



Marianna Zieleńska 

Wydział Socjologii UW

Daniel Płatek 

Instytut Studiów Politycznych PAN

Patryk Hubar-Kołodziejczyk 

Instytut Badań Literacki PAN

Agnieszka Karlińska 

NASK PIB

WYZWANIA I POTENCJAŁ ANALIZ CYFROWYCH DUŻYCH ZBIORÓW DANYCH TEKSTOWYCH W BADANIACH SPOŁECZNYCH NA PRZYKŁADZIE PRZYGOTOWANIA KORPUSU DOKUMENTÓW UNIJNYCH DOTYCZĄCYCH UBÓSTWA I WYKLUCZENIA SPOŁECZNEGO Z LAT 2001–2021

Artykuł analizuje potencjał i wyzwania związane z wykorzystaniem cyfrowych metod analizy dużych zbiorów tekstów w naukach społecznych na przykładzie badania unijnej polityki przeciwdziałania ubóstwu i wykluczeniu społecznemu w latach 2001–2021. Autorzy przedstawiają proces tworzenia korpusu dokumentów UE – od pozyskiwania danych z baz EUR-Lex, przez ich przetwarzanie, wzbogacanie metadanych i deduplikację, po analizę sieciową. Szczególną uwagę poświęcono wskaźnikom jako elementom infrastruktury kalkulacyjnej UE, które nie tylko opisują rzeczywistość społeczną, lecz także ją współtworzą, wpływając na definiowanie problemów i formułowanie interwencji politycznych. Analiza i wizualizacja sieci ujawniają zmiany w relacjach między wskaźnikami, tematami i instytucjami, trudne do uchwycenia tradycyjnymi metodami, a ich interpretacja została oparta na przeprowadzonych wywiadach i literaturze przedmiotu. Wyniki wskazują, że kluczową rolę w polityce UE odgrywa wskaźnik zagrożenia ubóstwem relatywnym

Marianna Zieleńska, Wydział Socjologii UW, mzielenska@uw.edu.pl, ORCID 0000-0001-6297-2671; Daniel Płatek, Instytut Studiów Politycznych PAN, dplatek21@gmail.com, ORCID 0000-0002-8253-3622; Patryk Hubar-Kołodziejczyk, Instytut Badań Literacki PAN, patryk.hubar@ibl.waw.pl, ORCID 0000-0001-5582-2042; Agnieszka Karlińska, NASK PIB, agnieszka.karlińska@nask.pl, ORCID 0000-0002-4846-7086. Tekst opublikowany na warunkach licencji Creative Commons Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska (CC BY-NC-ND 3.0 PL).

Artykuł powstał dzięki finansowaniu z dwóch źródeł: grantu „Stworzenie korpusu i ekstrakcja terminologii do analizy sieciowej i modelowania tematycznego w obszarze polityki walki z ubóstwem w UE”, finansowanego w ramach IDUB UW nr PSP-501-D135-20-0004316, oraz grantu „Tworzenie polityki walki z ubóstwem w Unii Europejskiej za pomocą miar ilościowych – analiza przypadku złożonego wskaźnika zagrożenia ubóstwem i wykluczeniem społecznym”, finansowanego w ramach NCN OPUS nr UMO-2019/33/B/HS5/01955.

(AROP), podczas gdy nowszy wskaźnik AROPE ma ograniczoną dyfuzję poza administracją unijną. Jednocześnie widoczna jest rosnąca rola agencji, instytutów badawczych, firm konsultingowych i środowisk akademickich, co świadczy o eksternalizacji ekspertyzy. Wprowadzanie nowych wskaźników sprzyja rozszerzaniu wpływu UE na kolejne obszary polityki społecznej. Autorzy proponują model pracy badawczej zgodny z zasadami FAIR i Linked Open Data, podkreślając znaczenie integracji metod cyfrowych z analizą jakościową oraz współpracy interdyscyplinarnej w badaniach procesów politycznych.

Słowa kluczowe: analiza sieci; korpus dokumentów UE; FAIR; Link Open Data; rządzenie przez wskaźniki; unijna polityka walki z ubóstwem

Challenges and Promises of Large Textual Data Sets' Digital Analysis in Social Research: The Case of Preparing a Corpus of EU Documents on Poverty and Social Exclusion from 2001 to 2021

The article explores the challenges promises of using digital methods to analyse large text datasets in the social sciences, focusing on the EU policies addressing poverty and social exclusion between 2001 and 2021. It outlines the creation of a corpus of EU documents, from acquiring data via EUR-Lex to processing, enriching metadata, deduplication and network analysis. Indicators are seen as elements of the EU's calculative infrastructure, which not only describe social reality but also shape it by framing problems and guiding policy interventions. Network analysis and visualisation reveal shifts in the relationships between indicators, topics and institutions that are hard to capture with traditional methods, while interpretation draws on the interviews and literature review. Findings show that at-risk-of-poverty rate (AROP) indicator is central to EU policy, while the newer indicator (AROE) is less present beyond administrative structures. The growing involvement of agencies, research institutes, consultancy firms and academia reflects the externalisation of expertise. Introducing new indicators extends the EU's influence into broader areas of social policy. The authors propose a research workflow aligned with the FAIR and Linked Open Data principles, highlighting the value of combining digital and qualitative methods and fostering interdisciplinary collaboration to better understand political processes.

Key words: network analysis; EU document corpus; FAIR; Link Open Data; governance by indicators; EU anti-poverty policy

Wprowadzenie

Jeszcze do niedawna zautomatyzowane metody analizy tekstu były wykorzystywane w naukach społecznych tylko w niewielkim stopniu. Większość programów komputerowych wspomagających jakościową analizę tekstu (ang. *Computer Assisted Qualitative Data Analysis Software*, CAQDAS) powielala manualne procedury badawcze. Istotne ograniczenie stanowił brak narzędzi do efektywnej pracy z danymi tekstowymi na dużą skalę (Grimmer i in., 2022).

W ostatnich latach nastąpił przełom w tym zakresie związany z tak zwanym zwrotem obliczeniowym (ang. *computational turn*). Doprowadził on do zmiany zarówno „sposobu obcowania z tekstem, który nie przypomina już linearnej, sekwencyjnej lektury”, jak i „sposobów jego analizowania i wizualizowania” (Szpunar, 2016, s. 359). Cyfrowe metody analizy tekstu pozwalają na spojrzenie „z dystansu” (ang. *distant reading*) – a więc poszukiwanie w dużych zbiorach ogólnych wzorców, trendów, relacji – i szczególnie dobrze sprawdzają się w zadaniach trudnych do wykonania manualnie, takich jak grupowanie dokumentów lub przypisywanie ich do określonych kategorii. Chociaż „komputery przewyższają człowieka, jeśli chodzi o systematyczną identyfikację struktur w dużych zbiorach danych” (Wiedemann, 2015, s. 23), wskazuje się, że użycie oprogramowania komputerowego nie eliminuje potrzeby analizy dokonanej przez człowieka i uważnej lektury (ang. *close reading*) tekstów, ale tę analizę uzupełnia. Grimmer, Roberts i Stewart (2022) zaznaczają, że aby wykorzystać potencjał wielu podejść i modeli statystycznych, niezbędna jest pogłębiona praca interpretacyjna, a połączenie dużych zbiorów danych, algorytmów i wiedzy merytorycznej prowadzi do bogatszych spostrzeżeń teoretycznych.

Cel artykułu stanowi przedstawienie potencjału wykorzystania cyfrowych metod analizy dużych zbiorów danych tekstowych w naukach społecznych oraz omówienie wyzwań pojawiających się na różnych etapach pracy z tego rodzaju danymi – od ich pozyskania, przez przetwarzanie i wzbogacanie, aż po analizę. Na przykładzie przygotowania i opracowania korpusu dokumentów unijnych dotyczących ubóstwa i wykluczenia społecznego z lat 2001–2021 pokazujemy, jak szeroki i złożony jest zakres problemów, z którymi muszą mierzyć się badacze i badaczki pracujący w nurcie socjologii cyfrowej.

Ilościowe badania polityki legislacyjnej Unii Europejskiej, oparte na dużych zbiorach danych pozyskanych z baz internetowych, są prowadzone od ponad dwóch dekad. Podejmowano w nich zróżnicowane zagadnienia: od efektywności i wyników procesu decyzyjnego (Golub, 1999, 2007; König, 2007; Häge, 2011), przez podział pracy w strukturach UE (Häge, 2007, 2008), po skutki rozszerzenia UE i roli reform instytucjonalnych (Toshkov, 2017; Rauh, 2021). Przedmiotem naszej analizy są długookresowe zmiany, które zaszły w infrastrukturze kalkulacyjnej unijnej polityki walki z ubóstwem w latach 2001–2021, a więc w zestawie wskaźników, procedur pomiaru oraz sieci aktorów zaangażowanych w definiowanie, monitorowanie i legitymizowanie politycznych priorytetów w obszarze ubóstwa. Szczególną uwagę zwracamy na rolę wskaźników – zarówno jako narzędzi diagnozy, jak i instrumentów wpływu instytucjonalnego – w strukturze zarządzania polityką społeczną UE. Analizujemy, w jaki sposób ich rozwój i zmiany konfiguracji aktorów zaangażowanych w ich produkcję oraz upowszechnianie odzwierciedlają głębsze procesy instytucjonalizacji polityki walki z ubóstwem.

W swoich badaniach wykorzystujemy dobrze już ugruntowane w naukach społecznych metody wywodzące się z analizy sieciowej, odzwierciedlające powiązania (krawędzie) między słowami kluczowymi a organizacjami posługującymi się nimi w swoich dokumentach. Aby lepiej zrozumieć uzyskane w ten sposób wyniki i pogłębić ich interpretację, osadzamy tę analizę w kontekście wcześniejszego, opartego na wywiadach badania jakościowego dotyczącego roli wskaźników w kształtowaniu polityki walki z ubóstwem. Jednocześnie pokazujemy, w jaki sposób analiza sieci może rozszerzyć wnioski z badań jakościowych dzięki ukazaniu ogólniejszych wzorców i umieszczaniu ich w dłuższej perspektywie czasowej. Naszym istotnym wkładem jest zaprezentowanie, w jaki sposób gromadzić, opracować i następnie wzbogacić nieustrukturyzowane dane tekstowe, żeby prowadzenie tego rodzaju analiz było możliwe, oraz do jakich nowych ustaleń można dzięki takiemu podejściu dojść. Wskazujemy także na potencjał łączenia spojrzenia „z dystansu” z pogłębionymi analizami jakościowymi. W ten sposób odpowiadamy na kluczowe wyzwania socjologii cyfrowej – takie jak integracja danych tekstowych z metodami analizy sieciowej, problemy reprezentatywności, a także trudności interpretacyjne związane z różnorodnością i kontekstualnością źródeł cyfrowych (Salganik, 2018; Veltri, 2020).

