania



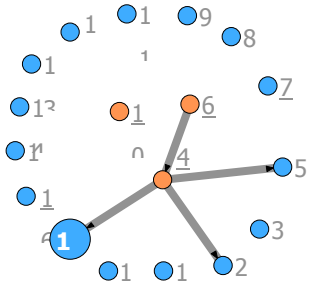
intensywna
średnia
słaba

Współpraca na etapie realizacji



intensywna
średnia
słaba

Współpraca po projekcie



współpraca po projekcie

lider partner partner badany partner niebadany kierunek relacji

X nr partnera zagranicznego X nr partnera polskiego

2.3.3 Możliwości zastosowania w ewaluacjach rozwoju regionalnego

Przytoczone wyżej przykłady pokazują w jaki sposób analiza sieciowa może być zastosowana w ewaluacji interwencji publicznych. Jak już podkreślano jest ona zazwyczaj stosowana łącznie z innymi metodami badawczymi (co jest zgodne z bardziej ogólnym postulatem triangulacji metod badawczych w celu podniesienia wiarygodności analiz). W przypadku programów rozwoju regionalnego analiza sieci znajduje przede wszystkim zastosowanie wszędzie tam gdzie istotnym elementem wdrażania programu (projektu) bądź też jego efektem jest sieć współpracy formalnej bądź nieformalnej między instytucjami lub osobami.

W tym miejscu warto dodać, że funkcjonowanie sieci można analizować z różnych perspektyw. Wynika to z tego, że można wyróżnić dwie uzupełniające się perspektywy analizy sieci organizacji (Provan et al., 2007):

- perspektywę poszczególnych organizacji działających w sieci oraz
- perspektywę sieci jako całości.

Perspektywy te bywają również określane jako podejście na poziomie mikro i makro (Wasserman, Galaskiewicz, 1994) albo perspektywa sieci egocentrycznej (ang. *egocentric network*) i perspektywa sieci jako całości (ang. *whole network*) (Kilduff, Tsai, 2003). W zależności od przedmiotu i celu danej ewaluacji należy odpowiednio uwzględniać jedną z perspektyw albo też obie równocześnie (Provan, Milward, 2001).

Podstawową zaletą analizy sieciowej jest to, że umożliwia ona ujmowanie nawet bardzo skomplikowanych i wielowymiarowych relacji między wieloma elementami w sposób ścisły i skwantyfikowany. Pozwala to na zwiększenie obiektywności analiz i co za tym idzie, wyników badania oraz rekomendacji.

Ta szeroka możliwość analizy i oceny relacji ma dwa ważne aspekty.

Po pierwsze, relacje można analizować w odniesieniu do systemu przestrzennego, czy społeczno-gospodarczego, na który działa dany program, jak też i w systemie w wdrażania samego programu (a więc relacji miedzy interesariuszami).

Po drugie, możliwość analizy i oceny relacji dotyczyć może zarówno relacji fizycznych - przestrzennych (np. sieci drogowe) w ujęciu czy to statycznym (połączenia) czy dynamicznym (przepływy), ale także - a może przede wszystkim - relacji "miękkich" - współpracy, wymiany informacji, obiegu innowacji.

Możliwości wizualizowania sieci (za pomocą odpowiedniego oprogramowania) dają szansę ciekawej i sugestywnej prezentacji wyników badania – co może znacząco podnieść jego użyteczność dla końcowych odbiorców. Wizualizowanie powiązań sieciowych może być też bardzo przydatne na analitycznym etapie badania, służąc np. jako sposób wyboru instytucji do pogłębionych analizy (można oczekiwać, że instytucja najsilniej powiązana z innymi będzie najbardziej „świadoma" tego, co dzieje się w całej sieci).

Nie bez znaczenia jest też to, że analiza sieci ma już solidną podbudowę teoretyczną, szczegółowe i ugruntowane metody oraz procedury badawcze (por. początek tego podrozdziału). Dostępne jest także wiele wersji oprogramowania komputerowego wspomagającego analizę sieciową, także oprogramowania dostępnego nieodpłatnie.

Analizy sieci mają również swoje **ograniczenia i wyzwania.** Bardzo istotną trudnością w prowadzeniu badań sieciowych jest to, że analizami należy objąć całość badanej populacji. Analiza sieciowa jest bardzo wrażliwa na braki danych i w zasadzie nie ma tu możliwości badania próby i przenoszenia wniosków na całą populację. Konieczność uwzględnienia w analizach wszystkich elementów tworzących daną sieć nastręcza oczywiście wiele kłopotów. Szczególnie w przypadku badań ankietowych gdzie uzyskanie stuprocentowej stopy zwrotu jest w zasadzie niemożliwe. Dlatego też zbierając dane sieciowe korzysta się raczej z wywiadów niż z ankiet. Innym sposobem jest korzystanie z danych świadczących o istnieniu formalnych powiązań (zaczerpniętych np. z dokumentów oficjalnych – takich jak np. przynależność do stowarzyszeń). Bardzo często stosuje się również informacje gromadzone w różnego rodzaju bazach danych (np. sieci współpracy instytucji w projektach badawczych bądź sieci współautorstwa publikacji naukowych, (Olechnicka, Płoszaj, 2008).

Z koniecznością uwzględnienia w badaniu całości sieci (brak możliwości próbkowania) wiąże się to, że przynajmniej w niektórych przypadkach rzetelna analiza sieciowa może wymagać o wiele większych nakładów niż standardowe badanie ilościowe. Chodzi tu nie tylko o nakłady finansowe, ale również czas potrzebny na zebranie i opracowanie materiału empirycznego.

Pewnym ograniczeniem może być też to, że danych sieciowych nie można analizować za pomocą standardowych metod statystycznych. Ograniczeniem będą tu z jednej strony potrzebne kompetencje do prowadzenia tego typu badań. Z drugiej

strony należy podkreślić to, że analiza sieci zazwyczaj wymaga posiadania i umiejętności obsługiwania specjalistycznego oprogramowania.

Kolejną kwestią komplikującą badania sieciowe są problemy związane z ich stroną etyczną. Otóż badania sieciowe (szczególnie osób) w przeważającej większości nie mogą być anonimowe na takiej samej zasadzie jak agregowane dane z ankiet. Niejako istotą analizy sieciowej jest to, że chodzi o uchwycenie relacji między konkretnymi a nie wyabstrachowanmi aktorami. Oczywiście zostały wypracowane pewne metody radzenia sobie z tego rodzaju problemami (Penuel et al., 2005). Jednakże planując badania sieciowe należy zwrócić należytą uwagę na zapewnienie ochrony prywatności badanych podmiotów.

3 PODSUMOWANIE

Głównym celem niniejszego badania było zidentyfikowanie i analiza praktyk z zakresu ewaluacji efektów programów rozwoju regionalnego. Poszukiwania objęły zarówno pole programów funduszy strukturalnych jak i szerszą, światową literaturę tematu. Podstawową przesłanką dla badania było dostarczenie inspiracji i pogłębienie krajowej bazy wiedzy dotyczącej metodologii prowadzenia badań ewaluacyjnych. Pytanie o to jakie podejścia i metody mogą nam dostarczyć wiarygodnych dowodów na temat efektów programów publicznych jest wyjątkowo ważkie w perspektywie nadchodzącej fali ewaluacji ex post efektów programów regionalnych. Jak dowodziliśmy we wstępie - jest też kluczowe dla budowania strategicznego myślenia o interwencjach publicznych i wzmacniania absorpcji strukturalnej.

Wyniki przeprowadzonego przeglądu praktyk czterech krajów UE o najwyższej kulturze ewaluacyjnej dostarczają bardzo ograniczonych inspiracji. Analiza 100 raportów, pokazała że większość badań z zakresu efektów programów rozwoju regionalnego współfinansowanych z UE, ma bardzo ubogi zakres i metodologię. Trzy podstawowe słabości to:

- brak spójnej, osadzonej w literaturze tematu koncepcji badawczej i ograniczanie metodologii do prostego zestawu narzędzi i metod;
- ograniczanie "skrzynki" metod do kilku podstawowych elementów - badań ankietowych beneficjentów końcowych, wywiadów, statystyk opisowych;
- brak podejmowania prób analizy relacji przyczynowej - efektów interwencji wychodzących poza prosty opis zaplanowanych produktów, ku rzeczywistym zmianom wywołanych programem (rezultatom, oddziaływaniu), w tym zmianom relacji.

W rezultacie powyższego, analizowane badania nie dostarczają wiarygodnych informacji na temat rzeczywistych efektów programów, ich wpływu na zakładaną i oczekiwaną zmianę społeczno-gospodarczą. Tym samym nie realizują podstawowej funkcji ewaluacji podsumowujących - weryfikowania założeń stojących za politykami i budowania szerszej wiedzy na temat mechanizmów rozwoju.

Za główne przyczyny tego stanu rzeczy (w odniesieniu do analizowanej populacji) uznać należy:

- niewielką formalną odpowiedzialność administracji krajowych za ocenę efektów (obowiązek prawny spoczywa na KE) połączoną z niską presją opinii

publicznej - skierowaną na samo wydawanie pieniędzy, a nie skuteczne ich wykorzystanie;

- strukturalny błąd zakresu badawczego zadań ewaluacji średniookresowych zawierający zarówno doradztwo w bieżących kwestiach wdrażania jak i ocenę efektów;
- przyjęcie zawężającej definicji ewaluacji efektów - rozumianej jako sprawdzenie stopnia wykonania celów programu (tzw. *objective oriented approach*) - a więc zorientowanej na rozliczenie menedżerów z zakładanego planu, a nie na identyfikację rzeczywistych - również tych nieprzewidzianych i "miękkich" zmian we wspomaganym układzie przestrzennym;
- dominację firm konsultingowych i przyjmowanie przez ewaluatorów wyłącznie roli konsultantów procesów, zamiast - co bardziej odpowiednie przy ewaluacjach podsumowujących - bardziej wymagającej roli nauczycieli-mentorów wspomagających procesy uczenia się i pogłębiających rozumienie teorii oraz praktyki interwencji z danego pola.

Z pierwszej części analizy wyniknęła więc konkluzja dotycząca trzech głównych wyzwań stojących przed metodologią ewaluacji efektów programów regionalnych:

1. znalezienie podejść, czy narzędzi ułatwiających konstruowanie spójnych, naukowo wiarygodnych podstaw metodologicznych rozumianych całościowo - jako koncepcja badawcza oraz zestaw metod i narzędzi;
2. znalezienie sposobów powiązania analiz efektów konkretnych programów UE z szerszymi teoriami i praktyką rozwoju - wyjście poza prosty opis produktów i mechaniczną ocenę celów programu, na rzecz analizy efektów, relacji przyczynowych i prób wpisania ich w szerszy kontekst mechanizmów rozwoju społeczno-gospodarczego;
3. znalezienie sposobów ustrukturalizowanej analizy i oceny wpływu na kapitał relacyjny uczestników programów UE i efektów tychże programów, ich poszerzających się zdolności uczenia się i współpracy, sieci powiązań - tak przestrzennych jak i instytucjonalno-społecznych - będących wartością dodaną interwencji publicznych.

W tym wypadku racjonalnym wydawało się wyjście poza ramy funduszy strukturalnych i poszukiwanie inspiracji w całym, szerokim polu ewaluacji interwencji publicznych. Dzięki takiemu podejściu, w drugiej części raportu, opierając się na

kompleksowym przeglądzie literatury światowej zidentyfikowano i przedstawiono dwie propozycje metodologiczne.

Pierwszą zaproponowaną inspiracją była "ewaluacja wspierana teorią" (ang. *theory driven evaluation* -TDE) oraz modele logiczne. TDE jest podejściem ewaluacyjnym (ang. *evaluation approach*) - a więc szerszą filozofią myślenia o badaniu i strategią jego strukturalizowania, podczas gdy modele logiczne są technicznym narzędziem jej obrazowania. Zarówno TDE jak i modele opierają się na założeniu że program można przedstawić w formie hipotezy (*jeśli... to...*). Zgodnie z tym podejściem analiza efektów powinna uwzględnić trzy rodzaje teorii składających się na ogólną teorię programu:

- szersze teorie społeczne (leżące u podstaw założeń programu);
- teorię zmiany (konkretne założenia, co do przyczyn które mają doprowadzić do osiągnięcia pożądanego celu) oraz
- teorię działania (obraną sekwencję wdrażania programu i osiągania produktów).

Ta proponowana inspiracja metodologiczna ma kilka zalet:

- pomaga wyraźnie przedstawić racjonalność programu - jego założenia na temat mechanizmu oczekiwanej zmiany społecznej, zidentyfikować kluczowe pola i połączenia logiczne (jak i relacje zadaniowe);
- zmusza do ustrukturalizowanego myślenie o programie i umożliwia stworzenie solidnej podstawy koncepcyjnej danego badania - postawienia pytań badawczych, logicznej struktury zbierania i analizy danych, stworzenia narracji raportu końcowego;
- zidentyfikowania i wyartykułowania teorii (nie zawsze uświadomionych) którymi kierowali się interesariusze oraz śledzenia przyczynowości;

Jako drugą inspirację zaproponowano analizę sieci (ang. *network analysis*). Teoria sieci koncentruje się na analizie relacji elementów połączonych ze sobą w system. Można analizować relacje między osobami oraz między organizacjami (instytucjami), a także powiązania przestrzenne - między miastami, regionami, państwami lub innymi całościami. Danych do analizy dostarczają specjalnie skonstruowane wywiady lub ankiety eksponujące kwestie siły relacji i współpracy (w przypadku ludzi lub instytucji) albo bądź dane statystyczne czy różnego rodzaju specjalistyczne zbiory danych (w przypadku jednostek przestrzennych, instytucji). Analiza prowadzona jest przy pomocy specjalistycznych programów komputerowych (część

z nich dostępna za darmo). Może dotyczyć zarówno perspektywy poszczególnych organizacji (elementów sieci) jak i układu jako całości. Wyniki analiz prezentowane są w formie graficznej.

Wartością tej inspiracji metodologicznej są:

- możliwość analizy i oceny relacji - zarówno w systemie przestrzennym czy społecznym, na który działa dany program, jak też i w systemie w wdrażania samego programu (a więc relacji interesariuszy);
- możliwość analizy i oceny relacji zarówno fizycznych - przestrzennych (np. sieci drogowe) w ujęciu statycznym (połączenia) jak i dynamicznym (przepływy), ale też - a może przede wszystkim - relacji "miękkich" - współpracy, wymiany informacji, obiegu innowacji;
- ujmowanie nawet bardzo skomplikowanych i wielowymiarowych relacji między wieloma elementami w sposób przejrzysty graficznie a jednocześnie ścisły i skwantyfikowany, co pozwala to na zwiększenie obiektywności analiz.

Naszym daniem obydwie zaprezentowane inspiracje mają także kluczową wartość - ujmują kwestie analizy efektów w sposób całościowy, wprowadzają systemowe, ustrukturalizowane myślenie oparte na mocnych podstawach badawczych. Tym samym mogą w znaczący sposób przyczynić się do pogłębienia i poszerzenia analiz ewaluacyjnych, wzbogacenia ich o relację z teorią, przy jednoczesnym zakorzenieniu w wiarygodnej empirii.