Realizacja tego rodzaju badań wymaga współpracy interdyscyplinarnej – łączącej kompetencje z zakresu językoznawstwa korpusowego oraz informatyki, a także pogłębioną znajomość badanych obszarów tematycznych. Takie podejście umożliwia nie tylko techniczne opracowanie dużych zbiorów danych, lecz także ich interpretację z uwzględnieniem złożonych kontekstów instytucjonalnych i społecznych, a w efekcie pełniejsze wykorzystanie potencjału metod cyfrowych w analizie procesów politycznych i społecznych.

Proponujemy modelowy przepływ pracy pozwalający na wykorzystanie zbiorów dokumentów UE i innych podobnych danych na potrzeby badań społecznych oraz gwarantujący możliwość ich ponownego użycia. Przyjęte podejście zakłada wykorzystanie mechanizmów Linked Open Data oraz zasad FAIR na wszystkich etapach – od analizy i wzbogacania danych, aż po udostępnianie wyników badania. W szczególności opieramy się na modelu 5-star Open Data zaproponowanym przez Tima Bernersa-Lee, który opisuje kolejne poziomy otwartości danych i ich gotowości do ponownego wykorzystania. Każdy kolejny poziom zakłada spełnienie warunków poprzednich:

- 1) udostępnienie danych w Internecie (w jakiegokolwiek formie) z licencją otwartą;
- 2) wykorzystanie formatu nadającego się do odczytu maszynowego (np. Microsoft Excel zamiast PDF);
- 3) zastosowanie formatu otwartego (np. pliki w formacie CSV zamiast Microsoft Excel);

- 4) udostępnienie danych w formatach zgodnych z założeniami Semantic Web;
- 5) powiązanie danych z innymi zbiorami dostępnymi w Internecie (Janowicz i in., 2014).

Równolegle realizujemy założenia FAIR Data Principles, według których dane powinny być:

- 1) możliwe do odnalezienia w Internecie poprzez jednoznaczne identyfikatory i bogate metadane;
- 2) dostępne za pomocą standardowych protokołów komunikacyjnych;
- 3) zgodne z odpowiednimi standardami i formatami, umożliwiającymi ich łączenie z innymi zbiorami;
- 4) udostępnione w sposób umożliwiający ich ponowne użycie w różnych kontekstach badawczych, z jednoznacznymi informacjami o licencjonowaniu i źródłach (Wilkinson i in., 2016).

Zbiory dokumentów UE

Pierwszym krokiem do realizacji badań przy użyciu metod cyfrowych jest pozyskanie lub utworzenie korpusu, a więc obszernego i możliwie reprezentatywnego dla danego języka lub jego odmiany zbioru tekstów zapisanych w formie elektronicznej (McEnery i Wilson, 2001; McEnery i in. 2006). W związku z brakiem publicznie dostępnych korpusów dokumentów EU dostosowanych do potrzeb środowiska naukowego dane gromadzone były przez poszczególnych badaczy lub zespoły badawcze na potrzeby konkretnych analiz. Jak słusznie zauważa Häge (2011), zakres zastosowania tak tworzonych zbiorów był więc bardzo ograniczony, a sam proces zbierania danych mało efektywny, ponieważ kolejni autorzy powielali te same kroki. Dlatego od połowy pierwszej dekady XXI wieku zaczęto tworzyć duże, szeroko dostępne zbiory danych umożliwiające badanie procesu podejmowania decyzji legislacyjnych, zmian w polityce UE i integracji europejskiej. Dane pozyskiwano najczęściej z baz CELEX, PreLex oraz EUR-Lex. Pierwsza z nich, prowadzona przez Urząd Publikacji UE od 1970 roku, rejestrowała wszystkie wydarzenia legislacyjne UE; druga, utworzona pod koniec lat 90. i podlegająca Komisji Europejskiej, służyła do monitorowania procesu przepływu dokumentów przedkładanych przez Komisję innym instytucjom unijnym od momentu złożenia wniosku aż po ostateczne przyjęcie aktu prawnego; trzecia natomiast, uruchomiona w 1998 roku i gruntownie zmodernizowana w 2004 roku, gromadzi dane od 1958 roku (Düro, 2009). Obecnie funkcjonuje ona jako centralna platforma udostępniająca dokumenty tworzone przez instytucje, agencje i organy UE, a także inne podmioty współpracujące, w tym instytucje rządowe i organizacje komercyjne działające w ramach

struktur unijnych. EUR-Lex umożliwia dostęp do publikacji, danych, wyników badań, ogłoszeń o zamówieniach publicznych oraz innych oficjalnych zasobów informacyjnych (Publications Office of the EU, 2024). Co istotne, obejmuje ona nie tylko dane legislacyjne i dokumenty prawne, lecz także narzędzia do ich przeszukiwania i przeglądania. CELEX i PreLex zostały zintegrowane z EUR-Lex: dane CELEX są dostępne za pośrednictwem linku do informacji o dokumencie, a większość danych PreLex za pośrednictwem linku do procedury (Blom-Hansen, 2019).

Pierwszym przedsięwzięciem zmierzającym do stworzenia zbioru danych UE dostosowanego do potrzeb naukowych był EU-Lex (König i in., 2006a, 2006b). Zawiera on połączone zasoby dokumentów legislacyjnych z lat 1984–2003 dostępne w bazach danych CELEX i PreLex, w szczególności decyzje, rozporządzenia i dyrektywy. Zbiór obejmuje 47 zmiennych – w pełni zakodowanych i opisanych – odnoszących się do kluczowych cech procesu podejmowania decyzji legislacyjnych (np. zastosowana procedura legislacyjna, reguły głosowania czy specyfika sektora), które mogą być wykorzystane w analizie statystycznej. Kolejnym wartym odnotowania zasobem jest European Union Policy-Making (EUPOL). Celem jego autorów było rozszerzenie zakresu możliwych analiz i zarazem zwiększenie ich replikowalności (Häge, 2011). Pierwsza wersja zawierała ponad trzydzieści typów dokumentów z lat 1975–2009 – zarówno legislacyjnych (np. dyrektywy, rozporządzenia czy decyzje), jak i nielegislacyjnych (np. dokumenty robocze, komunikaty czy raporty) – dostępnych w bazie danych PreLex wraz z opisującymi je 2600 zmiennymi dotyczącymi wydarzeń decyzyjnych oraz zaangażowanych podmiotów, a także odsyłaczami do dokumentów dostępnych w innych internetowych bazach danych. Mocną stroną EUPOL jest możliwość odtworzenia – dzięki obszernej dokumentacji – procesu gromadzenia danych oraz samodzielnego uzupełnienia zbioru o dane najnowsze za pomocą zautomatyzowanych procedur ekstrakcji informacji. Jego istotną wadą jest natomiast to, że obejmuje on surowe dane z PreLex, które wymagają dodatkowego przetworzenia. Najnowsza, piąta już wersja EUPOL (2022) zawiera komplet informacji dostępnych w bazie PreLex do 2014 roku, zapisanych w znormalizowanym formacie nadającym się do odczytu maszynowego. Z kolei na potrzeby zbioru EULO Hertz i Leuffen (2011) również pozyskali dane z PreLex, ale dodali informacje o CELEX pozyskane z EUR-Lex.

Trzy wymienione bazy mają istotne luki i ograniczenia, które zostały szeroko opisane przez Goluba (2023). Przede wszystkim są one niekompletne, zwłaszcza w zakresie dokumentów sprzed 1994 roku (Blom-Hansen, 2019). Ważny problem stanowią również braki w metadanych, w szczególności braki identyfikatorów, podstaw prawnych i dat, oraz nad- i niedoreprezentowanie – dokumenty są często błędnie klasyfikowane, a wiele pozycji jest połączonych

w jedną. Co więcej, informacje na temat dokumentów w EUR-Lex są niespójne z informacjami zawartymi w PreLex, nawet w zakresie metadanych tak podstawowych, jak tytuły czy daty. Problemy te dziedziczą wszystkie zbiory danych opracowane na bazie CELEX, PerLex i EUR-Lex. Rozbieżności między trzema źródłami próbował zminimalizować między innymi Rauh (2021) poprzez połączenie dokumentów pozyskanych z EUR-Lex z informacjami proceduralnymi, jednak udało mu się rozwiązać jedynie część wskazanych problemów.

W odpowiedzi na te ograniczenia Golub (2023) opracował zbiór danych EUPROPS, który znacząco zwiększa dostępność i kompletność informacji o propozycjach legislacyjnych UE z lat 1958–2021. Autor ręcznie przeanalizował archiwa Komisji i Rady oraz informacje z bazy EUR-Lex, uzupełniając brakujące dane i korygując błędy w metadanych. EUPROPS uwzględnia wnioski złożone przed ich oficjalną prezentacją oraz inne dokumenty legislacyjne pomijane przez poprzednie bazy danych. Pozwala na prześledzenie losów 96,7% z 29 964 propozycji legislacyjnych ze wskazanego okresu.

EUPROPS to obecnie najbardziej kompletne źródło danych dla badań nad procesami legislacyjnymi UE, jednak nadal nie jest źródłem optymalnym z perspektywy nauk społecznych. Przede wszystkim koncentruje się na oficjalnych dokumentach legislacyjnych, takich jak dyrektywy i rozporządzenia, pomijając raporty, dokumenty promocyjne czy artykuły naukowe, które dyrektywom i rozporządzeniom towarzyszą. W swoim projekcie, realizowanym równocześnie z EUPROPS, przyjęliśmy inny model udostępniania danych tekstowych UE (metadokumentów), zgodny z aktualnymi standardami przetwarzania i udostępniania danych i opierający się na zasadach FAIR oraz Linked Open Data (Karlińska i in., 2024). Opracowany przez nas zasób został dostosowany do potrzeb badaczy i badaczek, zwłaszcza z obszaru nauk społecznych, i oferuje możliwie szeroki zestaw metadanych.

Korpus dokumentów unijnych dotyczących ubóstwa i wykluczenia społecznego

Proces budowy korpusu rozpoczęliśmy od przeprowadzenia kwerendy w bazie EUR-Lex. Zdecydowaliśmy, że będzie obejmowała teksty od początku wdrażania strategii lizbońskiej, a dokładniej uzgodnienia na szczycie w Laeken w 2001 roku zestawu wskaźników mających służyć do regularnego monitorowania zagrożenia ubóstwem i wykluczeniem społecznym w UE. Uzgodniliśmy, że zbieramy teksty do momentu rozpoczęcia budowy korpusu, a więc do lipca 2022 roku. W celu pozyskania jak najszerszej reprezentacji tekstów i stworzenia możliwie najbardziej ogólnego korpusu, do wyszukiwania wykorzystaliśmy słowo kluczowe *poverty*, które jest oficjalnym terminem ze słownika EuroVoc

i którym oznaczane są wszystkie dokumenty dotyczące ubóstwa. Wyniki wyszukiwania zostały na wstępnym etapie zawężone na podstawie analizy dostępnych metadanych, w szczególności poprzez uwzględnienie takich informacji jak słowa kluczowe i przypisane tematy, a także rok publikacji. Po wstępnym zawężeniu wyników otrzymaliśmy 32 761 rekordów zawierających pełne teksty publikacji.

Metadane publikacji zostały pobrane automatycznie za pomocą skryptu napisanego w języku Python, który łączył techniki web scrapingu z wykorzystaniem otwartych interfejsów programistycznych udostępnianych przez serwis Urzędu Publikacji Unii Europejskiej. Skrypt ten umożliwił pobranie ustrukturyzowanych danych bibliograficznych, obejmujących między innymi tytuł publikacji, autorów, słowa kluczowe oraz odnośniki do pełnych wersji dokumentów. Na podstawie pozyskanych metadanych zidentyfikowaliśmy unikatowe linki do dokumentów w formatach PDF i HTML, a następnie automatycznie pobraliśmy pełne teksty dokumentów poprzez bezpośrednie połączenie z bazą danych serwisu. W kolejnym kroku dane bibliograficzne zintegrowaliśmy z odpowiadającymi im plikami tekstowymi, co pozwoliło na utworzenie spójnej bazy, w której każdej publikacji przypisano komplet powiązanych informacji.

Jednym z kluczowych etapów przygotowania korpusu było wzbogacenie metadanych dokumentów. Szczególną uwagę poświęciliśmy słowom kluczowym i tematowi, które stanowić mogą podstawę późniejszych analiz treści czy klasyfikacji dokumentów. Głównymi źródłami informacji wykorzystanymi w tym procesie były baza Wikidata oraz tezaurus EuroVoc – oba zasoby zostały dobrane ze względu na otwarty charakter, zgodność ze standardami LOD oraz możliwość integracji z danymi za pomocą otwartych interfejsów programistycznych. Wikidata to wielojęzyczna, otwarta baza wiedzy rozwijana przez Wikimedia Foundation, która stanowi obecnie jeden z największych grafów wiedzy na świecie (Diefenbach i in., 2021), obejmując miliony pojęć z niemal każdej dziedziny wiedzy. Zakres przechowywanych danych jest praktycznie nieograniczony i stale rozszerzany, zgodnie z zasadami crowdsourcingu (Waagmeester i in., 2020). EuroVoc to z kolei wielojęzyczny tezaurus tematyczny opracowany przez Urząd Publikacji Unii Europejskiej, wykorzystywany do opisu dokumentów instytucji UE. Jego struktura opiera się na hierarchicznej organizacji pojęć – każde hasło posiada przypisane terminy nadrzędne, podrzędne oraz hasła powiązane, co czyni go użytecznym narzędziem w porządkowaniu treści i grupowaniu dokumentów według tematów.

Na potrzeby wzbogacania danych opracowaliśmy skrypt w języku Python, który wykorzystując zapytania SPARQL, automatycznie pobierał dane z Wikidaty dla wykrytych słów kluczowych. Dane te obejmowały między innymi etykiety w różnych językach, pojęcia powiązane oraz identyfikatory do zewnętrznych serwisów (np. Library of Congress, OpenAlex, VIAF). W przypadku

EuroVoc rozpoczęliśmy od identyfikatorów dla wykrytych haseł tematycznych, a następnie – na podstawie relacji zawartych w tezaurusie – do każdego z terminów w oryginalnym zbiorze metadanych dodaliśmy pojęcia nadrzędne i podrzędne. Dzięki temu możliwe było przypisanie dokumentów nie tylko do konkretnych haseł, ale również do bardziej ogólnych kategorii tematycznych.

Na etapie przygotowania danych tekstowych do analizy zweryfikowaliśmy formaty pozyskanych dokumentów, oddzielając pliki PDF od publikacji zapisanych w innych formatach (np. HTML). Dla plików PDF przeprowadziliśmy analizę dostępności warstwy tekstowej. Następnie zastosowaliśmy oprogramowanie pozwalające na precyzyjne optyczne rozpoznawanie tekstu (OCR) oraz optyczne rozpoznawanie układu dokumentu (OLR).

W kolejnym kroku przeprowadziliśmy ekstrakcję treści, zapisując dane w dwóch formatach: TXT (do analiz ilościowych) oraz JSON (z uwzględnieniem struktury dokumentu). Zastosowaliśmy również mechanizm deduplikacji dokumentów, oparty na porównywaniu tytułów, dat publikacji i nazw autorów. Informacje te pochodziły zarówno z metadanych, jak i z pełnych tekstów. Deduplikacja miała na celu eliminację powtórzeń wynikających z błędów lub nieścisłości w danych źródłowych. W rezultacie udało się wyodrębnić 27 247 unikatowych dokumentów pochodzących z lat 2001–2022.

Szczegółowa analiza metadanych dokumentów wykazała, że słowa kluczowe pochodzące z tezausa EuroVoc nie stanowią będą dobrej podstawy do mapowania wymiarów ubóstwa istotnych dla kształtowania polityki UE. Opracowywano je według różnych schematów, przez co były bardzo niespójne i reprezentowały różny poziom szczegółowości. Przede wszystkim jednak tylko w niewielkim zakresie obejmowały informacje o wskaźnikach ubóstwa i wykluczenia społecznego. Dlatego kolejnym krokiem było wydobycie wskaźników bezpośrednio z polityk publicznych Unii Europejskiej oraz z raportów i analiz dotyczących ubóstwa i wykluczenia społecznego.

W tym celu opracowaliśmy manualnie listę wskaźników dotyczących ubóstwa. Proces ten był rozłożony na kilka kroków. Na początku sporządziliśmy listę wskaźników występujących w tabelach Eurostat, Social Protection Performance Monitor prowadzonych przez Komitet Ochrony Socjalnej, Joint Assessment Framework oraz Social Scoreboard prowadzonych przez Komisję Europejską. Zależało nam jednak na uwzględnieniu w analizie wskaźników wykraczających poza te wymienione w najważniejszych zbiorach i narzędziach UE, a jednocześnie chcieliśmy określić sformułowania, jakie używane są do opisu danego wskaźnika. Dlatego z ogólnej puli dokumentów wylosowaliśmy 400, które podaliśmy ręcznej analizie. Dokumenty losowaliśmy w taki sposób, aby zachowane były proporcje względem liczby dokumentów oraz ich typów występujących w poszczególnych latach. Ze względu na ogromne zróżnicowanie typów dokumentów podzieliliśmy je jedynie na trzy kategorie: dokumenty prawne,

pozostałe dokumenty możliwe do otagowania¹ oraz dokumenty niemożliwe do otagowania (zbyt mało powtarzalne, niemożliwe do identyfikacji jako typ w tak dużym zbiorze). Tabela 1 przedstawia szczegółowo dobór dokumentów.

Tabela 1. Dobór próby dokumentów do ekstrakcji wskaźników z podziałem na poszczególne lata

Rok	Liczba dokumentów w roku	Liczba wylosowanych dokumentów	Dokumenty prawne	Pozostałe dokumenty (otagowane)	Pozostałe dokumenty (nieotagowane)	Razem
2001	850	12	4	4	4	12
2002	534	8	2	3	3	8
2003	635	9	3	3	3	9
2004	769	12	4	4	4	12
2005	528	8	2	3	3	8
2006	697	10	3	3	4	10
2007	652	10	3	3	4	10
2008	791	11	3	4	4	11
2009	2150	32	10	11	11	32
2010	1041	15	5	5	5	15
2011	1200	17	5	6	6	17
2012	1250	18	6	6	6	18
2013	1750	26	8	9	9	26
2014	1500	23	7	8	8	23
2015	1400	20	6	7	7	20
2016	1510	23	7	8	8	23
2017	1550	23	7	8	8	23
2018	2000	29	9	10	10	29
2019	2010	29	9	10	10	29
2020	1700	26	8	9	9	26
2021	1900	27	8	9	10	27
2022	830	12	4	4	4	12
	27247	400	123	137	140	400

Źródło: opracowanie własne.

¹ Zidentyfikowaliśmy następujące kategorie dokumentów pozaprawnych: *leaflet, brochure, facts and figures, findings, sheet, project, evaluation, news, review, study, analysis, agreement, strategy, report, programme*.

Tabela 2. Wyekstrahowane wskaźniki i ich występowanie w poszczególnych latach

Nazwa wskaźnika	Liczba wystąpień w danym okresie			
	2001–2005	2006–2010	2011–2015	2016–2021
at risk of poverty rate	190	2 125	4 803	12 066
at risk of poverty or social exclusion	0	0	866	2 891
in-work at-risk-of-poverty rate	4	186	455	1 249
relative median at-risk-of-poverty gap	21	147	391	785
severe material deprivation rate	0	3	341	736
persons living in households with very low work intensity	0	19	314	681
at-risk-of-poverty threshold	7	59	151	553
absolute poverty	32	151	167	394
Material Deprivation Index	6	230	111	205
indicators of indebtedness and overindebtedness	0	23	133	379
Human Poverty Index	6	255	109	72
energy poverty	0	6	16	416
impact of social transfers (excluding pensions) on poverty reduction	0	45	29	291
at risk of poverty rate 65+	0	121	33	115
at risk of poverty rate for children	3	54	50	159
Multidimensional Poverty in Environment Index	0	62	70	130
persistent at-risk-of-poverty rate	19	34	72	130

Źródło: opracowanie własne.

Przeszkolony przez nas anotor na podstawie lektury wylosowanych dokumentów ręcznie identyfikował wskaźniki i sposób ich opisu, rozszerzając wyjściową listę. Następnie tak przygotowany zbiór został zweryfikowany przez eksperta domenowego, a kontrowersyjne pozycje (tzn. te, co do których istniały wątpliwości, czy dotyczą ubóstwa lub do jakiej miary się odnoszą) zostały zweryfikowane przez cały zespół badawczy.