Zamykając niniejsze opracowanie warto krótko nakreślić kierunki dalszych poszukiwań metodologicznych w polu ewaluacji. Widzę trzy główne elementy tych poszukiwań.

Po pierwsze, nowe metody wymagają testowania i adaptacji w polskich warunkach i to w odniesieniu do różnych typów programów, w badaniach o różnej skali. Nowe metody i podejścia mają swoje zalety ale należy również poznawać ograniczenia, które stawiają. Ten postulat dotyczy nie tylko inspiracji zidentyfikowanych i przedstawianych w niniejszym raporcie. W kontekście toczących się światowych debat filozoficzno-metodologicznych (paradygmaty pozytywizmu, konstruktywizmu, realizmu) warto zadawać pytania o ich praktyczne implikacje i podejmować refleksję nad warunkami brzegowymi zastosowania w programach funduszy strukturalnych

poszczególnych podejść, planów badawczych czy metod (np. efektu netto, czy quasi-eksperymentów).

Po drugie, mimo nieco rozczarowujących pierwszych efektów badań raportów zagranicznych nie należy rezygnować z dalszych poszukiwań inspiracji na gruncie europejskiej praktyki ewaluacyjnej. Potencjalnym obiektem zainteresowania mogą być ostatnie badania zlecane przez Komisję Europejską dla programów 2000-2006. Wartą dalszej eksploracji wydaje się także praktyka ewaluacji krajowych interwencji. Mechanizm transferu *know-how* z rozwiniętych pól ewaluacji programów krajowych, na pola programów UE najwyraźniej nie zadziałał (z omawianych już przyczyn). Nie oznacza to jednak, że administracje i ewaluatorzy europejskich krajów o długiej tradycji ewaluacji nie mają nic ciekawego do zaoferowania.

Wreszcie po trzecie, warto dalej śledzić trendy światowe. Na kolejnych konferencjach i w debatach toczonych na łamach periodyków branżowych pojawiają się już nowe pomysły. Pochodzą spoza pola funduszy UE czy nawet polityki regionalnej ale nie umniejsza to ich innowacyjności. Ich adaptowanie do pola polityki regionalnej UE może być polską wartością dodaną. Najnowszym przykładem może być dyskusja na temat zastosowania koncepcji systemów i tzw. myślenia systemowego i (ang. *system thinking*) do ewaluacji skomplikowanych programów publicznych (patrz: Evaluation and Program Planning 2008-31(3); Williams, Imam 2007). Dyskusja ta, jak na razie wciąż toczy się w dużej mierze w polu teorii (Cabrera et al., 2008; Hummelbrunner, 2007) choć pojawiają się już pierwsze sygnały zastosowań praktycznych - i to w polu programów regionalnych UE (Hummelbrunner, 2008).

Na zakończenie warto zauważyć, że skuteczne poszukiwania w polu metodologii ewaluacji efektów są szansą na wprowadzenie naszej nauki, jak i praktyki administracji publicznej w międzynarodową dyskusję - i to w roli innowatorów. Stare kraje UE nie są specjalnie zainteresowane rozwijaniem metod ewaluacji efektów programów regionalnych UE. Nie mają też specjalnie możliwości podjęcia takiej refleksji. Skala interwencji UE jest nieporównywalnie mniejsza do skali programów krajowych - jakiekolwiek próby budowania relacji przyczynowych byłyby w tym wypadku nadużyciem.

Tymczasem w przypadku naszego kraju siła finansowa programów UE jest nieporównywalnie większa od stricte krajowych inicjatyw prorozwojowych. Co więcej, jest prowadzona w sposób ustrukturalizowany (kierunki i cele) i wieloletni. Przy takiej skali interwencji publicznej możemy analizować związki przyczynowe i próbować

powiązać je ze zmianami w naszej przestrzeni. Jest to więc doskonała okazja do eksploracji i zrozumienia mechanizmów rozwoju regionalnego w Polsce jak i testowania najbardziej skutecznych metod ich wspierania. Patrząc rozpoczynającą się w cieniu światowego kryzysu debatę nad przyszłym budżetem UE, wydaje się, że jest to okazja niepowtarzalna. Kolejny budżet UE będzie już znacznie mniej hojny dla Polski. Zaś przy dużo mniejszych zasobach jeszcze większego znaczenia nabierze umiejętność kierowania środków na interwencje, które gwarantują rozwój i pozytywną zmianę społeczną. Bez ugruntowanej wiedzy na temat dotychczasowych efektów i mechanizmów zmian nie będziemy w stanie racjonalnie dokonywać takich strategicznych wyborów.

BIBLIOGRAFIA

Arnold, E., Clark, J. & Muscio, A. (2008) *What the evaluation records tells us about EU Framework Programme performance,* Brussels: Technopolis Group papers.

Bachtler, J. (2000) *The evaluation of regional policy in the European Union – main report,* Glasgow: European Policies Research Centre.

Bachtler, J. (2001) Quod erat demonstrandum? Ewaluacja polityki regionalnej. *Studia Regionalne i Lokalne*, **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. Report to the Commission of the European Communities (DG XVI),* Glasgow: European Policies Research Centre, University of Strathclyde.

Bachtler, J. & Wren, C. (2006) Evaluation of European Union Cohesion Policy: Research Questions and Policy Challenges. *Regional Studies*, **40(2)**, pp.143-153.

Ball, P. (2008) *Masa krytyczna. Jak jedno z drugiego wynika*. Kraków: Insignis.

Bamberger, M., Rugh, J. & Mabry, L. (2006) *RealWorld evaluation: working under budget, time, data, and political constraints*. Thousand Oaks, Calif. ; London: Sage.

Batorski, D. (2008) "Metody analizy sieci i ich zastosowanie w ewaluacji"; w: Haber, A. & Szałaj, M. (red.) *Środowisko i warsztat ewaluacji,* s.167-191. Warszawa: Polska Agencja Rozwoju Przedsiębiorczości.

Bendyk, E. (2004) *Antymatrix. Człowiek w labiryncie sieci*. Warszawa: Wydawnictwo W. A. B.

Bickman, L. (1987) Using program theory in evaluation. *New Directions for Program Evaluation*, **33**, 5-18.

Brik S. (2005) Application of Network Analysis in Evaluationg Knowledge Capacity. *New Directions for Evaluation*, **107**,

Cabrera, D., Colosi, L. & Lobdell, C. (2008) Systems thinking. *Evaluation and Program Planning*, **31(3)**, 299-310.

Chen, H.T. (1994) *Theory-Driven Evaluations*. Newbury Park CA: Sage Publications, Inc.

Chen, H.T. (2004) *Practical Program Evaluation: Assessing and Improving Planning, Implementation, and Effectiveness*. Thousands Oaks: Sage Publications, Inc.

Chen, H.T. (2005a) "Program theory"; w: Mathison, S. (red.) *Encyclopedia of Evaluation,* s.340-342. Thousand Oaks, Calif. ; London: SAGE Publications.

Chen, H.T. (2005b) "Theory-Driven Evaluation"; w: Mathison, S. (red.) *Encyclopedia of evaluation,* s.415-419. Thousand Oaks, Calif. ; London: SAGE Publications.

Chen, H.-T. & Rossi, P.H. (1983) Evaluating with sense: The theory-driven approach. *Evaluation Review*, **7**, 283-302.

Christie, C.A. (2008) "What have we learned about RCTs, Gold Standards, and credible evidence?", *European Evaluation Society Conference*, Lisbon, 2.10.2008.

Cooke, P. (2004) *Regional Innovation Systems: The Role of Governances in a Globalized World*. Routledge.

Council of the European Union (1999) Council Regulation (EC) No 1260/1999 of 21 June 1999 laying down General Provisions on the Structural Funds. *Official Journal of the European Communities*, **L161**, 26.6.1999, 1-45.

Cousins, B.J. & Leithwood, K.A. (1986) Current empirical research in evaluation utilization. *Review of Educational Research*, **56(3)**, 331-364.

CSES (2003) *Mid term evaluation of the Ease Wales 2000-2006 Objective 2 & Transitional Programme - Final report,* Sevenoaks: Centre for Strategy and Evaluation Services.

Daniluk, D. (2008) "Ewaluacja - kluczowy instrument oceny jakości wykorzystania krajowych środków publicznych - perspektywa zastosowania w kontekscie prac nad budżetem zadaniowym", *IV Ogólnopolska Konferencja Ewaluacyjna*, Warszawa.

Daponte, B.O. (2008) *Evaluation Essentials: Methods For Conducting Sound Research (Research Methods for the Social Sciences)*. San Francisco: Jossey-Bass.

Davis, R. (2005) *Moving from Logical to Network frameworks: a modular matrix approachto representing and evaluating complex programs,* 19.07.2005, Stanford: Creative Commons Attribution.

Davis, R. (2007) "Social Network Analysis as an Evaluation Tool: Experiences with International Development Aid Programmes", *The UK Social Network Conference*, Queen Mary College University of London, London, July 2007.

Dobrzyńska, A. (2008) *Tworzenie potencjału ewaluacyjnego w Polsce na przykładzie Funduszy Strukturalnych*. praca magisterska, Wydział Filozofii i Socjologii, Uniwersytet Warszawski,

Donaldson, S. (2008) "A practitioner's guide for gathering credible evidence in the Evidence-Based Global Society"; w: Donaldson, S.I., Christie, C.A. & Mark, M.M. (red.) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?,* s.239-251. Los Angeles: Sage Publications.

Donaldson, S.I. (2007) *Program Theory-Driven Evaluation Science: Strategies and Applications*. New York: Lawrence Erlbaum.

Donaldson, S.I., Christie, C.A. & Mark, M.M. (2008) (red.) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Sage Publications, Inc, Los Angeles.

Durland, M. & Fredericks, K. (2005) An Introduction to Social Network Analysis. *New Directions for Evaluation*, **107**,

EPRC & Fraser of Allander Institute (2000) *Methodologies used in the Evaluation of the Effectiveness of European Structural Funds: A Comparative Assessment. Final Report to Scotish Executive,* Glasgow: European Policies Research Centre, Fraser of Allander Institute.

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*, **40(2)**, pp.249-258.

European Commission. (1995a) *MEANS Handbook 1 - Organising intermediate evaluation in the context of Partnership*. Brussels: Commission of the European Communities.

European Commission. (1995b) *MEANS Handbook 2 - Identifying the structuring effects of community interventions*. Brussels: Commission of the European Communities.

European Commission. (1995c) *MEANS Handbook 3 - Measuring the employment effects of Community Structural Interventions*. Brussels: Commission of the European Communities.

European Commission. (1995d) *MEANS Handbook 4 - Applying the multi-criteria method to the evaluation of Structural Programmes*. Brussels: Commission of the European Communities.

European Commission. (1995e) *MEANS Handbook 5 - Quality assessment of evaluation reports: a framework*. Brussels: Commission of the European Communities.

European Commission. (1995f) *MEANS Handbook 6 - The ex post evaluation of Structural Funds Programmes: a methodological framework*. Commission of the European Communities, Brussels.

European Commission. (1995g) *MEANS Handbook 7 - The ex ante evaluation of Structural Funds Programmes: a methodological framework*. Brussels: Commission of the European Communities.

European Commission. (1997) *Evaluating EU expenditure programmes: A guide for ex post and intermediate evaluation*. Brussels: Office for Official Publications of the European Communities.

European Commission. (1999) *MEANS collection: evaluating socio-economic programmes*. Luxembourg: DG-Regio, Office for Official Publications of the European Communities.

European Commission (2000a) *Methodological Working Paper 3 - Indicators for Monitoring and Evaluation: An indicative methodology,* The New Programming Period 2000-2006 Methodological Working Papers; Brussels: European Commission - DG-Regio.

European Commission (2000b) *Methodological Working Paper 7 - Ex Ante Evaluation and Indicators for INTERREG (Strand A and B),* The New Programming Period 2000-2006 Methodological Working Papers; Brussels: European Commission - DG-Regio.

European Commission. (2001) *Project Cycle Management Manual*. Brussels: European Commission.

European Commission (2002a) *Methodological Working Paper 8 - The Mid Term Evaluation of Structural Fund Interventions,* The New Programming Period 2000-2006 Methodological Working Papers; Brussels: European Commission - DG-Regio.

European Commission (2002b) *Methodological Working Paper 8a - The Mid Term Evaluation of Structural Fund Interventions – INTERREG III,* he New Programming Period 2000-2006 Methodological Working Papers; Brussels: European Commission - DG-Regio.

European Commission (2003) *Evaluation in the Commission. Annual Evaluation Review 2002,* (SEC(2003)406), Brussels: European Commission - DG-Budget.

European Commission (2004a) *Evaluating EU activities - A practical guide for the Commission services,* Brussels: European Commission - DG-Budget.

European Commission (2004b) *Methodological Working Paper 9 - The Update of the Mid Term Evaluation of Structural Fund Interventions,* The New Programming Period 2000-2006 Methodological Working Papers; Brussels: European Commission - DG-Regio.

European Commission. (2004c) *Project Cycle Management guidelines, vol 1*. Luxembourg: Office for Official Publications of the European Communities.

European Commission (2006a) *Working Document no 1 - Indicative Guidelines on Evaluation Methods: Ex ante evaluation,* he New Programming Period 2000-2006 Methodological Working Papers; Brussels: European Commission - DG-Regio.

European Commission (2006b) *Working Document no 2 - Indicative Guidelines on Evaluation Methods: Monitoring and Evaluation Indicators,* The New Programming Period 2007-2013 Working Documents; Brussels: European Commission - DG-Regio.

European Commission (2006c) *Working Document no 3 - Commission methodological paper giving guidelines on the calculation of public or equivalent structural spending for the purposes of additonality,* The New Programming Period 2007-2013 Working Documents; Brussels: European Commission - DG-Regio.

European Commission (2007a) Evalsed - the resource for the evaluation of socio-economic development.
http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/index_en.htm, accessed: 10.10.2008.

European Commission (2007b) *Working Document no 5 - Indicative Guidelines on Evaluation Methods: Evaluation during the Programming Period,* The New Programming Period 2007-2013 Working Documents; Brussels: European Commission - DG-Regio.

European Commission (2007c) *Working Document no 6 - Measuring Structural Funds Employment Effects,* The New Programming Period 2007-2013 Working Documents; Brussels: European Commission - DG-Regio.

European Evaluation Society (2007) EES Statement: The Importance of a methodologically diverse approach to impact evaluation – in particular with respect to development aid and development interventions. http://www.europeanevaluation.org/, accessed: 20.06.2008.

Ferry, M. & Olejniczak, K. (2008) *Wykorzystanie ewaluacji w zarządzaniu programami unijnymi w Polsce,* Warszawa: Ernst & Young, Program "Sprawne Państwo".

Frechtling, J.A. (2007) *Logic modeling methods in program evaluation*. San Francisco, Calif.: Jossey-Bass.

Fredericks, K. (2005) Network Analysis of a Demonstration Program for the Developmentally Disabled. *New Directions for Evaluation*, **107**,

Freeman, L.C. (2004) *The Development of Social Network Analysis: A Study in the Sociology of Science*. North Charleston: BookSurge.

Furubo, J.-E., Rist, R.C. & Sandahl, R. (2002) *International atlas of evaluation*. New Brunswick, U.S.A.: Transaction Publishers.

GHK Consultants. (2006) *The Evaluation of Socio-Economic Development. The Evalsed Guide (2nd edition)*. London: GHK Consultants.

Górniak, J. (2007) "Ewaluacja w cyklu polityk publicznych"; w: Mazur, S. (red.) *Ewaluacja Funduszy Strukturalnych - perspektywa regionalna,* s.11-28. Kraków: Uniwersytet Ekonomiczny w Krakowie.

Hausner, J. (2007) "Czy należy pomagać regionom słabo rozwiniętym?", *Sympozjum Regional Studies Association - Sekcja Polska*, Warszawa, 18.04.2007.

Henry, G.T., Julnes, G. & Mark, M.M. (1998) Realist evaluation: An emerging theory in support of practice. *New Directions for Evaluation*, **78**,

Howell, E. & Yemane, A. (2006) An assessment of evaluation designs. Case studies of 12 large federal evaluations. *American Journal of Evaluation*, **27(2)**, 219-236.

Hummelbrunner, R. (2007) "Systemic Evaluation in the field of regional development"; w: Williams, B. & Iraj, I. (red.) *Systems concepts in Evaluation. An expert anthology,* Ann Arbor: American Evaluation Association.

Hummelbrunner, R. (2008) Monitoring and evaluation for strategic governance of Cohesion Policy, European Evaluation Society Conference, Lisbon, 2.10.2008

Jonak, Ł. (2007) "Analiza sieci społecznych i dynamika sieciowa"; w: Nowak, A., Winkowska-Nowak, K. & Rychwalska, A. (red.) *Modelowanie matematyczne i symulacje komputerowe w naukach społecznych - podręcznik dla studentów I roku,* Warszawa: Wydawnictwo SWPS Academica.

Kellogg Foundation (2004) *The logic model development guide,* Battle Creek: W.K. Kellogg Foundation.

Kierzkowski, T. (2002) Evaluation of socio-economic programmes in Poland - current situation and future challenges. *Annual Conference of Regional Studies Association "Evaluation and EU Regional Policy"*,

Kilduff, M., & Tsai, W. (2003) *Social Networks and Organizations*. London: SAGE.

Knowlton, L.W. & Phillips, C.C. (2008) *The Logic Model Guidebook: Better Strategies for Great Results*. Los Angeles: Sage Publications, Inc.

Kozak, M., Ledzion, B., Olejniczak, K. & Wojtowicz, D. (2009) *Ex Post Evaluation of Cohesion Policy Programmes 2000-06 Co-financed by ERDF. Working Package 11 – Task 2: National Assessment Report POLAND,* draft version - January 2009, Warsaw: EPRC & EUROREG.

Krajowa Jednostka Oceny (2005) "Rynek ewaluacji w Polsce - diagnoza stanu i perspektyw rozwoju", *I konferencja ewaluacyjna - Ewaluacja programów o charakterze społeczno-gospodarczym finansowanych z funduszy strukturalnych*, Ministerstwo Gospodarki i Pracy & Polska Agencja Rozwoju Przedsiębiorczości, Warszawa, 16 wrzesnia 2005.

Krajowa Jednostka Oceny (2008) Proces ewaluacji polityki spójności w Polsce. Podsumowanie dotychczasowych doświadczeń. Plany i wyzwania na przyszłość.

Lipsey, M.W., Crosse, S., Dunkle, J., Pollard, J. & Stobart, G. (1985) Evaluation: The state of the art and the sorry state of the science. *New Directions for Program Evaluation*, **27**, 7-28.

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.

Mathison, S. (2005) (red.) *Encyclopedia of evaluation* Sage, Thousand Oaks, Calif. ; London.

Mc Laughlin, J. & Jordan, G. (2004) "Using logic models"; w: Wholey, J.S., Hatry, H.P. & Newcomer, K.E. (red.) *Handbook of Practical Program Evaluation,* s.7-32. San Francisco: Jossey-Bass.

Olechnicka, A. & Płoszaj, A. (2008) *Polska nauka w sieci? Przestrzeń nauki i innowacyjności. Raport z badań,* Warszawa: Agnieszka Olechnicka.

Olejniczak, K. (2004) "Ewaluacja Programu Współpracy Przygranicznej Phare Polska-Niemcy - projekty"; w: Gorzelak, G., Bachtler, J. & Kasprzyk, M. (red.) *Współpraca transgraniczna Unii Europejskiej. Doświadczenia polsko-niemiecki,* s.294-318. Warszawa: Wydawnictwo Naukowe "Scholar".

Olejniczak, K. (2005) Teoria i praktyka ewaluacji w świetle doświadczeń europejskich. *Studia Regionalne i Lokalne*, **4(22)**, 35-58.

Olejniczak, K. (2007) "Ewaluacja jako narzędzie zarządzania w sektorze publicznym"; w: Gorzelak, G. & Tucholska, A. (red.) *Rozwój, region, przestrzeń,* s.361-396. Warszawa: Ministerstwo Rozwoju Regionalnego.

Olejniczak, K. (2008) *Mechanizmy wykorzystania ewaluacji. Studium ewaluacji średniookresowych INTERREG III*. Warszawa: Wydawnictwo Naukowe Scholar.

Olejniczak, K., Ledzion, B., Kościelecki, P., Płoszaj, A., Pander, Wojciech & Domaradzka, A. (2008a) *Ocena udziału polskich partnerów w projektach realizowanych w ramach programu INTERREG IIIB CADSES,* Warszawa: EGO s.c. dla MRR.

Olejniczak, K., Ledzion, B. & Pander, W. (2008b) *Ocena potencjału oraz stanu prac w ramach jednostek ewaluacyjnych funkcjonujących w Instytucjach Zarządzających oraz Instytucjach Pośredniczacych I i II stopnia w ramach NPR 2004-2006 i NSRO 2007-13,* Warszawa: EGO s.c. dla Krajowej Jednostki Oceny.

Patton, M.Q. (1997) *Utilization-focused evaluation: the new century text*. Thousand Oaks, Calif.: Sage Publications.

Patton, M.Q. (2008) *Utilization-focused evaluation. 4th edition*. Los Angeles, London: Sage Publications.

Pawson, R. (2002) Evidence-based Policy: The Promise of 'Realist Synthesis'. *Evaluation*, **8(3)**, 340–358.

Pawson, R. (2006) *Evidence-based policy: a realist perspective*. London: Sage.

Pawson, R. & Tilley, N. (1997) *Realistic evaluation*. London: Sage.

Penuel, W.R., Sussex, W., & Korbak, C. (2005) "Mapping the distribution of Expertise and Resources in School: Investigating the Potential of Using Social Network Analysis in Evaluation", *Joint Conference of The Canadian Evaluation Society and the American Evaluation Association*, Toronto, October 2005.

Petrosino, A., Rogers, P., Huebner, T. & Hacsi, T. (2000) Program Theory in Evaluation: Challenges and Opportunities. *New Directions for Evaluation*, **87(Fall)**,

Porter, M.E. (1998) *On Competition*. Harvard Business School Press.

Provan, K. G. & Milward, B. (2001) Do Networks Really Work? A Framework for Evaluationg Public Sector Organizational Networks. *Public Administration Review*, **61(4)**,

Provan, K.G., Fish, A., & Sydow, J. (2007) Interorganizational networks at the network level: A review of the empirical literature on whole networks. *Journal of Management*, **33**,

Rank, D. & Williams, D. (1999) Partial benefit/cost in the evaluation of the Canadian Networks of Centres of Excellence. *Evaluation and Program Planning*, **22**,

Reichardt, C. & Mark, M.M. (2004) "Quasi-exprimentation"; w: Wholey, J.S., Hatry, H.P. & Newcomer, K.E. (red.) *Handbook of Practical Program Evaluation,* s.126-149. San Francisco: Jossey-Bass.

Rogers, P. (2005) "Logic model"; w: Mathison, S. (red.) s.232-235. Thousand Oaks, Calif. ; London: SAGE Publications.

Rossi, P.H., Freeman, H.E. & Lipsey, M.W. (1999) *Evaluation: a systematic approach. 6th edition*. Thousand Oaks, Calif. ; London: Sage Publications.

Rozporządzenie Rady (WE) (2006) Rozporządzenie Rady (WE) nr 1083/2006 z dnia 11 lipca 2006 r. ustanawiające przepisy ogólne dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego oraz Funduszu Spójności i uchylające rozporządzenie (WE) nr 1260/1999. *Dziennik Urzędowy WE*, **OJ L210**, 31.7.2006,

Russ-Eft, D.F. & Preskill, H.S. (2001) *Evaluation in organizations: a systematic approach to enhancing learning, performance, and change*. Reading, Mass. ; Oxford: Perseus.

Shadish, W.R., Cook, T.D. & Campbell, D.T. (2001) *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

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*, Ministerstwo Gospodarki i Pracy & Polska Agencja Rozwoju Przedsiębiorczości, Warszawa, 16 wrzesnia 2005.

Stufflebeam, D.L. & Shinkfield, A.J. (2007) *Evaluation theory, models, and applications*. San Francisco, Calif.: Jossey-Bass ; Chichester : John Wiley [distributor].

Swianiewicz, P., Herbst, J., Lackowska, M., & Mielczarek, A. (2008) *Szafarze darów europejskich. Kapitał społeczny a realizacja polityki regionalnej w polskich województwach*. Warszawa: Wydawnictwo Naukowe Scholar.

Szałaj M., Ledzion B. (2008) "Ewaluacja bieżąca sieci Regionalnych Ośrodków Europejskiego Funduszu Społecznego", w: Olejniczak, K., Kozak, M. & Ledzion, B. (2008) (red.) *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki.* Wydawnictwa Pedagogiczne i Naukowe, Warszawa.

Szlachta, J. (2006) "Polskie doświadczenia w zakresie ewaluacji", *II Konferencja Ewaluacyjna - Ewaluacja ex-ante jako narzędzie doskonalenia zarządzania strategicznego w administracji publicznej*, Ministerstwo Rozwoju Regionalnego & Polska Agencja Rozwoju Przedsiębiorczości, Warszawa, 22 listopada 2006.

Szlachta, J. (2007) "Wprowadzenie"; w: Haber, A. (red.) *Ewaluacja ex post. Teoria i praktyka badawcza,* Warszawa: Polska Agencja Rozwoju Przedsiębiorczości.

Tavistock Institute, GHK Consultants & IRS. (2003) *The Evaluation of Socio-Economic Development. The Evalsed Guide (1st edition)*. London: Tavistock Institute.

Toulemonde, J. (2000) Evaluation culture(s) in Europe: differences and convergence between national practices. *The Danish Evaluation Society Annual Conference*,

United Way. (1996) *Measuring Program Outcomes: A Practical Approach*. USA: United Way of America.

Wasserman, S., & Faust, K. (2007) *Social network analysis: methods and applications*. New York: Cambridge University Press.

Wasserman, S., & Galaskiewicz, J. (1994) *Advances in Social Network Analysis: Research from the Social and Behavioral Sciences*. Newbury Park.: Sage Publications.

Weiss, C.H. (1997) How can theory-based evaluation make greater headway? *Evaluation Review*, **21(4)**, 501-524.

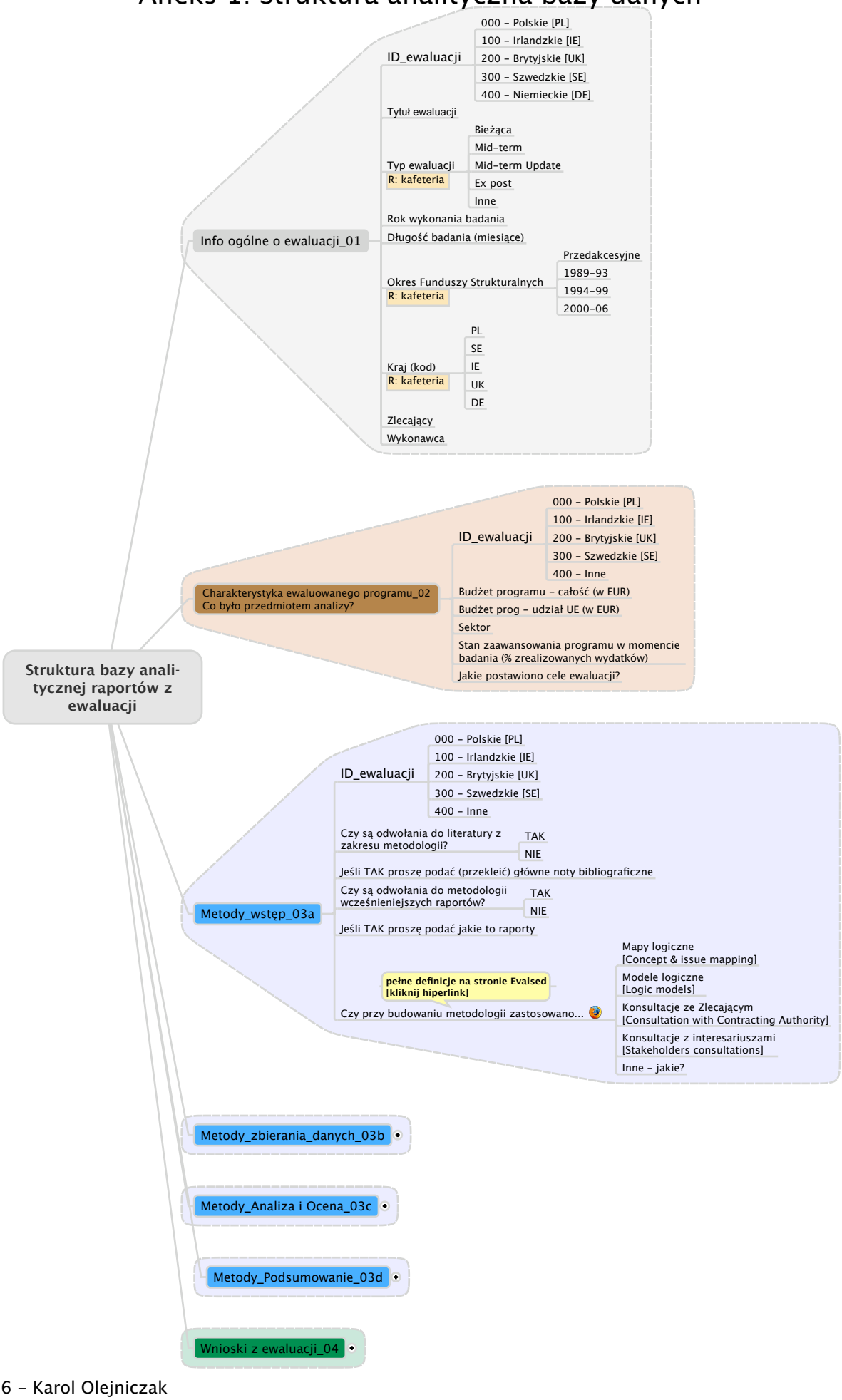
Weiss, C.H. (2004) On Theory-Based Evaluation: Winning Friends and Influencing People. *Evaluation Exchange*, **IX(4)**, 2-3.

Weiss, C.H. & Bucuvalas, M.J. (1980) Truth Tests and Utility Tests: decision-makers' frame of reference for social science research. *American Journal of Sociological Review*, **45(April)**, 302-313.