Listę tę wykorzystaliśmy jako podstawę do automatycznego przeszukiwania tekstów, co umożliwiło identyfikację dokumentów odwołujących się do poszczególnych wskaźników. Ze względu na różnorodność zapisów występujących w dokumentach (m.in. odmiany gramatyczne, różne formy zapisu terminów), zarówno pełne teksty dokumentów, jak i listę wskaźników poddaliśmy procesowi lematyzacji, sprowadziliśmy wszystkie wyrazy do małych liter oraz oczyściliśmy z elementów technicznych. Usunęliśmy między innymi interpunkcję, zbędne białe znaki, nagłówki stron, numeracje oraz inne powtarzające się fragmenty tekstu o charakterze redakcyjnym. Dzięki przygotowaniu wskaźników

i tekstów w jednolity sposób uzyskaliśmy możliwość ich bezpośredniego dopasowywania. W przygotowanym korpusie wyszukiwaliśmy wystąpienia każdego ze wskaźników w każdej publikacji, co pozwoliło nam zidentyfikować dokumenty, w których omawiano konkretne miary ubóstwa, a następnie powiązać te informacje z autorami, instytucjami odpowiedzialnymi za publikację oraz innymi elementami metadanych. Ostatecznie zawęziliśmy analizę do lat 2001–2021. Zdecydowały o tym dwa czynniki: po pierwsze, rok 2022 był niepełny, zbieranie dokumentów zakończyliśmy bowiem w lipcu. Dodatkowo chcieliśmy zachować spójność analizowanych okresów, koncentrując się na realizacji i podsumowaniu dwóch zintegrowanych strategii: strategii lizbońskiej oraz strategii Europa 2020 (ze względu na pandemię COVID-19 przyjęliśmy dłuższy czas podsumowywania strategii Europa 2020, a za najważniejszy moment tego procesu można uznać szczyt w Porto w 2021 roku dotyczący przyszłości polityk społecznej w UE). Ostatecznie do korpusu analizy weszło więc 26417 tekstów. Poniżej przedstawiamy wykaz wskaźników wyekstrahowanych w korpusie wraz z częstotliwością wystąpień (tabela 2).

Wskaźniki jako infrastruktura kalkulacyjna unijnej polityki walki z ubóstwem

Aby pokazać potencjał wykorzystania cyfrowych metod analizy dużych zbiorów danych tekstowych w naukach społecznych, odwołaliśmy się do metod analizy sieciowej. Naszym celem było pokazanie powiązań (krawędzi) między słowami kluczowymi a organizacjami posługującymi się nimi w dokumentach włączonych do korpusu. Analiza ta stanowi uzupełnienie szerszego projektu badawczego, którego celem jest zrozumienie roli wskaźników w kształtowaniu unijnej polityki walki z ubóstwem². W ramach projektu przeprowadzono między innymi 45 indywidualnych wywiadów pogłębionych z przedstawicielami instytucji unijnych i państw członkowskich, w tym Komisji Europejskiej, Parlamentu Europejskiego i Komitetu Ochrony Socjalnej³, oraz ekspertami, partnerami społecznymi i pozarządowymi, biorącymi udział w tworzeniu i realizacji unijnej polityki walki z ubóstwem. Wywiady miały na celu zrozumienie, w jaki sposób ci aktorzy posługują się wskaźnikami ubóstwa w codziennej pracy oraz jak je postrzegają i kategoryzują w ramach tych praktyk. Niniejsza analiza sieciowa stanowi ich rozwinięcie i pozwala uchwycić z szerszej perspektywy wzorce

² Projekt „Tworzenie polityki walki z ubóstwem w Unii Europejskiej za pomocą miar ilościowych – analiza przypadku złożonego wskaźnika zagrożenia ubóstwem i wykluczeniem społecznym” finansowany w ramach NCN OPUS nr UMO-2019/33/B/HS5/01955.

³ Komitet Ochrony Socjalnej stanowi ciało doradcze ministrów państw członkowskich UE w Radzie ds. Zatrudnienia, Polityki Społecznej, Zdrowia i Spraw Konsumenckich (EPSCO). Gromadzi przedstawicieli państw członkowskich oraz KE.

relacji, które umykają jakościowym analizom. Dzięki temu identyfikujemy nie tylko najczęściej stosowane wskaźniki, ale i ich instytucjonalne otoczenie, zmiany w czasie oraz mechanizmy rozbudowy unijnej infrastruktury kalkulacyjnej.

Przyjęliśmy perspektywę, zgodnie z którą wskaźniki odgrywają fundamentalną rolę w unijnej polityce walki z ubóstwem – nie są jedynie narzędziami opisu rzeczywistości społecznej, lecz aktywnie ją współtworzą (Bandola-Gill i in., 2022; Bruno, 2009; Louis i Maertens, 2021). Konstruuje ramy, w których definiowane są problemy społeczne, oceniana jest skuteczność działań oraz formułowane są interwencje. Wpływają przy tym na relacje między instytucjami unijnymi, a także na dystrybucję władzy w obrębie całej Unii. Wskaźniki stanowią zatem element szerszej infrastruktury kalkulacyjnej (Kurunmäki i in., 2019; Kurunmäki i Miller, 2013; Mehrpouya i Samiolo, 2019; Reilley i Scheytt, 2019) – zestawu narzędzi (np. baz danych i programów do administrowania nimi, wskaźników używanych do monitorowania sytuacji), praktyk i standardów, które umożliwiają zarządzanie polityką walki z ubóstwem poprzez liczby.

Ich znaczenie staje się szczególnie wyraźne w kontekście ograniczonego mandatu UE w zakresie polityki społecznej – większość kompetencji w tej dziedzinie pozostaje w gestii państw członkowskich. Unia może w tych obszarach jedynie wspierać, koordynować i uzupełniać działania krajowe, ale nie posiada uprawnień legislacyjnych. Właśnie dlatego wskaźniki pełnią funkcję narzędzi wpływu politycznego – pozwalają instytucjom unijnym, zwłaszcza Komisji Europejskiej, oddziaływać na politykę społeczną państw członkowskich mimo ograniczonych możliwości formalnej regulacji.

Siła wskaźników wynika nie z mocy prawa, lecz z logiki danych, porównań i standardów, które wpływają na sposób postrzegania problemów i oceny polityk publicznych. W tym sensie wskaźniki i generowane na ich podstawie benchmarki, rankingi oraz scoreboardy nie wymuszają decyzji, ale kreują ramy ich podejmowania, sprawiając, że pewne rozwiązania stają się bardziej „racjonalne”, „efektywne” lub „oparte na dowodach”. W dalszej części artykułu dzięki analizie sieciowej pokazujemy, jak ta infrastruktura kalkulacyjna rozwijała się w czasie: które wskaźniki odgrywały rolę centralnych węzłów sieci, w jaki sposób dystrybuowana była wokół nich aktywność różnych instytucji oraz jak zmieniała się logika pomiaru i interwencji w polityce unijnej. Pozwoli to przejść od opisu ogólnych ról wskaźników do empirycznego rozpoznania ich funkcji jako instrumentów selekcji, legitymizacji i wpływu (Armstrong, 2012; Lelie i Vanhercke, 2013; Zeitlin, 2009; Zeitlin i in., 2005).

Wskaźniki jako narzędzie negocjacji politycznych i legitymizacji

Wywiady z przedstawicielami wspomnianych instytucji unijnych ujawniły, że uzgadnianie wskaźników to nie tylko kwestia techniczna, ale przede wszystkim

proces negocjowania priorytetów politycznych. To, co zostaje objęte monitorin-
giem, zyskuje widoczność, natomiast zjawiska pozostające poza ramami pomia-
ru mogą łatwo wypaść z agendy politycznej. Wskaźniki pełnią funkcję narzędzi
selekcji – decydują o tym, które problemy społeczne są uznane za istotne, a któ-
re pozostają w cieniu. Nie tyle odwzorowują rzeczywistość, ile ją kształtują –
poprzez definiowanie obszarów uznawanych za wymagające interwencji. Odgry-
wają także ważną rolę w rozgrywkach instytucjonalnych: Komisja Europejska,
mimo ograniczonych kompetencji w zakresie polityki społecznej, wykorzystuje
wskaźniki jako instrument wpływu i legitymizacji. Umożliwiają one wskazywa-
nie zaniedbanych obszarów, ocenę skuteczności krajowych strategii i uzasadnia-
nie potrzeby reform (Boswell, 2008; Bruno, 2009; Lelie i Vanhercke, 2013; The-
dvall, 2006). Jak ujął to jeden z przedstawicieli Komisji Europejskiej:

*dopóki nie masz porównywalnych danych z państw członkowskich, to jesteś kompletnie
bezbronny – nie masz jak powiedzieć, które z nich powinny zrobić więcej.* (EC4)

Z badań jakościowych wyłania się wyraźna wizja infrastruktury kalkulacyj-
nej jako narzędzia rozbudowy kompetencji ponadnarodowych. W szczególności
Komisja Europejska legitymizuje swoje działania przez odwołanie do polityki
opartej na dowodach i do epistemicznej kultury liczb (Boswell, 2008; Thedvall,
2006) – przekonania, że statystyki nadają procesom decyzyjnym charakter tech-
niczny i obiektywny. Elementy tej kultury są stopniowo przyswajane również
przez inne instytucje, takie jak Rada UE czy Parlament Europejski.

Opisane procesy nie są jedynie deklaratywne – mają odzwierciedlenie
w strukturze powiązań między wskaźnikami i instytucjami. Przedstawiona w dal-
szej części artykułu analiza sieciowa pozwala uchwycić te zjawiska empirycznie.
Odpowiada między innymi na pytania:

- 1) Które wskaźniki, a wraz z nimi wymiary ubóstwa, są centralne z punktu
widzenia kształtowania polityki walki z ubóstwem?
- 2) Jak aktorzy na poziomie unijnym różnią się pod względem zestawów wskaź-
ników, które promują jako centralne dla polityki walki z ubóstwem, oraz czy
różnice te mogą być powiązane z potencjalnymi napięciami wynikającymi
z konkurujących priorytetów w zakresie przeciwdziałania ubóstwu?
- 3) Jakie zmiany zachodzą na przestrzeni lat 2001–2021? Jakie widać prze-
sunięcia w relatywnym znaczeniu poszczególnych wskaźników ubóstwa?