Williams, B. & Imam, I. (red.) (2007) *Systems concepts in Evaluation. An expert anthology,* Ann Arbor: American Evaluation Association.

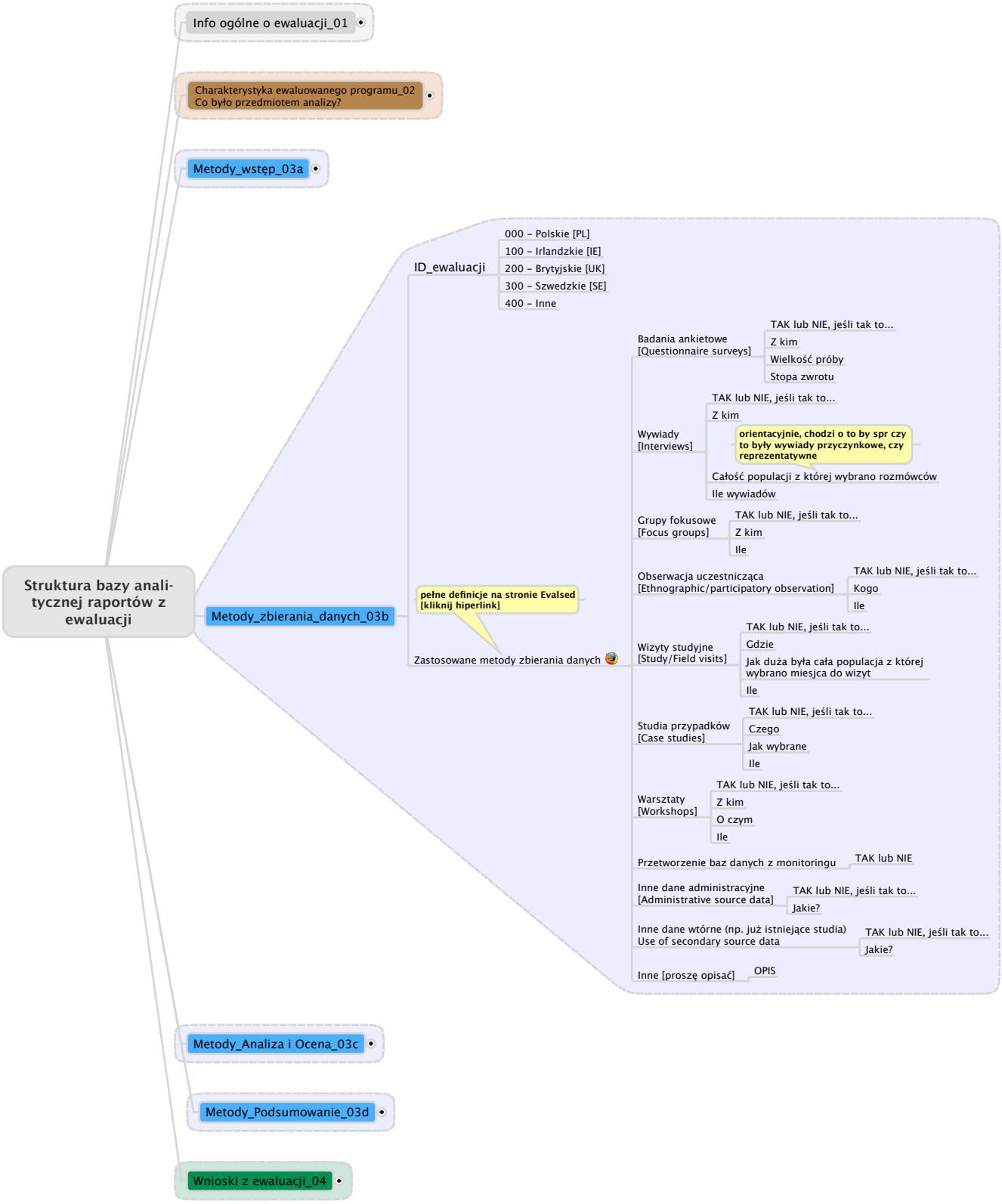
Zawicki, M. (2007) Trzy płaszczyzny ewaluacji – z praktyki polskiej administracji publicznej. *Zarządzanie Publiczne*, **2(2)**, 113-124.

Aneks 1: Struktura analityczna bazy danych



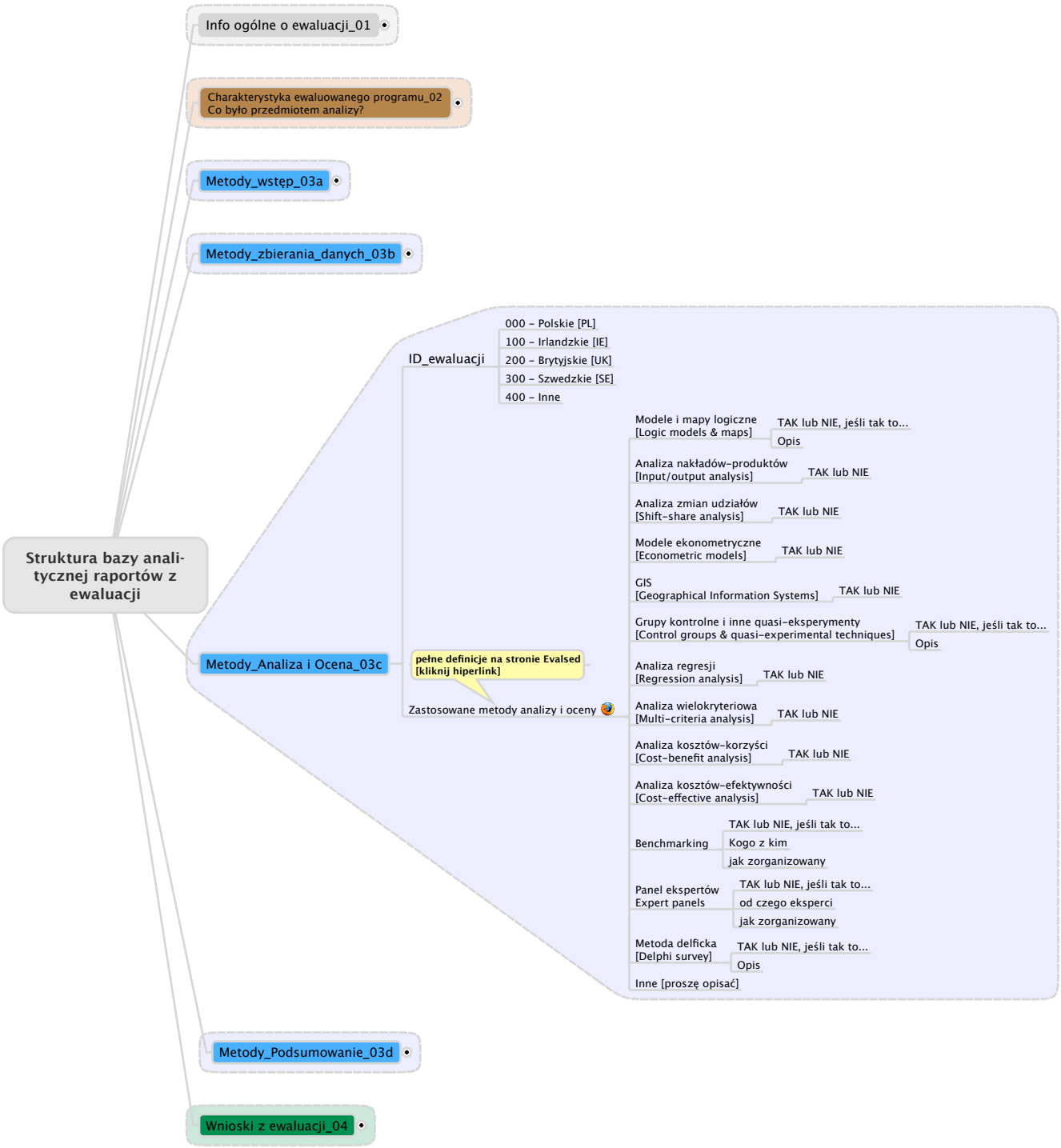
09-01-16 - Karol Olejniczak

Aneks 1: Struktura analityczna bazy danych



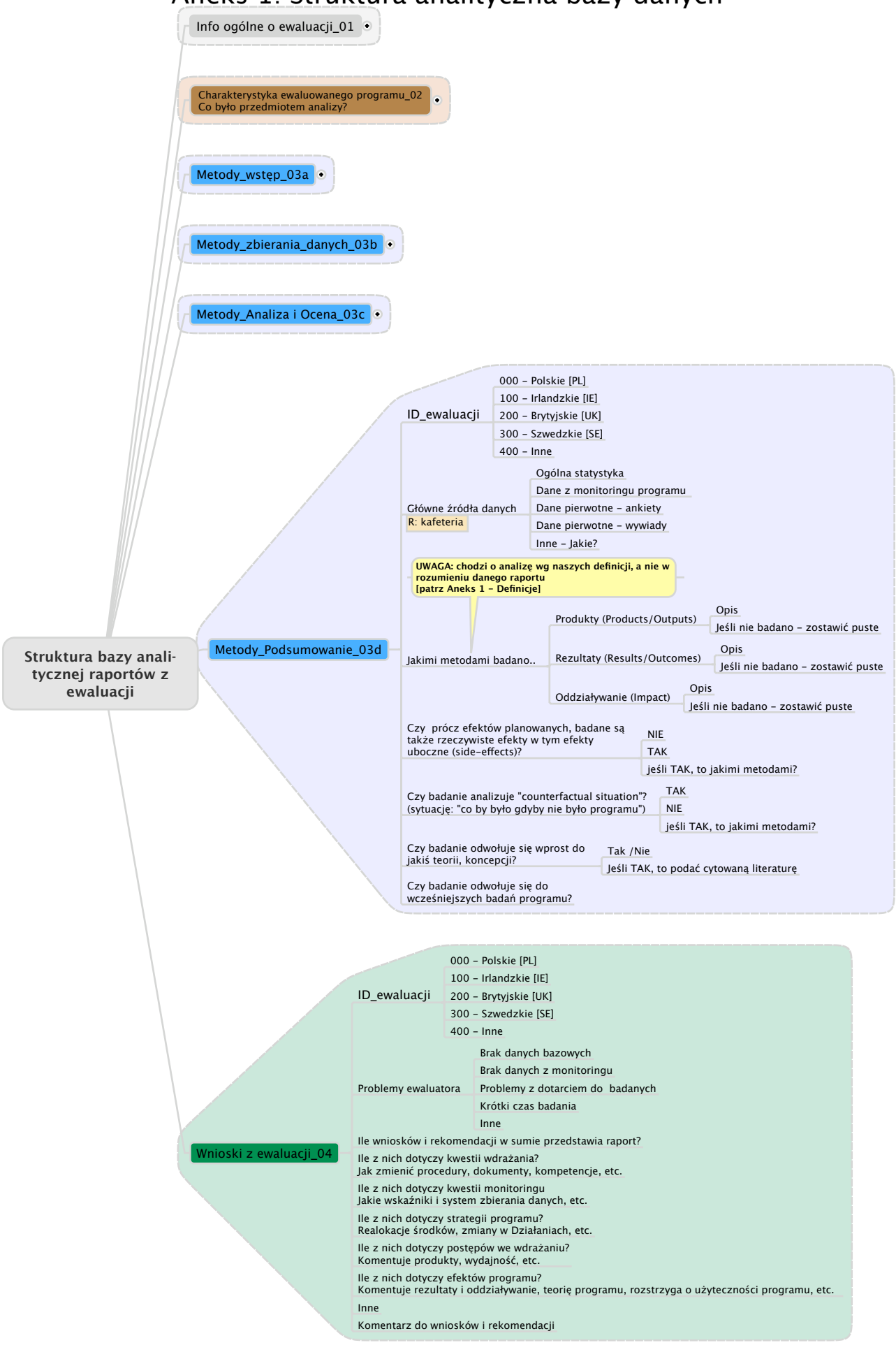
09-01-16 - Karol Olejniczak

Aneks 1: Struktura analityczna bazy danych



09-01-16 - Karol Olejniczak

Aneks 1: Struktura analityczna bazy danych



09-01-16 – Karol Olejniczak

Aneks 2

Praktyka ewaluacji w polu funduszy strukturalnych

Ekspertyzy krajowe Irlandii, Niemiec, Szwecji i Wielkiej Brytanii

dr Richard Boyle
Institute of Public Administration, Dublin

Evaluation practice in the field of structural funds in Ireland

Introduction

Ireland is a parliamentary democracy with a written constitution. It has a population of just over 4 million. In the 1970s and into the 1980s evaluation in Ireland was limited in scope and largely peripheral to decision making. Ireland had a reasonably well-developed system for the evaluation of individual projects, but it did not have a strong tradition of evaluation of policies and programmes.

One significant exception to this scenario of limited interest in evaluation in the 1980s is in relation to EU expenditure in Ireland. Since 1989, evaluation of the EU Structural Funds has been a formal requirement of those receiving the assistance, and led to further developments in evaluation practice in Ireland in the 1990s. The funds were applied through a number of operational programmes conducted under a joint Irish-EU agreed Community Support Framework (CSF) plan.

Outside of EU expenditure, there has been a renewed interest in evaluation of public expenditure in Ireland in the late 1990s and early 2000s. As the economy has dramatically improved, the interest has been in examining the value being obtained for the significant increases in public expenditure that have taken place. As part of a more wide-ranging programme of management reforms, there has been an emphasis on programme review and evaluation. To this end, in 1997, the government approved a series of what are called 'value for money and policy reviews' to be carried out as part of a system of comprehensive programme evaluation. The success of this initiative has been patchy so far, with some organisations and programmes lending more support than others to the process (Boyle, 2002; 2005).

In examining evaluation capacity development in Ireland with regard to evaluating the Structural Funds, it is useful to look at developments covering the last three Community Support Frameworks (CSF) periods: 1989-1993, 1994-1999 and 2000-2006. As noted by Hegarty (2003):

> During both the 1989 to 1993 and 1994 to 1999 periods, Ireland was treated as an Objective 1 region. Total structural funds expenditure in Ireland amounted to 4.2 billion euros under the CSF 1989 to 1993 and 5.8 billion euros under the CSF for 1994 to 1999, equivalent to about 1.7 per cent of average GDP

over the entire period. In the 2000 to 2006 programming period, Ireland was designated as two NUTS II regions: the Border, Midland and Western (BMW) region (which enjoyed Objective 1 status to 2006); and the Southern and Eastern (S&E) region (which qualified for transitional Objective 1 funding to 2005). Including the 4 per cent performance reserve, the total structural funds allocation to Ireland under the CSF 2000 to 2006 amounts to 3.2 billion euros (equivalent, on an annual basis, to about 0.4 per cent of 2001 GDP).

Demand side

The NDP/CSF Evaluation Unit is the key body involved in evaluation of the National Development Plan (NDP) and Community Support Framework (CSF). The Unit is an independent unit under the aegis of the Department of Finance. The Unit undertakes or commissions evaluations of the NDP/CSF operational programmes and also provides advice and assistance to the European Commission, government departments, regional assemblies and other bodies on programme monitoring and evaluation issues generally. The Unit is part funded by the Structural Funds.