Dzięki analizie sieci możliwe jest wyjście poza jednostkowe perspektywy
rozmówców i uchwycenie bardziej trwałych wzorców związanych z wykorzy-
stywaniem wskaźników. Umożliwia to także analizę zmian w dłuższej perspek-
tywie czasowej – niezależnie od indywidualnej pamięci i ocen. Jednocześnie jed-
nak pełne zrozumienie znaczenia wyników analizy sieciowej i wizualizacji sieci
byłoby niemożliwe bez wiedzy pochodzącej z wywiadów – to ona pozwala osa-
dzić wzorce sieciowe w konkretnych praktykach i kontekstach instytucjonalnych.

Procesy negocjacji i legitymizacji, dostrzegalne w wywiadach, znajdują odbicie również w analizie sieciowej. Pozwala ona zobaczyć, które wskaźniki zostały trwale zakorzenione w unijnej infrastrukturze kalkulacyjnej, a które – mimo politycznych ambicji – pozostały marginalne lub niedystrybuowane. W dalszej części pokazujemy, jak logika kompromisów i legitymizacji odciska się na strukturze relacji między wskaźnikami i instytucjami.

Przygotowanie matrycy do analizy

Martyce sieci zostały opracowane na podstawie wyekstrahowanych wskaźników. W efekcie uzyskaliśmy nieskierowane, dwumodalne sieci ważone zawierające instytucję (nazwę instytucji), która przygotowała dany dokument, i wyekstrahowaną frazę. Powstała sieć semantyczna (Pfiffner, 2021), gdzie wartość krawędzi odzwierciedla liczbę wystąpień danego wskaźnika w dokumentach instytucji. Dla większej klarowności przedstawionych grafów sieci dokonaliśmy odcięcia wartości od średniej wag powiązań w sieci, dodając dwa razy odchylenie standardowe. Pozwoliło to na ukazanie najsilniejszych powiązań między wskaźnikami a instytucjami w każdym okresie. Analiza została wykonana przy pomocy programu Visone (<https://visone.ethz.ch/>).

Trzy instytucje występowały pod kilkoma postaciami: Komisja Europejska, Parlament Europejski oraz Komitet Regionów. Funkcjonowały w dokumentach albo pod wymienionymi nazwami nadrzędnymi, albo jako konkretne jednostki wchodzące w skład tych większych całości (np. Directorate-General Employment, Social Affairs and Inclusion, Committee on Employment and Social Affairs). Ponieważ zależało nam na przyjrzeniu się relacjom między zróżnicowanymi podmiotami wytwarzającymi dokumenty w ramach polityki walki z ubóstwem UE, zdecydowaliśmy się na agregację tych poszczególnych jednostek pod parasolem instytucji nadrzędnych. Jednocześnie dopuściliśmy dwa wyjątki: Eurostat oraz Joint Research Centre jako specjalistyczne jednostki zajmujące się między innymi produkcją wskaźników i danych wchodzące w skład Komisji Europejskiej postanowiliśmy rozpatrywać oddzielnie. Chodziło o to, by zobaczyć, w jakim stopniu posługiwanie się wskaźnikami przez Komisję będzie zawężone do tych ciał, które skupiają się na produkcji wiedzy opartej na danych statystycznych.

Analizę podzieliliśmy na cztery okresy, zgodne z dynamiką wdrażania strategii lizbońskiej oraz strategii Europa 2020: 2001–2005, 2006–2010, 2011–2015 oraz 2016–2021. Każda ze strategii przechodziła tak zwane śródkresowe podsumowania, z którymi wiązały się modyfikacje ich realizacji. Jednocześnie podział ten pozwolił nam na utrzymanie mniej więcej równych przedziałów czasowych, w ramach których prowadzimy porównania.

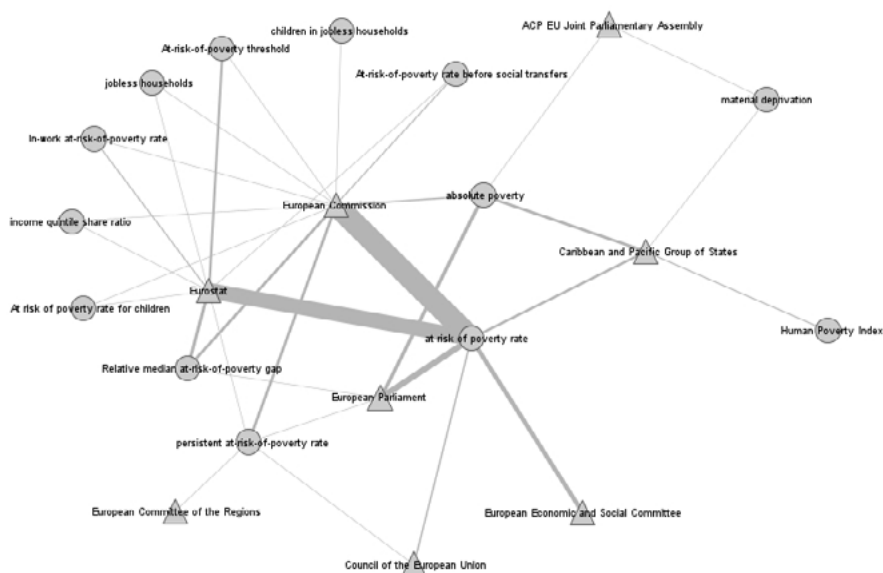
Rozrost sieci wskaźników i centralna pozycja AROP

Centralne wskaźniki. AROP jako fundament infrastruktury kalkulacyjnej

Analiza sieciowa pokazuje, że w centrum sieci przez cały ten czas niezmiennie pozostawał wskaźnik zagrożenia ubóstwem relatywnym (ang. *at-risk of poverty rate*, AROP), wprowadzony w obecnej formie wraz ze strategią lizbońską jako jeden z podstawowych wskaźników społecznych uzgodnionych przez państwa członkowskie na szczycie w Laeken w 2001 roku⁴.

W pierwszym okresie analizy (2001–2005) wystąpiło 8 aktorów instytucjonalnych i 13 wskaźników. Biorąc pod uwagę miarę centralności stopnia (ang. *degree centrality*) dla węzłów wyrażoną w procentach⁵, wyróżniała się Komisja Europejska (centralność na poziomie 13%) i Eurostat (centralność na poziomie 11%), które najczęściej posługiwały się właśnie wskaźnikiem AROP.

Graf 1. Sieć powiązań aktorów instytucjonalnych i wskaźników 2001–2005



Legenda: trójkąty – aktorzy instytucjonalni; koła – wskaźniki.

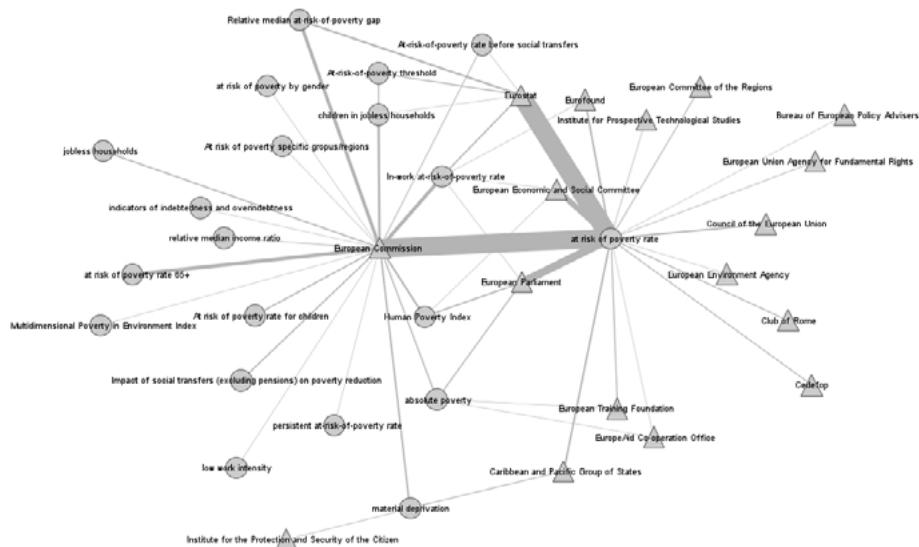
Źródło: opracowanie własne.

⁴ Wskaźnik zagrożenia ubóstwem relatywnym (AROP) to odsetek osób, których ekwiwalentny dochód do dyspozycji (po transferach socjalnych) jest niższy od 60% mediany krajowych dochodów do dyspozycji (Eurostat).

⁵ Zastosowano tu tylko miarę liczby powiązań, bez uwzględnienia wag. Oznacza to, że dany węzeł ma, przykładowo, 13% połączeń z maksymalnej liczby możliwych. Centralność została obliczona dla każdej kategorii węzłów oddzielnie.

W drugim okresie (2006–2010) powiększona sieć zawiera 20 wskaźników i 17 aktorów instytucjonalnych. Komisja Europejska zwiększyła jednak swoją centralność do 18% i była powiązana ze wszystkimi dwudziestoma wskaźnikami, z AROP jako dominującym (centralność na poziomie 31%).

Graf 2. Sieć powiązań aktorów instytucjonalnych i wskaźników 2006–2010



Legenda: trójkąty – aktorzy instytucjonalni; koła – wskaźniki.

Źródło: opracowanie własne.

AROE jako wskaźnik kompromisu

Od 2010 roku konfiguracja sieci uległa zmianie: pojawił się nowy, złożony wskaźnik ubóstwa lub wykluczenia społecznego (ang. *at risk of poverty or social exclusion*, AROPE), mający połączyć różne wymiary ubóstwa w duchu kompromisu między Komisją a państwami członkowskimi (Copeland i Daly, 2012). Składa się on z trzech komponentów: wskaźnika relatywnego ubóstwa dochodowego (AROP), mierzącego odsetek osób żyjących poniżej 60% mediany krajowego dochodu; deprywacji materialnej (SMD)⁶, czyli braku dostępu do podstawowych dóbr i usług; oraz niskiej intensywności pracy (LWI), odnoszącej się do bardzo ograniczonego udziału dorosłych członków gospodarstwa domowego w rynku pracy.

⁶ Obecnie, po rewizji wskaźnika, wprowadzono jego nową wersję Severe Material and Social Deprivation (SMSD).