The NDP/CSF evaluation unit was established when evaluation arrangements under the 1994-1999 CSF were strengthened from previous arrangements, with the formal requirement for prior appraisal, mid-term review and ex-post evaluation. In this context, the Department of Finance recognised the need for improved co-ordination of evaluation activities. To achieve better co-ordination, it was decided in 1996 to establish a central evaluation unit based in the Department of Finance. This unit was funded under the Structural Funds CSF Technical Assistance operational programme. The main tasks identified for the unit were:

- setting up and maintaining a central register of evaluations;
- reviewing the methodology used in the evaluations carried out with a view to identifying best practice and making recommendations for future evaluations;
- reviewing the work programmes for evaluation of the operational programmes with a view to identifying the scope for co-ordinating evaluations and avoiding duplication;
- examining requests for evaluations from the regional authorities and making recommendations in this matter;
- reviewing the adequacy of the performance indicators for the CSF and making recommendations for their improvement or for alternative indicators where appropriate;
- considering the arrangements for mid-term review of the CSF and the operational programmes and making recommendations in the matter.

In practice, and particularly from the perspective of evaluation capacity development, the central evaluation unit played a key role in the identification and dissemination of good practice in evaluation. One particularly helpful role was the production of good practice guides and reviews of practice. One review, of the ongoing evaluation function in the CSF in 1998, is widely seen as promoting a common shared understanding of the purpose and focus of evaluation of the Structural Funds in Ireland. A particularly helpful part of the review from this perspective was the outlining of the approach that should be taken to ongoing evaluation work appropriate in a CSF context. Five key evaluation questions were identified:

- *Rationale* (is there a market failure?)
- *Continued relevance* (to what extent do the objectives remain relevant in the light of external developments?)
- *Effectiveness* (are we meeting our objectives?)
- *Efficiency* (are benefits commensurate with costs, and could it be delivered more economically?)
- *Impact* (what are net effects or changes in the socio-economic situation that can be attributed to the programme?)

These five evaluation questions have subsequently been used to guide evaluation activity and provide a common approach to the evaluation of the Structural Funds. Hegarty (2003) notes the importance of this agreed approach for evaluation practice:

> The adoption of this evaluation approach has provided a valuable 'anchor' for the NDP/CSF evaluation process. In particular, the adoption and communication of this schema has helped create a greater awareness and understanding of the purpose and focus of evaluation work among programme managers and other stakeholders. The key evaluation questions underpin both the design of evaluation work programmes and terms of reference for individual evaluation projects.

Taking the 2000-2006 period as an example of how demand is structured, the NDP/CSF evaluation unit developed an interim evaluation work programme for the period 2001-2003 (up to the mid-term evaluation). The work programme comprised some 16 proposed evaluation projects, and was adopted by the NDP/CSF technical assistance monitoring committee in 2001 (this committee was chaired by the managing authority of the NDP/CSF and included representatives of the operational programme managing authorities and the European Commission, acting in an advisory capacity). A number of projects were selected on the basis of concerns raised in the ex ante evaluation. The work programme was subject to revision due to delays

in the start up of some programmes and in recruitment of evaluators to the unit. Seven evaluations were actually carried out between 2001 and 2003.

With regard to the mid-term evaluation process, a mid-term evaluation planning group was set up in mid 2002 at NDP/CSF level to coordinate the process. This group brought together representatives of the operational programme managing authorities and the relevant Commission directorates under the chairmanship of the NDP/CSF managing authority. The NDP/CSF evaluation unit acted as secretariat to the group. The group also functioned as the steering committee for the mid-term evaluation of the NDP/CSF. The committee agreed an overall timetable for the evaluation process and to core terms of reference for the operational programme level evaluations.

Supply side

In the 1989-1999 periods, the institutional arrangements for CSF evaluation were quite complex. Six of the nine operational programmes appointed an external evaluator (typically a consultancy firm); one operational programme had an independent evaluation unit; one operational programme had an independent evaluation unit plus an external evaluator; one operational programme had neither an independent unit nor an external evaluator; and the ESF evaluation unit, the management structure of which was not integrated with CSF management structures, covered activity under six operational programmes.

The independent evaluation units are of particular interest. They represent the development of specific evaluation expertise within the public service, that was not there before. They also, as a review of the ongoing evaluation function found, contributed:

> … in a general way to the fostering of a more evaluative and analytical culture within the public service … The extent of our reliance on a limited pool of available and suitable qualified private consultants is already problematic. Without the units, we would be more reliant on this small pool (CSF Evaluation Unit, 1998, p. 30).

The development of these independent evaluation units is discussed below, including the central CSF evaluation unit mentioned above.

Establishment of independent evaluation units

The European Social Fund (ESF) evaluation unit was established in January 1992 on the initiative of the European Commission, to evaluate the effectiveness of all human

resource development measures supported by the ESF. It was located in the Department of Labour. In August 1993, the industry evaluation unit became operational. It was responsible for monitoring and evaluation of measures financed through the operational programme for industrial development. It was located in the Department of Industry and Commerce. In late 1993, major changes in the structuring of government departments saw the main functions of the Departments of Labour and Industry and Commerce, including the two evaluation units, merged in a new Department of Enterprise and Employment (now entitled the Department of Enterprise, Trade and Employment). Also, in 1995, an analysis and evaluation unit was established to monitor and evaluate aspects of the operational programme for agriculture, rural development and forestry (hereafter, the unit is referred to in the text as the agriculture evaluation unit).

These three evaluation units were separate and independent from the Irish national administration. Studies carried out under their auspices were independent and not subject to veto by the national administration. The ESF evaluation unit was 100 per cent funded by the EU Commission. The Industry evaluation unit was 75 per cent funded by the EU Commission, with the remaining funds coming from the national administration.

The staffing structure for the units varied. The ESF evaluation unit had one manager, seven evaluators, and five support staff, though staff turnover and slow recruitment meant that actual capacity was often below this level. The industry evaluation unit staff complement fluctuated from eight staff in the early days (five evaluators and three support staff) to five later on (four evaluators, one of whom managed the unit, and one support staff). The agriculture evaluation unit had a staff complement of one manager, two evaluators and one support staff, but actual employment levels were often lower, with one evaluator being common. Staff retention was an ongoing problem for the units.

A key criterion for employment in the units was work experience. This was particularly important for the credibility of evaluators going out into the field, to the executive agencies, to gather data. Young graduates straight out of third-level institutions were not seen as having sufficient credibility to deal with senior managers in agencies. A background of working experience in analysis and evaluation was seen as enhancing the standing of the evaluator, and subsequently their ability to conduct evaluation studies.

For the 2000-2006 period, the Department of Finance reviewed and revised evaluation arrangements. A decision was made not to continue the departmental based independent evaluation units. Instead, the central evaluation unit was expanded, and renamed the NDP/CSF evaluation unit. The monitoring committee for the technical assistance programme sets the work programme for this central unit.

Central evaluation unit supply side issues

During the 1994-1999 period, the staffing of the central evaluation unit was comprised of one senior evaluator who managed the unit, one evaluator and support staff. It was agreed that the evaluators would be employed on a temporary, one-year, non-pensionable contract basis in the first instance, with the possibility of extension for further periods (with the option of secondment if an existing public servant took up a post).

For the 2000-2006 period, the unit had a staff complement of five evaluators and two administrative staff. Evaluators were employed on three-year contracts. The unit had a budget of 4.6 million euros over 2000-2006. The NDP/CSF evaluation unit has conducted some ongoing evaluation work itself. The unit has also contracted out evaluation work to independent, private suppliers of evaluation.

Private sector suppliers

The main evaluation suppliers contracted by the NDP/CSF evaluation unit have been relatively small, specialised providers (either small firms, or smaller parts of larger consultancy and accountancy firms, usually with a strong focus on economic evaluation expertise). One organisation, the Economic and Social Research Institute, through the development of a macro-economic model capable of being used to assess the impact of Structural Funds in the Irish economy, has developed particular strengths for macro-level evaluation of the Structural Funds at the ex-ante, mid-term and ex-post stages. Thus there is a dependence on a relatively small pool of evaluators in Ireland. However, the counterpoint to this issue is that the small number of firms involved have, over the years, built up specialised expertise which is tailored to the requirements of Structural Funds evaluation, with a good working knowledge both of the techniques needed and also the social and political context within which the evaluations take place. The smallness of the market for Structural Funds evaluators within Ireland therefore has both weaknesses but also some strengths with regard to evaluation capacity development.

Products

Ex ante, ongoing and ex post evaluations are carried out:

- Ex-ante evaluation is carried out before programmes are agreed with the main focus being on the question of programme rationale. The ex-ante evaluation of the 2000-2006 NDP/CSF was a two-part process. Firstly, the ESRI was commissioned to prepare an independent analysis of investment needs for the Irish economy for the 2000 to 2006 period. At a later stage, a formal ex-ante evaluation was carried out by the CSF Evaluation Unit (now NDP/CSF Evaluation Unit) in November 1999 and submitted to the European Commission with the NDP.
- Ongoing (including mid-term) evaluation. Mid-term evaluation of all operational programme is carried out, as well as of the NDP/CSF. This is a requirement under the EU regulations. The main objective of these evaluations, which feed into the mid-term review process, was to assess the need for changes to the programmes at the half way stage in the light of progress to date and changes in the socio-economic situation. During the 2000-2006 period these evaluations were commissioned from external consultants.
- In addition to the mid-term reports, other evaluations are also carried out of particular areas or themes under the different operational programmes. These reports are commissioned or undertaken by the NDP/CSF Evaluation Unit, under an agreed work programme.
- Ex-post evaluation is carried out after the completion of the programming period. The main focus is on the question of the impact of the programmes. Details of the results of the ex-post evaluation for the 1994-1999 NDP/CSF are available on the European Commission website.

In practice, in the 1994-1999 period, with regard to the conduct of ongoing evaluation, independent units and external evaluators operated quite differently. The work of the independent units tended to focus on in-depth measure level evaluations. The work of external evaluators tended to be in the form of analysis of individual issues, shorter reports of broad focus, and oral advice. Some of the external evaluators adopted a 'watchdog' role whereby they raised questions and issues of concern at Monitoring Committee meetings and elsewhere (CSF Evaluation Unit, 1998).

The three evaluation units focused on the evaluation of measures that made up the operational programmes, rather than investigating broad policies. Judgements were made as to the process of running various measures and programmes and the effectiveness of particular interventions. The studies had a clear public accountability focus, being concerned with the benefits gained from the expenditure of public

money. In terms of timescale, evaluation studies in the units would normally be expected to take between six and twelve months to complete.

The 1994-1999 period represented the high point with regard to the number of evaluation reports produced, largely due to the significant number of ongoing evaluation reports produced by the independent evaluation units. The number of evaluation reports dropped in the 2000-2006 period.

Regarding dissemination and the follow up of reports, in the 1994-1999 period, the case of the industry evaluation unit provides a best practice example. After the report had been considered by the steering group and subsequently finalised, it was circulated to the responsible person(s) in the Irish administration to obtain a formal response to the recommendations contained in the report. A paper setting out the main recommendations and the official response (e.g. whether a recommendation was accepted and the approach and time frame for implementation, or whether rejected and the reasons why), plus a summary of the report, formed the basis for deliberations by the monitoring committee. Copies of full reports were available to members of the committee on request. The intention was that the monitoring committee take decisions on which recommendations it wished to have implemented. These recommendations would then be the main focus of follow-up reports – either specifically or through the annual reports – to the committee on implementation of the recommendations.

For the 2000-2006 period, learning from the industry evaluation unit experience and, importantly, in line with detailed procedures set out in the CSF, completed evaluation reports had to be formally considered by the relevant operational programme monitoring committees. The procedures required that the relevant managing authority must, following consultation with implementing bodies, submit a formal response on each evaluation recommendation to the monitoring committee. Subsequently, the managing authority is obliged to report back to the monitoring committee at regular intervals on progress made in implementing agreed recommendations. These procedures, introduced for the 2000-2006 period, represented an important innovation in the evaluation system compared with the previous CSF when evaluation reports often received little or no formal consideration by monitoring committees.

Evaluation reports are compiled by the NDP/CSF evaluation unit, and from 2000, all reports are accessible online from http://www.ndp.ie

With regard to quality, steering committees oversaw quality issues, but these steering committees varied in their approach. Steering arrangements for the industry

evaluation unit were widely seen as being good practice, with some of the other committees being over large and in practice focused on policy as opposed to quality of evaluation issues. The industry evaluation unit steering committee agreed terms of reference for individual evaluations and approved draft reports from a quality perspective. The membership of the committee was small, including representatives from the main departments and agencies with responsibility for the operational programme, the NDP/CSF evaluation unit, the European Commission, and three independent technical experts. The steering committee did not concern itself with policy issues as such but focused on methodological and quality-related issues. It was widely regarded as contributing to enhanced quality of evaluation reports.

Evaluation utilisation

There is some evidence that evaluation has influenced both programme design and the allocation of Structural Fund resources. The mid-term evaluation of the CSF 1994-1999, for example, made a significant contribution to the mid-term review and clearly influenced decisions taken on the re-allocation of Structural Funds by the Commission in conjunction with the national authorities (Hegarty and Honahan, 1998). The evaluation also influenced a number of non-financial decisions. For example concerns about deadweight in some areas classified as targeted subsidies was reflected in a commitment to reduce aid rates to the private sector in a number of areas and to focus training and capacity expansion grants more on smaller firms.

The ex ante evaluation of the 2000-2006 NDP had some influence in shaping both the NDP and the CSF. The initial analysis of investment priorities influenced a wide-ranging public debate on the priorities of the NDP and was the subject of much discussion amongst policy makers and in debate around the design of the NDP (Hegarty and Fitzgerald, 2000).

Interim and ongoing evaluation work, focused more on programme management and implementation, has generally met with a good response from programme managers. Some of the reports have also led to considerable media comment and public debate. For example an interim evaluation report of NDP/CSF investment in the road network drew attention to a significant escalation in the estimated cost of completion of the national roads programme.

Perspectives

In general terms, the disbandment of the departmental-based evaluation units and centralisation of evaluation efforts into one evaluation unit based in the Department of Finance has led to some reduction in evaluation capacity for the evaluation of

Structural Funds within the public sector. The central evaluation unit itself, even with the extended contracts for evaluators, has had some difficulties in maintaining staffing levels. And the kind of detailed evaluative information provided by the departmental-based units has not, in the main, been replaced by the operational programme level evaluations carried out on contract by private suppliers.

A number of general lessons for those interested in building evaluation capacity emerge from the Irish experience to date. These lessons can be divided into (a) drivers that encourage evaluation capacity development, and (b) blockages to evaluation capacity development (taken from Boyle, 2005).

Drivers of evaluation capacity development

Among the key drivers are:

- An external 'push' to institutionalise evaluation has had a significant effect on evaluation practice. As Hegarty (2003) states: "…it is clear that the requirements of EU regulations have helped promote an evaluation culture and capacity in Ireland." This external push to promote evaluation has been a key determinant in systematising the approach to evaluation compared to what had been a very ad hoc approach previously. The requirement to develop evaluation was taken seriously.
- The central support function of the NDP/CSF evaluation unit has been influential in promoting and developing evaluation. While small in terms of staffing complement, the unit has played a significant role in developing and spreading an evaluation culture. Two aspects of their work, in particular, stand out from the perspective of evaluation capacity development. One is the development and spread of good practice, through the use of guidelines, the promotion of common approaches, the development of networks and the like. The second is the review role, taking a periodic overview of development and promoting subsequent change and development.
- With regard to the issue of review, it is clearly important that evaluation capacity and practice is subject to formal, periodic scrutiny. The review of the ongoing evaluation function in the 1994-1999 CSF period by the central evaluation unit was influential in terms of identifying strengths and weaknesses in evaluation capacity and suggesting ways forward.
- The independent evaluation units established during the 1989-1993 and 1994-1999 CSF periods were an important and innovative approach to building evaluation capacity where evaluation was not a significant function previously.

Barriers to evaluation capacity development

Among the main barriers are:

- Without the strong external push and formal requirement for evaluation, demand may vary. In the case of the 2000-2006 CSF period, where Structural Funds were a much smaller part of the total National Development Plan, the decision was made to abolish the departmental based independent evaluation units and develop the central evaluation unit. In the changeover period, the evaluation expertise built up in the departmental based units was lost to the system, and developing and maintaining the staff complement of the enhanced central unit has proved problematic.
- As a small country, the number of potential suppliers of professional evaluation services in Ireland is limited. There is a strong reliance on a small pool from which to draw evaluation expertise. The corollary of this point, however, is that a strong working relationship can be developed with professional experts who have developed a good working knowledge of the specific requirements of evaluation practice.
- Without formal procedures and practices to integrate evaluation into decision-making, the impact of evaluation may be limited. This, in turn, may lead to decision-makers questioning the merits of putting effort into building evaluation capacity.
- On a detailed point of employment of evaluators to evaluation units, the experience of the independent evaluation units was that one-year rolling contracts were problematic. These led to high turnover of staff, and there were subsequent delays in recruitment. The NDP/CSF evaluation unit, to ensure some stability in the employment base, provided three-year contracts, but even here retention problems were encountered.

References

Boyle, R. (2002), ' Two-Tiered Approach: Evaluation Practice in the Republic of Ireland', in J.E. Furubo, R.C. Rist and R. Sandahl (eds.), *International Atlas of Evaluation*, New Brunswick, NJ: Transaction Publishers, pp. 261-272

Boyle, R. (2005), Evaluation Capacity Development in the Republic of Ireland, ECD Working Paper Series 14, Washington D.C.: The World Bank (www.worldbank.org/ieg/ecd)

CSF Evaluation Unit (1998), *Review of Ongoing Evaluation Function in the Community Support Framework (CSF) for Ireland 1994-1999*, Dublin: CSF Evaluation Unit

Hegarty, D. (2003), 'Framework for the Evaluation of the Structural Funds in Ireland', paper prepared for Fifth European Conference on the Evaluation of the Structural Funds, Budapest, 26-27 June (conference organised by the European Commission

Hegarty, D. and P. Honahan (1998), 'Mid-Term Evaluation of CSF for Ireland', paper prepared for European Conference on Evaluation Practices in the Field of Structural Policies, Seville, March.

Hegarty, D. and J. Fitzgerald (2000), 'Ex Ante Evaluation Process for the 2000 to 2006 Period in Ireland', paper prepared for Fourth European Conference on Evaluation of the Structural Funds, Edinburgh, September.

Evaluation practice in the field of Structural Funds in Germany

metis

supporting good governance

This paper was produced for Karol Olejniczak, Ph.D.
Assistant Professor, EUROREG - University of Warsaw

by Dipl. Geogr. Lukas Wortmann, metis GmbH – supporting good governance

Evaluation practice in the field of Structural Funds in Germany

Content

1 Evolution of evaluation 4

2 Demand side 6

3 Supply side 7

4 Products 8

5 Use of evaluation 10

6 Perspectives 11

7 References 12

1 Evolution of evaluation

The content

This paper aims to provide insight into the evaluation practice and system in the field of Structural Funds in Germany. Section 1 describes the tradition of evaluation in Germany. Section 2 deals with the situation on the side of institutions contracting evaluations in the field of Structural Funds. The situation on the side of evaluators will be discussed in section 3. Section 4 is concerned with the population of the executed studies and the practice of results dissemination. Section 5 focuses on the trends in utilization of evaluation reports and the practice of use of evaluation studies. Finally, section 6 aims to identify possible developments in the field of evaluation studies in Germany. A full listing of the literature reviewed is presented in section 7.

1 Evolution of evaluation

The beginning of evaluation practice in Germany dates back to the 1970ies. Two main factors for its appearance can be identified. First, the development in the United States, where evaluation has become increasingly important since the mid 1960ies. The second factor initiating evaluation efforts in Germany was the national political agenda, thus an internal impulse.

These changes in the political agenda started in the early 1970ies. The social-liberal government in Germany foresaw a neo-Keynesian economic policy with large investments in infrastructure and an extension of the Welfare State. The complexity of support programmes and the high degree of policy interventions generated an increased need for information. Above all, this information was of high importance to the government and the parliament in order to monitor public expenditures. The high demand for control and analysis of policy interventions and the use of public funds required an instrument that could respond to those needs.

In this context, evaluation was regarded as an important analytical tool for policy development and implementation. The regional policy Framework Plan ("*Rahmenplan*") of 1969 included evaluation provisions for the first time. Evaluation in the framework of Structural Funds only started several years later.

At the beginning of the development of evaluation in Germany, the driving force came from the national, i.e. federal level. The states ("*Länder*") at the regional level and the communities at the local level followed with their own approaches to evaluation.

Ministries elaborated specific evaluations according to their policy field. Due to this fact, evaluation in Germany has evolved in a highly distinctive technical way.

Evaluations were mainly conducted in the field of education policy, urban development policy, and infrastructure policy, i.e. mainly in the policy fields that have a close connection to regional policy.

From the mid 1970ies to the end of the 1980ies, evaluations of economic related policies became more important, e.g. labour market policy, technology funding policy, and promotion of small and medium sized enterprises. During this period, the implicit function of evaluations was increasingly regarded as a way to lower public expenditures.

From the beginning of the 1990ies, the assumed inefficiency of the public sector became the main focus of evaluation.

A broad political consensus over the last 30 years supported the afore-mentioned evaluation practice during this time.

2 Demand side

Germany has a "subsumed" system of Structural Fund delivery (Taylor/Bachtler/Polverari, 2001). This means that projects are appraised and selected using established domestic resource allocation mechanisms. Structural Fund budgets are integrated alongside domestic ones, and decision-making is often delegated to single competent agencies rather than reserved for dedicated Structural Fund committees.

There is no strong "programme identity". This means that Structural Funds programmes often have little profile as entities in themselves, "limiting the chances that evaluating their design and management will be a high priority among the stakeholders or that this will be seen as a legitimate task for evaluation" (Taylor/Bachtler/Polverari, 2001). The accent is on delivering the Structural Funds through domestic institutional and policy channels, and minimising the need to create new or distinctive methods to deploy these resources.

Due to Germany's federal structure, programmes are developed and evaluated mainly at the level of "*Länder"*. The main contracting institutions are the ministries of the "*Länder"*.

Responsibilities for evaluation lie within the departments responsible for Structural Funds, following the implementation logic in Germany. Each "*Land"* deals with evaluation on an individual basis, no coordination at federal level is foreseen. However, coordination has taken place at a scientific level (see "*DeGEval*" in section 3).

Evaluation studies were planned in past Structural Funds periods with an ad-hoc need-driven approach. In Structural Funds period 2007-13, annual reports are foreseen in order to proceed from ad-hoc evaluations to a concept of on-going evaluations.

3 Supply side

Due to a large number of "*Länder"* and programmes, Germany has a big and active evaluation community with a wide range of evaluators specialized in Structural Funds evaluation. The position of evaluation is relatively established regarding the supply side, i.e. the availability of evaluators experienced in the field of Structural Funds is good. Networking takes place to some extent, an evaluation society, the "*Deutsche Gesellschaft für Evaluation*" ("*DeGEval"*), was established in 1997. "*DeGEval"* aims to promote a continuous dialogue on the theories, models, and methods between professionals involved in evaluation research and teaching, evaluation implementation, and commissioning of evaluation. It builds a communication platform for all those involved or interested in evaluation by means of organization of annual conferences, its presence in the German Journal of Evaluation, and email services among others.

Evaluation services are provided by consultancies and research institutions. Consultancies have become more dominant, whereas research institutes nowadays are not as dominant as they used to be in the past.

A few big evaluation companies, and a variety of small research institutions and consultancies are involved in evaluations.

Principally, fluctuation is low and entry barriers for newcomers are high. Most of the evaluators are in the business for more than ten years, some of the big evaluators even more than 30 years. Therefore, evaluation in the field of Structural Funds can not be regarded as an "ocean of opportunities". It is a large field of activities which is, however, shared by a community of established evaluators. This established community capitalize on the know-how built-up in the frame of already accomplished evaluations in the field of Structural Funds.

4 Products

The number and types of evaluative studies increased from the beginning of the 1990ies when the new "*Länder*" became Objective 1 regions providing for a variety of new programmes. There is an overall growing trend that can be identified throughout the programming years.

Evaluation of programmes is by far the most common evaluation in terms of subject, since programmes of "*Länder"* dominate the field of Structural Funds in Germany. Structural Funds regulations foresee obligatory ex ante evaluation for each programme. The same applies to interim and updated interim evaluations during the last programming period; this type has been substituted by on-going evaluations that are currently commissioned. Structural Funds regulations do not require ex post evaluation by the programmes. This is why only few programmes provide such a type of evaluation.

The following analysis is based on presentations held at "*DeGEval´s*" conference „*10 Jahre DeGEval – Zukunft der Evaluation*" on the 10th until 12th October 2007 in Dresden (Grajewski/Schmüdderich 2007, Schwab 2007). The quality of evaluation has increased over the programming periods. Even though evaluation research is lacking, the quality of reports has increased as well as the expertise of evaluators. Evaluators benefit from their experiences gained through previously accomplished evaluations in the field of Structural Funds ("learning-by-doing"). Contracting institutions notice a high quality of evaluation reports in terms of comprehensibility and clarity of recommendations. They assess the quality of evaluation reports in technical terms even more positively. Furthermore, they miss a stronger relation to practical experience. Altogether, they wish for a more user-friendly type of evaluation report.

Even though the quality of evaluation has increased in technical terms, the methodology has not changed. Evaluators still follow the same data-orientated approach. They focus on the evaluation of monitoring data, qualitative interviews, and written interviews of final beneficiaries. Other methods like scoring-methods or participative approaches, for example, are not considered. Governance aspects are largely not covered. However, an analysis how decision-making is organised at different levels and how these levels interact with each other would enhance the quality of evaluation reports. Present emphasis lies on indicators and results and not on the way how these results have been achieved. Therefore, the approach seems to be rather result-orientated than process-orientated. On the other hand, contracting institutions state, that evaluation results and recommendations are most significant for their work.

Contracting institutions have become more professional regarding the steering of evaluation processes and preparation of tenders. Communication processes between contracting institutions and evaluators have improved in the frame of the European Social Fund. With regard to the European Regional Development Fund, information exchange mechanisms are missing.

The most common dissemination strategy seems to be the internet. Evaluations are mostly available on the web sites of contracting institutions, i.e. mostly federal state ministries. However, this concerns most often only the latest evaluation reports. Very few evaluation reports are published in academic journals.

A central database does not exist. There is no opportunity to access all conducted evaluations in the field of Structural Funds in one place.

5 Use of evaluation

Evaluation mainly follows the regulations and guidelines of Structural Funds. The main user groups are the institutions responsible for programme management who also contract evaluation. Evaluations are mostly used for reporting and verifying effects. In contrast to Anglo-Saxon countries, results are mainly discussed internally in the technical departments, but not in a wider audience.

The share of reports being used in the frame of public relations is rather small. To a much lesser extent, results are discussed at the political level. According to that, the influence on political decision-making processes is limited. Evaluators consider mid term evaluations and on-going evaluations, to be most influential, followed by ex ante evaluations on the second place and ex post evaluations on the third place. There exists an upward trend concerning the influence of evaluation reports. Results of evaluations are increasingly recognized by administrations and at the political level. This applies particularly to reports in the frame of the European Social Fund.

Meeting regulatory requirements seems to be the principal driving force of evaluation practice. Evaluation reports are largely carried out with a summative focus, i.e. where the emphasis is on justification of resources used and achievement of impacts. Programme managers would benefit more from a formative focus, i.e. where the emphasis is on improvement of operational practices and increase the effectiveness of the programmes delivered.

A stronger participation of the political level in the discussion of results would be desirable. The upward trend to a stronger recognition of evaluation reports by the political level can be considered a positive development. Apart from the influence of politics, a wider utilization of evaluation reports in public discussions would positively affect the development of evaluation research which ultimately could deliver inputs on a bigger scale to evaluation practice and practitioners.

6 Perspectives

Evaluation reports present an upward trend in Germany regarding number and quality. The latter can foremost be attributed to the experiences of evaluators made in the field of Structural Funds but also to contracting institutions that have professionalized the steering of evaluation processes and preparation of tenders. This trend is likely to continue in the future programming periods.

The “subsumed” system of Structural Funds delivery in Germany is unlikely to change in the foreseeable future. This may influence also the further development of evaluation as a management tool in the field of Structural Funds. There is no strong programme identity in a “subsumed” system of Structural Funds. It is therefore unlikely that evaluating the design and management of programmes will be considered a legitimate task for evaluation or that it becomes a high priority among the stakeholders. Evaluation conceived as a management tool, however, requires formative techniques, focussing on programme design and delivery.

For that purpose, interim and on-going evaluations are the most relevant stage, taking place during implementation and “aiming to improve policy design and delivery over the remainder of the programme” (Taylor/Bachtler/Polverari, 2001). The on-going evaluation in Germany 2007-13 may be seen as an opportunity in this regard.

On-going evaluations could also provide for a more intensive contact between programme administrations and evaluators during the evaluation process, resulting also in a more intensive communication of results. A larger participation of different stakeholders could entail a wider discussion of results of evaluations.

7 References

Eser T (2001) Evaluation und Qualitätsmanagement – Anforderungen und Konsequenzen für die EU-Strukturpolitik, Informationen zur Raumentwicklung, Heft 6/7, 327-339.

Geberth F (2008) Laufende Bewertung, monitoring und andere Prozessketten – Abgrenzungen, Berührungspunkte und Überschneidungen. Beispiele aus der Praxis für ELER: Maßnahmen- und Entwicklungsplan Ländlicher Raum Baden-Württemberg 2007-2013 (MEPL II), presentation held at the Frühjahrsworkshop des AK Strukturpolitik der DeGEval, 12th and 13th of June in Düsseldorf.

Grajewski R and K Ortner (2007) Evaluation der 2. Säule in Deutschland und Österreich – Mehr Gemeinsamkeiten als Unterschiede? Presentation held at the Jahrestagung „10 Jahre DeGEval – Zukunft der Evaluation“, 10th and 12th October in Dresden.