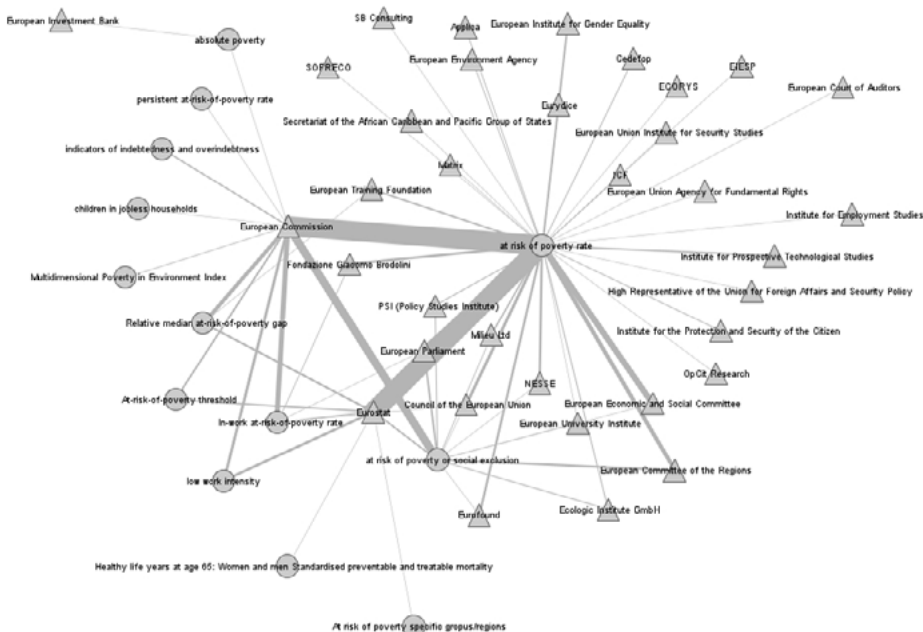
Choć AROPE uzyskał status wskaźnika wiodącego w strategii Europa 2020, jego relatywna centralność pozostała ograniczona, a dyfuzja – znacznie słabsza niż w przypadku AROP. Jak ujął to jeden z członków Komitetu Ochrony Socjalnej:

Był to rodzaj kompromisu politycznego, który był niezbędny, aby cel mógł odzwierciedlać wpływ różnych polityk, jakie państwa członkowskie mogłyby wdrażać w celu ograniczenia zmniejszenia ubóstwa w UE lub w ich własnym kraju. (SPC3)

Pomimo formalnego statusu AROPE jako głównego wskaźnika realizacji celu redukcji ubóstwa w strategii Europa 2020 na podstawie analizy sieci można stwierdzić, że nie osiągnął on porównywalnej do AROP pozycji. W sieciach z lat 2011–2015 oraz 2016–2021 jego centralność i liczba powiązań pozostały wyraźnie niższe.

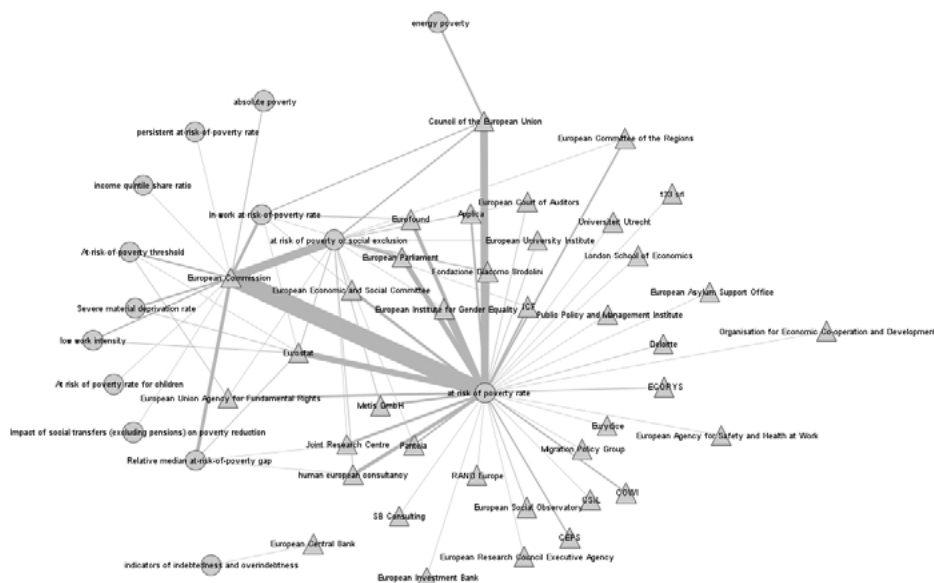
W trzecim okresie (2011–2015) sieć rozrosła się do 35 aktorów, ale wystąpiło w niej 13 wskaźników, a więc ich liczba zmniejszyła się w stosunku do poprzedniego okresu. W centrum nadal plasował się AROP (centralność na poziomie 36%). Tym razem najsilniejsza reprezentacja tego wskaźnika występuje w dokumentach Eurostatu i Komisji Europejskiej. Obok niego pojawił się AROPE jako nowy, lecz wciąż wtórny węzeł.

Graf 3. Sieć powiązań aktorów instytucjonalnych i wskaźników 2011–2015



Legenda: trójkąty – aktorzy instytucjonalni; koła – wskaźniki.

Źródło: opracowanie własne.

Graf 4. Sieć powiązań aktorów instytucjonalnych i wskaźników 2016–2021

Źródło: opracowanie własne.

Kolejne okresy przynoszą wzrost liczby instytucji i wskaźników oraz idącą za tym zmianę zagęszczenia powiązań. Proces ten obserwować można na podstawie miary gęstości sieci⁷. W pierwszym okresie (2001–2005) wynosiła ona 0,32, w drugim okresie (2006–2010) – 0,13, w trzecim (2011–2015) zaś i czwartym (2016–2021) – odpowiednio 0,15 i 0,14. Gęstość powiązań od drugiego okresu dwukrotnie zmalała, a potem utrzymywała się prawie na tym samym

⁷Policzyliśmy gęstość sieci po dychotomizacji (dla sieci binarnych, gdzie każda wartość została zastąpiona 1). Gęstość została policzona dla sieci dwumodalnych ze wzoru $m/(N \text{ wierszy} \times N \text{ kolumn})$, gdzie m to liczba niezerowych komórek (powiązań).

poziomie. Oznacza to, że pomiędzy wskaźnikami i instytucjami odsetek potencjalnych połączeń zmalał, podczas gdy w pierwszym okresie był on najwyższy przy najmniejszych rozmiarach sieci.

Niższa gęstość sieci w trzech ostatnich okresach wiąże się z pojawieniem się w niej nowych aktorów instytucjonalnych – wraz z rozrostem sieci zmniejsza się jej wewnętrzna spójność. Istotne jest jednak to, że od końca 2005 roku średnia wartość połączeń w sieci znacząco wzrosła⁸. Można to interpretować jako zwiększenie częstotliwości wzmiankowania wskaźników w dokumentach przez badane instytucje w kolejnych okresach. Innymi słowy, choć aktorzy instytucjonalni są obecnie mniej ze sobą powiązani niż w początkowej, niewielkiej sieci, częściej odwołują się do wskaźników.

Ograniczona dyfuzja AROPE poza strukturami administracji UE

Porównanie sieci instytucjonalnych związanych z AROP i AROPE w latach 2011–2015 oraz 2016–2021 uwidacznia istotne różnice w stopniu instytucjonalizacji i dyfuzji obu mierników. Wskaźnik AROP, funkcjonujący w polityce unijnej od początku realizacji strategii lizbońskiej, charakteryzuje się znacznie szerszym zakresem oddziaływania. Wokół niego gromadzą się zarówno centralne instytucje unijne (Komisja Europejska, Parlament Europejski, Rada UE), jak i liczne agencje UE (m.in. Eurostat, JRC, FRA, Cedefop, ETF, EEA), uczelnie (EUI, LSE, Utrecht University), organizacje pozarządowe i partnerzy społeczni (np. Fondazione Giacomo Brodolini, Social Platform), a także wiele firm doradczych i badawczych (ICF, RAND Europe, Deloitte, CSIL, Matrix, Ecorys).

Tabela 4. Aktorzy połączeni z AROP w latach 2011–2015 i 2016–2021

2011–2015	
Typ organizacji/ instytucji	Organizacja/instytucja
Instytucje centralne UE i ich jednostki	European Commission, European Parliament, Council of the EU, Eurostat, European Economic and Social Committee, European Committee of Regions, Institute for the Protection and Security of the Citizen (IPSC) jednostka badawcza Joint Research Centre (JRC), Institute for Prospective Technological Studies jednostka badawcza Joint Research Centre (JRC), High Representative of the Union for Foreign Affairs and Security Policy

⁸ Policzyliśmy także gęstość ważoną, która jest sumą wszystkich wartości powiązań podzieloną przez liczbę możliwych powiązań. Jest to więc średnia wartość powiązań w sieci. Zastosowano znormalizowaną miarę gęstości dla matryc w każdym okresie (Wasserman i Faust, 2012). W pierwszym okresie (2001–2005) wynosiła ona 0,013, w drugim okresie (2006–2010) – 0,13, w trzecim (2011–2015) zaś i czwartym (2016–2021) – odpowiednio 0,28 i 0,72.

Agencje UE	European Union Agency for Fundamental Rights (FRA), European Court of Auditors (ECA), Eurydice, European Environment Agency (EEA), European Institute for Gender Equality (EIGE), European Training Foundation (ETF), European Centre for the Development of Vocational Training (Cedefop), European Union Institute for Security Studies (EUISS), European Foundation for the Improvement of Living and Working Conditions (Eurofound)
Organizacje pozarządowe / partnerzy społeczni	–
Firmy doradcze / agencje badawcze	ICF Consulting Group, Ecorys Research and Consulting, Op Cit Research, Applica, SB Consulting, Matrix, Milieu Ltd, SOFRECO
Uniwersytety	European University Institute (EUI)
Think tanki / instytuty i sieci badawcze	Fondazione Giacomo Brodolini, Network of Experts in Social Sciences of Education and Training (NESSE), Ecologic Institute GmbH, Institute for Employment Studies, Policy Studies Institute (PSI), European Institute of Education and Social Policy (EIESP)
Inne organizacje	Secretariat for the African, Caribbean and the Pacific Group of States (ACPS)
2016–2021	
Typ organizacji/instytucji	Organizacja/instytucja
Instytucje centralne UE i ich jednostki	European Commission, European Parliament, Council of the EU, Eurostat, Joint Research Centre (JRC), European Investment Bank (EIB), European Economic and Social Committee, European Committee of Regions,
Agencje UE	European Union Agency for Fundamental Rights (FRA), Eurydice, European Research Council Executive Agency (ERCEA), European Asylum Support Office (EASO), European Agency for Safety and Health at Work (EU-OSHA), European Court of Auditors (ECA), European Institute for Gender Equality (EIGE), European Foundation for the Improvement of Living and Working Conditions (Eurofound)
Organizacje pozarządowe / partnerzy społeczni	–
Firmy doradcze / agencje badawcze	ICF Consulting Group, Ecorys, Deloitte, RAND Europe, Centre for Industrial Studies (CSIL), Applica, Metis GmbH, t33 Srl, Human European Consultancy, Panteia, COWI
Uniwersytety	European University Institute (EUI), Utrecht University, London School of Economics and Political Science (LSE)
Think tanki / instytuty badawcze	Fondazione Giacomo Brodolini, European Social Observatory (OSE), Public Policy and Management Institute (PPMI), Migration Policy Group (MPG), Centre for European Policy Studies (CEPS)
Inne organizacje	Organisation for Economic Co-operation and Development

Źródło: opracowanie własne.