Grajewski R (2007) Erfahrungen aus den Ex-ante-Evaluationen von ELER, EFRE und ESF für die Förderperiode 2007 bis 2013. Ergebnisse einer Befragung von EvaluatorInnen, presentation, 1st and 2nd of March, Braunschweig.

Grajewski R (2008) Begleitung und laufende Bewertung im Rahmen der ELER-Programme 2007 bis 2013, presentation held at the Frühjahrsworkshop des AK Strukturpolitik der DeGEval, 12th and 13th of June in Düsseldorf.

Schwab O (2007) Evaluierung der Strukturfonds in Deutschland und Österreich seit den 1990er Jahren: Wirkungen und Perspektiven. EFRE – Deutschland, presentation held at the Jahrestagung „10 Jahre DeGEval – Zukunft der Evaluation“, 10th and 12th October in Dresden.

Taylor S, J Bachtler and L Polverari (2001) Structural Fund evaluation as a programme management tool: comparative assessment and reflections on Germany, Informationen zur Raumentwicklung, Heft 6/7, 341-357.

Toepel K (2000) Evaluation in der Regionalpolitik, Vierteljahreshefte zur Wirtschaftsforschung, 69(3), 395-405.

Wollmann H (2008) Kontrolle in Politik und Verwaltung: Evaluation, Controlling und Wissensnutzung, in K Schubert, N Bandelow (Eds) Lehrbuch der Politikfeldanalyse, Oldenbourg, München/Wien, 2nd edition.

Evaluation practice in the field of structural funds in Sweden

by Dr. Kai Böhme, Spatial Foresight GmbH

1. Introduction

The approach to and terminology of evaluations used in the Structural Funds context has been new to Sweden when it joined the European Union in 1995. However, the Structural Funds evaluation culture meet in Sweden a policy environment with a long history of assessing policy interventions and using external expert knowledge in policy processes. Especially in the field of regional development there was already earlier a practice of programme evaluation prior to EU membership.

Sweden employs at national level a centuries-old system of official investigatory commissions. The deliberations of government are supplemented by these commissions in which parliamentarians and representatives of political parties, professions, experts and interest groups work out proposals before they are submitted to Parliament. The results of these commissions are documented in a series of official government publications, so called Report of the Government Commission, called *Statens offentliga utredningar* (SOU) (Böhme, 2002).

Reinert (1987) describes these widely used procedures not only as a method for achieving high-quality expert advice for policy-making, but also as a key element in the process of reaching consensus, which has an important role in Swedish decision and policy-making.

Certainly this national system is changing over time and there are some suspicions that the influence of expert knowledge and evidence has been reduced over the last decades. At the same time, also a growing influence of the evaluation approaches employed by the European Commission can be noted.

With regard to regional policy interventions there have been two reports in 2007 which characterise the current state of the debate in Sweden (NUTEK, 2007 and Bager-Sjögren et.al., 2007). Both highlight that in fields where the market is absent, there is a need for the public administration to use instruments to assess the effectiveness of its interventions. Monitoring and evaluations are considered as such instruments, despite the fact that there are no explicit or common standards in Sweden on how these monitoring and evaluation exercises are to be understood or undertaken. The main common dominator is that they assess results in relation to the aims set out when the policy-intervention has been launched.

Nutek (2007) describes the organisation of monitoring and evaluation as four major moments:

1. Planning

   This moment consists of (a) the formulation of the goals for the interventions, (b) the identification of what shall be changed by the intervention, (c) the definition of indicators to measure the change, and (d) the establishment of a monitoring and evaluation plan. For more complex interventions it includes also (e) the data collection describing the statues quo when the intervention starts, and (f) a full-fleshed ex-ante evaluation. The various steps are usually to be carried out by people involved in the managing of the intervention in question.

2. Monitoring

   Monitoring concerns the continuous documentation of the progress made over time. This concerns both the implementation and the results. Are payments made according to the plan? Have the planned activities been carried out? Are there possibilities for improvements? Etc.

3. Evaluation of effects

   When there is a need, the effects of an intervention are assessed. This done based on the monitoring information and additional assessments of effects. In many cases it appears sufficient to use one single indicator to judge the effects. In other cases more complex considerations are necessary. That is when an evaluation of the effects is carried out. The

challenge is to establish a causal and measurable link between the interventions and the effects.

4. Ex-post evaluation
   When an intervention has been finalised the last step is usually an ex-post evaluation. This can focus on the results, effects, organisation, the implementation and the administration. This ex-post evaluation is to be carried out by an independent party.

The above reflects in rough terms the ideal standard for monitoring and evaluation within Swedish regional policies. The degree to which it is carried out according to this varies and depends on the type of intervention.

The standard description has been influenced by the Structural Funds evaluation practice in the European Union.

The following discussion of evaluation practices focuses on the ERDF programmes, excluding INTERREG which follows a slightly different logic because of its cross-border and trans-national character. Also the ESF programmes have different settings.

2. Demand side

The organisation, management and implementation structures for Structural Funds have been changed – or reinvented – for every programming period in Sweden. Thus the current structures (2007-2013 period) are not the same as the ones in the previous (2000-2006) period, which where different from the first Structural Funds period in Sweden (1995-1999).

The overall actors remained however largely the same, i.e. the main actors are the Ministry of Enterprise, Energy and Communications (*Näringsdepartementet*) and NUTEK, the Swedish Agency for Economic and Regional Growth.

The Ministry of Enterprise, Energy and Communications has the overall responsibility for the design and supervision of the implementation of EU

policies. This ministry is also chairing the three Monitoring Committees for the eight regional ERDF programmes 2006-2013.

NUTEK acts as Secretariat for the Monitoring Committees for ERDF programmes and is in charge of maintaining the overview of the programmes' operations. It is furthermore responsible for training and co-ordination, reporting to the central Government, national-level publicity, evaluation and monitoring.

As regards the evaluations, largely two cases can be distinguished (a) evaluations which are mandatory according to the Structural Funds regulations, and (b) evaluations which are outside the official evaluation cycles, such as national thematic evaluations.

As concerns the mandatory evaluations NUTEK plays a central role. During the previous (2000-2006) period, NUTEK was largely responsible for drawing up the Terms of Reference for the mid-term evaluations. In principle, all evalutions have been tendered simultaneously and the same evaluator has been contracted for all them, although they have been carried out as separate evaluations following the same template. A similar procedure was followed for the up-date of the mid-term evaluation of the 2000-2006 programmes. Also the ex-ante evaluations of the 2007-2013 programmes have been handled centrally and all eight of them have been carried out by the same evaluation team. In conclusion, the mandatory evaluations have been handled centrally and although there are separate evaluations for each programme they have been carried out by the same evaluation team using the same template.

For the 2007-2013 NUTEK has elaborated an evaluation plan which has been co-ordinated with the monitoring committees. There will be on-going evaluations at three levels. The first is concerning the overall programme organisation, and is covering all eight regional ERDF programmes as well as the national ESF programme. The second is concerning programme implementation. Three contracts, one for each Monitoring Committee, will work with similar tasks but for different regions. At the third level, there will also be evaluations of the progress and results of the largest projects.

During the previous periods, there have been a series of thematic evaluations in addition to the mandatory ones. Usually, they have been commissioned by NUTEK and concerned either all ERDF programmes or groupings of them.[1]

Furthermore, other national agencies as e.g. the Swedish National Rural Development Agency (Glesbygdsverket)[2] or ITPS, the Swedish Institute for Growth Policy Studies[3], have conducted or commissioned thematic evaluation of ERDF programmes concerning their specific policy-fields. Also here, these evaluations concerned usually either all ERDF programmes or groupings of them.

Project evaluations do not have a prominent position in the field of Swedish Structural Funds evaluation and are rather rarely spotted.

3. Supply side

The central structure of the demand side is also reflected on the supply side. Organisations with a capacity to conduct a large number of evaluations simultaneously have an advantage, as there is a tendency to commission evaluations for all regional programmes to the same evaluator.

Thus, the group of potential evaluators is rather limited – in particular keeping in mind that Sweden has only 9 million inhabitants.

Most evaluations are conducted by private consultancy firms, as far as the evaluations are not conduct by national agencies, as was the case e.g. for the ex-ante evaluation of the 2006-2013 ERDF programmes and a number of thematic evaluations. University departments are more usual as sub-contractors and not so visible as lead partners for the evaluations.

---

1 Examples for this are the studies of the 2000-2006 period: (a) Lisvmiljö som attrakationskraft (b) Regionala förutsättningar för innovation, (c) Strukturfonder, entreprenörskap och företagandets villkor.

2 An example for this is the study of the 2000-2006 period: Strukturfondsprojekt med anknytning till gles- och landsbygders utveckling.

3 An example for this is the study of the 1995-1999 period: Effektutvärdering.

4. Products

The overall impression is that the number of Structural Funds evaluations is increasing from one period to the next. This regards in particular the non-mandatory thematic and on-going evaluations.

For the 1995-1999 period there was only one major thematic evaluation, i.e. the study of effects carried out by ITPS.

In the next period, 2000-2006, there have already been four national thematic evaluations, i.e. three by NUTEK[4] and one by the Swedish National Rural Development Agency (Glesbygdsverket)[5].

In the current, 2007-2013, period it appears that there will a wide range of different continuous evaluations, which largely replace mid-term evaluations and their updates which have been mandatory during the previous period.

The type of evaluations is predominantly programme evaluations carried out uniformly across all ERDF programmes in Sweden (or groupings of them). Thematically they can be targeted to specific national policies, such as entrepreneurship, innovation, living conditions or rural development.

The quality of the evaluations varies considerably, although one might argue that generally it is increasing as experience with the Structural Funds and the specific type of evaluations linked to them is growing.

The fact that often evaluations are carried out for all or many ERDF programmes simultaneously and by the same evaluator, provides a potential to cross-programme analysis and comparisons which is not fully utilised. Indeed, the advantages of this approach is mainly seen in the economy of scales, so that man-power used for the evaluations can be reduced as several general parts and the methodological approach needs only to be developed

---

4 Examples for this are the studies of the 2000-2006 period: (a) Lisvmiljö som attrakationskraft (b) Regionala förutsättningar för innovation, (c) Strukturfonder, entreprenörskap och företagandets villkor.

5 An example for this is the study of the 2000-2006 period: Strukturfondsprojekt med anknytning till gles- och landsbygders utveckling.

once and thereafter can be applied across all programmes.

The dissemination of the evaluations targets firstly those who are involved in the national Structural Funds community. Thus it can be argued that the target group of the evalutions are mainly the actors who actually can learn from the evaluations and possibly improve their activities in the future.

Furthermore, many of the official evaluations are published by NUTEK or the national agency which commissioned and/or conducted them. These publications or mostly available free of charge as pdf-files downloadable from the internet and partly also in print (in a limited edition). Usually, the official websites are re-organised when a new programming period starts, which often implies that information on the previous programming periods, incl. the access to the evaluations, is no longer easily available on the internet. In conclusion, the main evaluations of the on-going programmes are usually easily available, but are replaced by more current information as the programmes move on. Furthermore, there is no central database or publication platform for Structural Funds evaluations.

5. Evaluation utilization

The main audience of the evaluations conducted is the national Structural Funds community. In particular the Ministry of Enterprise, Energy and Communications that has the overall responsibility for the design and implementation of EU policies, and NUTEK that acts as Secretariat for the Monitoring Committees and is in charge of maintaining the overview of the programmes' operations. Certainly the Monitoring Committees and regional programme offices as well as the regional partnerships are also part of the audience.

The main purpose of the evaluations seems to be mainly on the reporting and justification of the programmes. Only to a minor degree they are used for discussions on the nature and rationality of EU interventions and for actually planning and improving the programmes. Indeed, it seems that the potential to learn and improve the programmes which is offered in many evaluations is not

fully utilised in Sweden.

The simultaneous evaluation of all or several Swedish ERDF programmes by the same evaluation team would offer a potential for learning across programmes which is not used very much. Also ambitions to look for experience beyond the national borders for learning and improving national Structural Funds practices are only rarely seen.

Thus the opportunity for critical self-reflection which is offered by the evaluations is not fully used. The evaluations do not leave any major traces in the wider debate. This might partly be explained with the Swedish administrative and policy culture where changes are made rather slowly and by consensus of a larger community. Thus it is often difficult to trace what exactly triggered a change or development in the policy context.

6. Perspectives

When Sweden joined the EU in 1995 it had a tradition of using external assessments and expert knowledge in policy making. However, this was rather different form the approach to evaluations in Structural Funds. Swedish regional policy have rather quickly adapted to the European evaluation approach. Thus the mandatory evaluations are carried out and in addition a wide range of thematic and on-going evaluations. Most evaluations cover all ERDF programmes or groupings of them – although formally they might be conducted separately for every programme. Evaluations at project level are not so usual.

So far, the opportunities to learn and improve which are provided by the evaluations have not been fully used. Often the evaluations had a main focus on reporting and justifying. In the current period, a lot of effort is put into on-going evalutions and more process oriented evaluations. This might lead to a change of how evaluations results are received and used. Thus by the end of the current, 2006-2013, period the picture might have changed to the fact that evaluations are an integrated moment of learning and self-reflection in the programme management.

Martin Ferry, PhD,
European Policies Research Center,
University of Strathclyde

Evaluation practice in the field of Structural Funds in the United Kingdom

1. Introduction[1]

The United Kingdom (UK) is generally regarded as having one of the most developed evaluation cultures in the European Union. The principles of deciding objectives, setting targets, monitoring and evaluating progress and impact have become established in UK policy-making over the past two decades [BACHTLER, 2001]. Policy-makers have invested heavily in evaluations of the efficiency and effectiveness of regional policy since the late 1970s, initiated in response to increasing criticism concerning the lack of a proper evaluation of the costs and benefits of the policies pursued. The commitment of the Government (particularly the Treasury) to evidence-based policy making emphasizes the role of evaluation: "...effective policy making must be a learning process which involves finding out from experience what works and what does not and making sure that others can learn from it too. This means that new policies must have evaluation of their effectiveness built into them from the start." [CABINET OFFICE, 2000] In particular, evaluation is needed to improve the success of policies in achieving Departmental objectives, emphasised by the introduction of Public Service Agreement Targets set by Departments in consultation with the Treasury. Evaluation is expected to play an important role both in enhancing public accountability by providing accurate information about performance and in improving decision making by highlighting 'good practice' [MARTIN AND SANDERSON, 1999].

Evaluation of UK domestic policies preceded evaluation of Structural Funds. The development of an evaluation culture was initiated independently of the Structural Funds. Indeed, as with other aspects of the Structural Funds programming process, there has been a perception amongst UK policy-makers that domestic and Structural Funds evaluations were completely separate, administered according to different principles, under the jurisdiction of different authorities and producing different types of knowledge according to different timetables. However, some policy-makers in the UK have identified distinctive elements in Structural Funds evaluation that can now be seen in domestic approaches: its

[1] This report is informed by research carried out in the United Kingdom as part of the *IQ-Net* project, managed by the European Policies Research Centre, University of Strathclyde. See *www.**eprc**.strath.ac.uk/**iqnet**/*

programmatic nature; the role of policy learning; and the approach to assessing macroeconomic effects from diverse microeconomic interventions [RAINES, 2006].