Natomiast AROPE, mimo ambicji syntetycznego ujęcia problematyki ubóstwa i wykluczenia społecznego poprzez połączenie trzech wskaźników składowych, pozostał miernikiem o bardziej ograniczonym zakresie stosowania. Jego

Tabela 5. Aktorzy połączeni z AROPE w latach 2011–2015 i 2016–2021

2011–2015	
Typ organizacji/institucji	Organizacja/institucja
Instytucje centralne UE i ich jednostki	European Commission, European Parliament, Council of the EU, Eurostat, European Committee of the Regions, European Economic and Social Committee
Agencje UE	European Foundation for the Improvement of Living and Working Conditions (Eurofound)
Organizacje pozarządowe / partnerzy społeczni	–
Firmy doradcze / agencje badawcze	Milieu Ltd
Uniwersytety	European University Institute (EUI)
Think tanki / instytuty i sieci badawcze	Network of Experts in Social Sciences of Education and Training (NESSE), Policy Studies Institute (PSI), Ecologic Institute GmbH
Inne organizacje	–
2016–2021	
Typ organizacji/institucji	Organizacja/institucja
Instytucje centralne UE i ich jednostki	European Commission, European Parliament, Council of the EU, Eurostat, European Economic and Social Committee, Joint Research Centre (JRC), European Committee of the Regions
Agencje UE	European Court of Auditors (ECA), European Institute for Gender Equality (EIGE), European Foundation for the Improvement of Living and Working Conditions (Eurofound), European Union Agency for Fundamental Rights
Organizacje pozarządowe / partnerzy społeczni	–
Firmy doradcze / agencje badawcze	Applica, Panteia, Human European Consultancy, Metis GmbH, ICF Consulting Group
Uniwersytety	European University Institute (EUI)
Think tanki / instytuty i sieci badawcze	Fondazione Giacomo Brodolini
Inne organizacje	–

Źródło: opracowanie własne.

użytkowanie skupia się głównie wokół Komisji Europejskiej, Eurostatu, Eurofound oraz wybranych agencji unijnych. Dopiero w drugim analizowanym okresie (2016–2021) w sieci AROPE zaczynają pojawiać się firmy doradcze i think tanki, jednak ich obecność pozostaje incydentalna i znacznie słabsza niż w przypadku AROP. Również liczba zaangażowanych uczelni i organizacji pozarządowych jest ograniczona.

Różnice te mogą wynikać zarówno z historycznego zakorzenienia AROP w infrastrukturze monitoringu społecznego UE, jak i z problemów konceptualnych związanych z AROPE. W literaturze przedmiotu wskazuje się, że konstrukcja AROPE – łącząca wskaźniki dochodowe, deprywacyjne i związane z intensywnością pracy – nie została poprzedzona pogłębioną analizą relacji między tymi wymiarami. W efekcie wskaźnik ten może nietrafnie identyfikować najbardziej zagrożone grupy, a jego złożona struktura utrudnia interpretację oraz praktyczne zastosowanie (Gábos i Goedemé, 2016; Maître i in., 2013; Nolan i Whelan, 2011).

Z tej perspektywy AROPE jawi się przede wszystkim jako narzędzie politycznego kompromisu, mającego pogodzić różne podejścia do polityki społecznej obecne w państwach członkowskich. Choć jego konstrukcja odzwierciedla ambicję ujęcia ubóstwa w sposób bardziej złożony i wielowymiarowy, to w praktyce jego znaczenie analityczne i zdolność mobilizacji instytucjonalnej mają ograniczenia.

Podobna sytuacja dotyczy komponentów wskaźnika, a więc LWI i SMD. Choć z czasem zakorzeniają się mocniej wśród kluczowych aktorów unijnych, nie osiągają jednak podobnej pozycji do AROP. W następnej części przeanalizujemy je wraz z innymi wskaźnikami, które ujawniają się w sieciach, zwracając przede wszystkim uwagę na to, do jakich szczegółowych obszarów ubóstwa się odnoszą.

Komplementarne wskaźniki i profesjonalizacja pomiaru

Poza wskaźnikami centralnymi istotną rolę odgrywają wskaźniki komplementarne, które pozwalają uchwycić różne wymiary i przyczyny ubóstwa. Pomiędzy okresem 2001–2005 a 2006–2010 następuje wyraźny wzrost liczby wskaźników (z 12 komplementarnych wskaźników + AROP do 18 komplementarnych wskaźników + AROP), po czym w latach 2011–2015 spada do poziomu zbliżonego do początkowego (11 komplementarnych wskaźników + AROP i AROPE), a w ostatnim okresie rośnie nieznacznie (12 komplementarnych + AROP i AROPE). Zmiany te interpretujemy jako przejście od początkowej fazy wdrażania pierwszych uzgodnionych w ramach unijnej polityki walki z ubóstwem wskaźników; poprzez okres intensywnej ekspansji wskaźników, głównie związanych z pomiarem ubóstwa relatywnego, a więc AROP, do konsolidacji wokół AROPE, który skupia w sobie podejście redystrybucyjne, deprawacyjne i aktywizacyjne; a w ostatnich latach – do poszerzenia zakresu pomiaru o nową kategorię ubóstwa energetycznego.

Skok w drugim okresie wiąże się częściowo z tym, że aktorzy (głównie Komisja Europejska) stosują wskaźnik AROP do diagnozy ubóstwa specyficznych

grup – dzieci, osób starszych – oraz zróżnicowania grup ze względu na płeć. Jest to również moment pojawiania się wskaźnika *low work intensity*, który w następnych okresach zastąpi *jobless households* i stanie się komponentem AROPE.

Przez cały okres badania utrzymuje się stała grupa pięciu wskaźników komplementarnych. Większość tych miar, a zatem *at-risk-of-poverty threshold*, *relative median at-risk-of-poverty gap*, *persistent at-risk-of-poverty rate*, uzupełniają AROP o nowe wymiary podejścia redystrybucyjnego, które pozwalają ocenić głębokość i trwałość ubóstwa oraz wysokość progu ubóstwa. Dopiero w ostatnim okresie w sieci pojawia się natomiast *income quintile share ratio* (S80/S20), informujący o różnicy dochodów między najbogatszymi i najbiedniejszymi gospodarstwami domowymi, który osadza ubóstwo w szerszym kontekście dystrybucji dochodu i nierówności społecznych.

Innym stale występującym wskaźnikiem jest *in-work at risk of poverty*, zwracający uwagę na relatywne ubóstwo wśród osób pracujących. Za każdym razem współwystępuje z innymi wskaźnikami nawiązującymi do podejścia aktywizacyjnego, akcentującymi rolę zatrudnienia w walce z ubóstwem: początkowo z *people living in jobless households*, a następnie z *low work intensity*. Od 2011 roku obszar niskiej intensywności pracy zostaje także włączony do wskaźnika AROPE, co dodatkowo wzmacnia to podejście.

Warto zaznaczyć, że współwystępowanie wskaźników redystrybucyjnych i aktywizacyjnych odzwierciedla dwa podejścia do polityki społecznej UE, które pozostają w pewnym napięciu: głównym celem pierwszego z nich ma być łagodzenie skutków ubóstwa poprzez transfery, drugiego – raczej jego prewencja poprzez zwiększanie aktywności zawodowej. Jak wspomniano już wcześniej, oba podejścia połączone zostaną w ramach wskaźnika AROPE.

Wreszcie, stale występują w sieci również wskaźniki wywodzące się z podejścia absolutnego. Początkowo ograniczone do dokumentów z obszaru polityki rozwojowej i krajów Afryki, Karaibów i Pacyfiku (AKP), po 2016 roku zaczęły pojawiać się w dokumentach Komisji Europejskiej niełączących się z tą polityką poprzez aktorów instytucjonalnych. Może to oznaczać, po pierwsze, rosnące zainteresowanie uchwyceniem skrajnego ubóstwa w państwach członkowskich UE, a po drugie – proces umiędzynarodowienia metod pomiarowych i adaptacji globalnych ram, takich jak Cele Zrównoważonego Rozwoju (SDG 1: End Poverty), do unijnej polityki społecznej. Warto przy tym odnotować eksperymenty prowadzone przez Komisję Europejską z wykorzystaniem budżetów referencyjnych jako podstawy do określania minimalnych potrzeb życiowych (Menyhért i in., 2021).

W tym kontekście warto zwrócić uwagę na obecność w unijnej infrastrukturze kalkulacyjnej wielowymiarowych wskaźników ubóstwa opracowanych przez ONZ, takich jak *Human Poverty Index* (HPI) i jego następcę – *Multidimensional*

Poverty Index (MPI). Występują one jednak tylko w pierwszych trzech okresach, co może wiązać się z przesunięciem optyki na kwestię ubóstwa absolutnego, a także zakorzenianiem się w UE własnej miary złożonej, czyli AROPE.

Podjęcie deprywacyjne jest obecne przez cały badany okres, choć jego znaczenie i formy ewoluują. Odwołanie od miar deprywacji w pierwszym okresie ma związek z polityką rozwojową UE i krajami AKP. Dopiero w następnym okresie deprywacja przesuwają się w wewnętrzny obszar zainteresowania Komisji Europejskiej. Jak wskazują wywiady, jest to również okres intensywnej pracy nad własną unijną miarą deprywacji materialnej, w związku z rozszerzeniem UE o państwa Europy Środkowo-Wschodniej i Południowej (EX4). W państwach tych różnice dochodowe były mniejsze niż w tak zwanej starej UE, przez co AROP nie poradził sobie z uchwyceniem różnic w poziomie ubóstwa materialnego. Dopiero miary pokazujące brak dostępu do dóbr i aktywności pozwalały lepiej diagnozować sytuację. Podjęcie deprywacyjne, podobnie jak aktywizacyjne, zostało wzmocnione w trzecim okresie dzięki wprowadzeniu wskaźnika AROPE, który obejmuje komponent *severe material deprivation* (SMD). W ostatnim okresie SMD zaczyna również funkcjonować samodzielnie jako miara wykorzystywana w dokumentach Komisji Europejskiej i Eurostatu. Jego uzupełnieniem jest wskaźnik *overindebtedness*, który rozszerza analizę o aspekt braku dostępu do dóbr i usług wynikający z nadmiernego zadłużenia. Wskaźniki te stały się istotnym uzupełnieniem wskaźników dochodowych, pokazując materialne wymiary wykluczenia społecznego.