2. Demand side

Against a backdrop of devolution and decentralisation, UK government authorities have increasingly moved towards providing an overall evaluation framework or common guidelines for evaluation within which there is some flexibility for UK nations and regions to operate [see POLVERARI et al 2007]. The mid-term evaluations for 2000-06 were strongly led from national level (in England) and by the Devolved Authorities in Scotland and Wales. The Office of the Deputy Prime Minister (now the Department for Communities and Local Government) as Managing Authority for the English programmes, coordinated the contracts for the 12 English regions. ODPM set the timetable, and the evaluations were funded through the programmes' Technical Assistance, match-funded by ODPM. Again, all were carried out by independent consultants. The Scottish Executive coordinated contracts for the four Scottish programmes; and the Welsh European Funding Office (WEFO) coordinated contracts for the two Welsh regions. Terms of reference were relatively standardised across the UK, largely based on the original interpretation of the Commission's WP8 by the UK Department for Trade and Industry (DTI, now Department for Business and Regulatory Reform). Programme managers in the regions (i.e. the Devolved Authorities in Scotland and Wales and the Government Offices in the English regions) were encouraged to add local issues of relevance to the Terms of Reference.

The national level also played an important role in managing the UMTEs, with ODPM specifying both the scale and scope of the Commission requirements as well as the balance of work between the regional Government Offices and the ODPM. The regional Government Offices were responsible for the first three points of the Commission guidelines (review of MTE recommendations, analysis of outputs and results, and the analysis of impacts and likely achievement of objectives), while the ODPM carried out a 'ground-clearing' exercise entailing a comprehensive review of past evaluations to cover identified gaps. Data (for England) was aggregated and the results synthesised by the ODPM's evaluation unit. Some Government Offices undertook the evaluation work 'in-house', others hired consultants. In Scotland, the Scottish Executive coordinated and carried out the UMTEs, with input from Programme Management Executives and some tasks commissioned from consultants (e.g. a Social Inclusion Survey for the ESF UMTE). In Wales, WEFO also commissioned the UMTE to be carried out externally.

For 2007-13, the programmes in England have recieved some evaluation guidance from the national level, in the form of a CLG user manual, based on Commission guidelines. Regional Development Agencies, who have assumed responsibilities for programme management from the regional Government Offices, are expected to carry out their own ongoing evaluations. The RDA are keen to consolidate their evaluation culture, including domestic and Structural Funds aspects. Although RDAs are finalising their evaluation plans, it seems likely that there will be MTEs, (possibly around 2010) and that these will supplemented with ongoing *ad hoc* thematic evaluations e.g. on innovation, local economic growth or the implications of current domestic administrative reforms at local, sub-regional and regional levels.

In Scotland, a Monitoring and Evaluation Group (MEG) has been established. It is chaired by the Scottish Government and includes Scottish Enterprise, Highlands and Islands Enterprise, as well as Intermediary Administration Bodies. It will oversee what evaluation work will be needed (Scotland-wide), with a budget to spend on evaluation. It will draw up evaluation workplans and specifications and commission work externally. There is planned to be more focus on *ad hoc* thematic evaluations than in the previous period. In Wales, the monitoring and evaluation plan was presented to the PMC in September 2007 in draft format. Building on experience under the 2000-06 programmes, WEFO decided to cover both Convergence and Competitiveness programmes under the plan.

3. Supply side

In keeping with the United Kingdom's well-established and expanding evaluation culture, there is a wide range of evaluators. Large multi-national consultancies (such as ECORYS, ECOTEC, PIEDA etc) are well represented. However, smaller, specialist agencies (e.g. Fraser Associates) are also very active. Beyond consultancies, research institutes (such as EPRC) are prominent in the field of Structural Funds evaluation. As the UK evaluation field expands and becomes more sophisticated, there are increasing opportunities for new actors to become established in the field.

4. Products

Over programming periods, evaluations have become more frequent and more sophisticated as there has been a move towards more strategic and iterative approaches to the evaluation of European (and domestic) regional policy in the United Kingdom. As well as ex-ante, mid-term and ex-post evaluations, thematic and (on-going or 'rolling' studies) have been part of the evaluation approach in

parts of the UK for several years. From the late 1990s, Scottish programming authorities were some of the first to create an evaluation strategy for Structural Funds. This included 'rolling evaluation', complemented by thematic evaluations in areas such as sustainable development, gender mainstreaming and RTDI, as well as investment in methodological research, and the production of guidance on evaluation (EPRC and FRASER OF ALLANDER, 2000. North East England was a strong advocate of the ongoing evaluation approach during the 2000-06 period; after the MTEs they kept two consultancy firms on a retainer/stand-by to carry out *ad hoc* studies when required. They studied emerging issues as a rolling evaluation process. It is, therefore, not surprising that the use of ongoing evaluations instead of mid-terms is generally supported in the UK. The approach is considered to introduce a degree of flexibility as it was felt that the last mid-terms had taken place too early in the 2000-06 programming period. With increasing flexibility in Commission regulations for 2007-13 it is very likely that ongoing evaluations will be prominent.

In terms of focus, some basic trends can be identified. First, there is an increasing emphasis on Lisbon-type themes. This matches the increasing focus on research and development, innovation and entrepreneurship in many English programmes (e.g. in North East England, the RDA is considering evaluation of innovation connectors as part of its ongoing, thematic evaluation approach). Second, there is increasing focus on how to evaluate less tangible or 'softer' programme impacts (e.g. concerning self-confidence or self-esteem in employability-related measures). Third, there is increasing focus on evaluation at project level. Again this is related to the decision of several UK programmes to focus on the implementation of larger, more strategic projects. For instance, the evaluation strategy developed by WEFO requires that an evaluation plan must be developed as part of each project application's business case (WEFO provides advice and guidance on this). External evaluation will be required for all projects receiving £2m grant or more, or for projects implementing innovative or experimental actions.

Evaluation results are disseminated through a variety of means: bulletins; published papers; meetings, seminars, conferences and workshops and the monitoring and evaluation pages of the programming authorities website. Some innovative approaches to disseminating the knowledge produced by evaluation have emerged in the UK. For instance, in the Highlands and Islands Special Transitional programme 2000-6 MTE, a series of thematic Working Papers were produced throughout the course of the evaluation. This meant that the results of each research task could be disseminated immediately, without having to wait for the Draft Final or Final Report to be prepared. The Working Papers approach was also a dynamic one, allowing evaluators to engage in a dialogue with interested parties and stakeholders through dissemination events organised as each Paper became available. There is no central database that collects UK Structural Funds evaluations.

5. Evaluation utilization

The main audiences for the Structural Funds evaluations are the Commission, programme managers, stakeholders and beneficiaries. Different evaluation reports are utilised in different ways:

- Planning new interventions

In recent years, evaluation has become more readily accepted as a learning tool, to assist preparations for the next programme period. Such policy learning has been prominent, for example, in Scottish approaches to evaluation. In the mid-term updates of the current programmes, extensive use was made in Scotland of the learning gained by project sponsors and key programme partners through a number of stakeholder events at which assessments of programme performance were discussed. There has also been learning in terms of how to undertake Structural Funds evaluations, for example, through the UK-wide Structural Funds Evaluation Group, which has acted as a forum for discussing the challenges of Structural Funds evaluation and innovative ways of meeting those challenges. [RAINES, 2006]

- Operational management (problem solving)

Operational evaluations will be a particular priority for several programmes in the UK where implementation arrangements have changed significantly in the 2007-13 period. Delivery mechanisms in Scotland have changed from the previous period. In response to substantially reduced levels of Structural Funds, the management and implementation system has been rationalised. The new arrangements will be reviewed in 2010. The implementation of major strategic projects will also be evaluated. In Wales, the implementation process has also changed (e.g. the reorientation of divisions within WEFO and the introduction of a new computer system are two major process-oriented changes). Process issues will be examined from a top-down perspective only.

- Wider discussion on the nature and rationality of the EU interventions

The UK is often associated with those Member States seeking a significant reorientation of Cohesion policy. It argues that EU-funded regional policy should finance only the poorest regions in the newer Member States of Central and Eastern Europe (EU10) and that the more prosperous Member States, including the UK, should renationalise regional policy and funding. Critical to the UK's approach to the EU debate is a clear understanding of the value added by the Structural Funds in terms of quantitative impacts and more widely with respect to the design, management and delivery of economic development policies. Thus, 'Community added value' in evaluation of the impact and added value of enterprise and regeneration activity funded by Structural Funds in the UK has been increasingly emphasised. This applies to evaluations of individual

programmes but also to broader, thematic evaluations. For instance, a report in 2003 conducted on behalf of the UK government evaluated of the added value and costs of the Structural Funds in the UK [ECOTEC, 2003].

- Reporting & verifying effects

As is the case in other Member States, impact assessment has been a challenging area for Structural Funds evaluations in the United Kingdom. Methodological issues and the timing of evaluations mean that Structural Funds evaluations are much stronger in assessing programme processes than effects. Some UK programme managers have noted a focus in Commission evaluation requirements on the physical and financial progress and performance of programmes and management and implementation processes , rather than on effects. Moreover, in the UK Structural Funds programmes fund interventions in a wide range of policy fields, involving different types of stakeholders. Evaluating the impact of all of these activities is very challenging. Many of these interventions involve 'softer' impacts that are challenging to measure. For instance, increasing employability as part of labour market interventions may involve 'improving self-esteem', which is less conducive to evaluation of effects than 'jobs created' or 'qualifications gained'.

Another challenge relates to timing issues. Evaluations have encountered significant difficulties in responding to Commission requirements to meaure progress at specific times in the programme life-cycle. This often involves identifying the impacts of programmes over timescales which are not long enough to produce robust conclusions on the effects of policy actions [EPRC and FRASER OF ALLANDER, 2000]. Up to the current programming period, the Commission model of ex-ante, mid-term and ex-post evaluations, coupled in many cases with delays in launcing programmes, have made it difficult to make evaluation results available at the appropriate stage of the programming cycle. Ex post evaluations have tended to come too late to inform the next generation of programmes, and mid-term evaluations come too early for there to be any real progress to measure.

- Any trace of clear use & influence of the reports (on public discussions, decision-making, policies)?

As noted above, policy learning has been prominent in UK approaches to evaluation. In the mid-term updates of the 2000-6 programmes, extensive use was made of learning gained in the preparation of the 2007-13 programmes . For this purpose, stakeholder events were often organised at which assessments of programme performance were discussed. In Wales, both the UMTE for 2000-

06 and ex ante evaluation for 2007-13 had a strong focus on carrying forward lessons learned into the new programmes. Programme Monitoring Committees were asked to provide evidence-based opinions on what had worked well in the 2000-6 period and why. WEFO also organised a research projwect on measuring 'soft outcomes' (e.g. improved motivation or self-esteem) in the 2000-6 objective 1 and 3 programmes. Based on this, the ex-ante evaluation of the Welsh programmes for 2007-13 included a report on lessons learned in an annex. This process informed some important changes in the 2007-13 programmes, notably the shift towards larger, more strategic projects under the Welsh Strategic Frameworks model.

6. Perspectives

As noted above, the UK has a well-established evaluation culture that encompasses domestic interventions and Structural Funds programmes. Drivers for increasing this include: commitment to 'evidence-based policy-making, more sophisticated appreciation of the 'regional challenge' and processes of devolution and decentralisation. As a result, evaluation has developed a more iterative and strategic approach. In the UK, evaluation is embedded as an integral part of the policy-making process. At the same time, this evolution presents significant challenges, particularly for Structural Funds evaluations. How do you measure subtle, complex or intangible policy impacts? How do you encourage specific evaluation approaches within devolved or decentralised bodies while maintaining an overarching perspective? In an era of 'joined-up' policy making, how can the individual contribution to regional development of a variety of institutions and initiatives from various levels and policy fields be individually identified and evaluated? [MCVITTIE and SWALES, 2004] How can greater alignment be achieved between evaluation of domestic actions and Structural Funds interventions? New Structural Funds implementation arrangements for 2007-13 include greater flexibility for programme managers to develop evaluation strategies that suit thein own purposes (in terms of timing and theme) and this should help to adress some of these issues.

References

BACHTLER, J. (2001) 'Quod erat Demonstrandum? The Evaluation of Regional Policy' in Eckey HF, Hecht D, Junkernheinrich M, Karl H, Werbeck N and Wink R (Editors) *Ordnungspolitik als konstruktive Antwort auf wirtschaftspolitische Herausfoerderungen - Festschrift zum 65. Geburtstag con Paul Klemmer,* Lucius & Lucius Verlag, Stuttgart.

CABINET OFFICE (2000) *Professional Policy Making in the 21st Century*

ECOTEC (2003) *Evaluation of the Added Value and Costs of the European Structural Funds in the UK*, Final Report to the Department of Trade and Industry (DTI) and the Office of the Deputy Prime Minister (ODPM).

EPRC & THE FRASER OF ALLANDER (2000) *Methodologies used in the Evaluation of the Effectiveness of European Structural Funds: A Comparative Assessment,* Report to the Scottish Executive, European Policies Research Centre, University of Strathclyde, Glasgow.

MARTIN, S. AND SANDERSON, I. (1999) 'Evaluating public policy experiments: measuring outcomes, monitoring progress or managing pilots?' *Evaluation*, Vol. 5, No. 3, 245-58

MCVITTIE, E. AND SWALES, J.K. (2003) 'Constrained Discretion in UK Monetary and Regional Policy' paper presented at the *Regional Policy in the UK* seminar, LSE, November 2003, organised by the ESRC Urban and Regional Economics Seminar Group.

POLVERARI, L. et al (2007) Making Sense of European Cohesion Policy: 2007-13 On-Going Evaluation And Monitoring Arrangements *IQ-Net Thematic Paper No. 21(2)*

RAINES, P. (2006) 'The 'Trojan Horse' effect and the evaluation of SF', *Regional Studies* Volume 40, Number 02 / April 2006 pp 285-288.

The central organisation of the evaluation work by NUTEK with influence from the Ministry will most likely remain. So will also the fact that the Structural Funds community and the number of potential evaluators is rather limited, which does not come as a surprise considering the limited size of the country and in particular the limited amount of Structural Funds money allocated to Sweden.

References

Bager-Sjögren, L., F. Calidoni, H. Gadd & A.M.M. S. Quoreshi (2007) Utvärdera effekter av offentligt stöd till näringslivet. Stockholm: ITPS.

Böhme, K. (2002) Nordic Echoes of European Spatial Planning. Stockholm: Nordregio.

Moxens Steineke, J. (2007) Top-down or bottom up? Methods for evaluating empolyment effects in Nordic Structural Funds programmes. In: K. Lähteenmäki-Smith (ed.) Learning through evaluations: the Nordic experience. Stockholm: Nordregio.

NUTEK (2007) NYTTA – Nuteks system för uppföljning och utvärdering. Stockholm: Nutek.

Polverari, L., J. Bachtler & R. Michie (2003) Taking Stock of Structural Fund Implementation. Current Challenges and Future Opportunities. IQ-Net.

Reinert, A. (1987) Organisierte Formen gesellschaftspolitischer Aktivierung als Problem der Sozialdemokratie in Schweden und der Bundesrepublik Deutschland. Wuppertal.

Plus various Swedish ERDF evaluation reports.