Nowym obszarem pomiaru staje się *energy poverty* (tzw. ubóstwo energetyczne), który w ostatnim analizowanym okresie pojawia się w sieci jako wyodrębniona kategoria wskaźników. Choć nie posiada jednej ustalonej definicji, jego obecność może świadczyć o wstępnym etapach instytucjonalizacji tego zjawiska jako elementu polityki społecznej UE. W analizowanej sieci zajmuje ono jednak peryferyjną pozycję i pozostaje słabo powiązane z kluczowymi instytucjami – co ogranicza jego polityczną sprawczość. W kontekście Zielonego Ładu ubóstwo energetyczne zyskuje na znaczeniu – wskazuje na to, jak transformacja klimatyczna redefiniuje politykę społeczną UE. Osoby żyjące w ubóstwie często nie mają dostępu do „czystej” energii i są szczególnie narażone na rosnące koszty życia. Pomiar w tym obszarze może zatem pełnić funkcję legitymizującą dla nowych interwencji politycznych.

Zmiany w zestawie wskaźników pokazują, że poszerzanie wachlarza miar nie jest jedynie procesem technicznym wynikającym z profesjonalizacji pomiaru. Jak wskazują przedstawiciele instytucji unijnych (EC3, EC4, SPC1, SPC3), wprowadzanie nowych wskaźników pełni także funkcję polityczną, stanowiąc strategię rozszerzania wpływu UE na kolejne obszary polityki społecznej. Wskaźniki dotyczące rynku pracy, deprywacji, nierówności czy sytuacji dzieci stają się narzędziem nie tylko diagnozy, lecz także legitymizacji interwencji

w sferach wcześniej pozostających w gestii państw członkowskich. W efekcie pomiar ubóstwa staje się formą ponadnarodowego zarządzania społeczną agendą UE, w której instytucje europejskie definiują problemy społeczne i wyznaczają kierunki ich rozwiązywania.

Rola Komisji Europejskiej w zarządzaniu infrastrukturą wskaźników

Komisja Europejska odgrywa kluczową rolę w rozwijaniu i utrzymywaniu infrastruktury kalkulacyjnej unijnej polityki walki z ubóstwem. Jej centralność utrzymywała się w każdym okresie i wynosiła odpowiednio 13%, 16%, 17%, 18%. To nie zaskakuje, jeśli weźmie się pod uwagę, że Komisja Europejska jest głównym aktorem odpowiedzialnym za wdrażanie wskaźników i monitorowanie zmian w obszarze walki z ubóstwem

Wskaźniki pełnią funkcję narzędzi miękkiej władzy (Bruno i in., 2006; Porte i in., 2001; Lelie i Vanhercke, 2013), pozwalając Komisji wpływać na polityki narodowe poprzez dane, standardy i porównania, a nie poprzez twarde normy prawne. Dzięki temu Komisja wyznacza priorytety, ustala ramy polityk i pośrednio dyscyplinuje działania państw członkowskich, wzmacniając zarządzanie oparte na wskaźnikach (ang. *governance by indicators*). Jednocześnie wskaźniki nie pełnią wyłącznie funkcji instrumentalnej związanej z realizacją celów politycznych, ale mają także wymiar symboliczny. Umożliwiają Komisji legitymizowanie roszczeń politycznych w obszarach, w których nie posiada formalnego autorytetu legislacyjnego, i pozwalają jej pozycjonować się jako ciało neutralne politycznie oraz eksperckie, działające w na podstawie dowodów empirycznych (Boswell, 2008).

Po 2016 roku obserwujemy rosnącą eksternalizację ekspertyzy (Gornitzka i Sverdrup, 2015; Metz, 2013; Năstase i Radulova, 2025) – część funkcji analitycznych i doradczych zostaje powierzona wyspecjalizowanym agencjom (np. JRC, EIGE, EASO), firmom konsultingowym (np. Deloitte, RAND Europe) oraz środowisku akademickiemu (LSE, Utrecht University). Taka strategia umożliwia Komisji nie tylko zwiększenie zasobów wiedzy, ale także wzmocnienie wizerunku aktora eksperckiego – opartego na naukowości i technokratycznej racjonalności. Zewnętrzni partnerzy nadają pozory niezależności procesom definiowania problemów społecznych i sposobom ich pomiaru, co dodatkowo zwiększa siłę przekonywania Komisji Europejskiej wobec państw członkowskich.

Rozbudowa sieci aktorów wspierających produkcję wiedzy o ubóstwie sprzyja kształtowaniu polityki społecznej jako europejskiego problemu publicznego przekraczającego kompetencje narodowe. Komisja, organizując tę infrastrukturę

wokół danych, raportów i wskaźników, wzmacnia swoją sprawczość w obszarach formalnie poza jej legislacyjnym zasięgiem – sprawując *de facto* władzę poprzez wiedzę i standardy.

Wnioski

W artykule pokazaliśmy, że analiza sieciowa oparta na korpusie ponad 26 tysięcy dokumentów unijnych z lat 2001–2021 i towarzyszące jej wizualizacje sieci wnoszą nową perspektywę do badań nad unijną polityką walki z ubóstwem. Umożliwiają one analizę współwystępowania wskaźników, tematów i instytucji oraz zmian w relacjach między nimi w dłuższym horyzoncie czasowym, co trudno uchwycić za pomocą tradycyjnych metod jakościowych i ilościowych. Jednocześnie wywiady i literatura przedmiotu okazały się niezbędne do interpretacji tych procesów oraz osadzenia ich w szerszym kontekście unijnej infrastruktury kalkulacyjnej i polityki walki z ubóstwem.

Analiza gęstości sieci wykazała zwiększającą się fragmentaryzację i rozproszenie relacji wraz ze wzrostem liczby uczestników sieci. W pierwszym okresie (2001–2005) gęstość sieci była najwyższa (0,32) przy niewielkiej liczbie aktorów i wskaźników, co wskazuje na silne powiązania pomiędzy instytucjami. W kolejnych okresach, w związku z dochodzeniem nowych aktorów, gęstość spadła gwałtownie (do 0,13 w latach 2006–2010) i utrzymywała się na niskim, stabilnym poziomie (0,14–0,15). Jednocześnie wzrosła średnia wartość połączeń, co można interpretować jako częstsze odwoływanie się instytucji do wskaźników pomimo mniejszej wewnętrznej spójności sieci.

Analiza centralności wskazała, że przez cały badany okres dominującą pozycję zachowywał wskaźnik AROP, wokół którego skupiała się aktywność zarówno instytucji centralnych UE, jak i organizacji zewnętrznych. W latach 2011–2015 pojawił się złożony wskaźnik AROPE, nowa miara powiązana z celem redukcji ubóstwa, jednak jego centralność pozostała wyraźnie niższa, a dyfuzja ograniczona głównie do struktur administracji UE. Wywiady i literatura przedmiotu pozwoliły wyjaśnić, że przyczyną mniej chętnego wykorzystywania go przez aktorów spoza administracji UE w porównaniu do AROP mogą być trudności interpretacyjne wynikające z jego konstrukcji łączącej trzy wymiary: redystrybucyjny, deprywacyjny i aktywizacyjny.

Wizualizacja sieci pozwoliła również uchwycić proces eksternalizacji ekspertyzy. Uwidoczniała to, co było trudno dostrzegalne z perspektywy badań jakościowych jak ten proces wiązał się z poszerzaniem pola instytucjonalnego – od jednostek Komisji Europejskiej po agencje, firmy konsultingowe i środowiska akademickie. Jednocześnie delegowanie produkcji wiedzy eksperckiej na zewnątrz buduje technokratyczną legitymizację w obszarze polityki społecznej.

Ta zróżnicowana koalicja aktorów wzmacnia ponadnarodowy wymiar polityki społecznej, nadając jej pozór neutralności i oparcia na dowodach.

Przemiany konfiguracji wskaźników pokazują, że unijna infrastruktura kalkulacyjna ewoluowała od pierwszych zestawów opartych na AROP, przez fazę rozbudowy i eksperymentów, aż do etapu konsolidacji, w którym pomiar został skupiony wokół trzech wymiarów – redystrybucyjnego, deprywacyjnego i aktywizacyjnego – zintegrowanych w złożonym wskaźniku AROPE. W ostatnich latach widoczny jest także rozwój nowego obszaru pomiaru, związanego z ubóstwem energetycznym, co wskazuje na dalsze poszerzanie zakresu unijnej polityki społecznej.

Pojawienie się nowego obszaru monitoringu, ubóstwa energetycznego, wskazuje na adaptację infrastruktury kalkulacyjnej do wyzwań transformacji klimatycznej. Badania jakościowe pozwoliły zidentyfikować je jako obszar dyskutowany przez poszczególnych aktorów i widoczny w dokumentach oraz wnioskować o pojawiającym się napięciu między celami społecznymi a środowiskowymi. Jednak analiza występowania wskaźnika w dokumentach ujawniła, że temat ten nie został jeszcze silnie zakorzeniony w strukturze wskaźnikowej i relacjach instytucjonalnych – występował raczej na obrzeżach sieci. Pokazuje to, że mimo rosnącej obecności dyskursywnej ubóstwo energetyczne nie osiągnęło jeszcze statusu ustabilizowanej kategorii kalkulacyjnej, co ogranicza jego polityczną sprawczość i potencjał interwencji.

Warunkiem przeprowadzenia tej i podobnych analiz jest utworzenie i odpowiednie przygotowanie korpusu. Praca z dużymi zbiorami danych tekstowych, takimi jak dokumenty unijne, wiąże się z wieloma wyzwaniami na różnych etapach procesu badawczego. Pierwszym wyzwaniem jest pozyskanie danych, które wymaga zaawansowanych technik automatyzacji, takich jak web scraping i integracja z interfejsami API, a także radzenia sobie z ograniczeniami w zakresie dostępnych metadanych. Kolejnym wyzwaniem jest przetwarzanie danych tekstowych, w tym integracja różnych formatów (PDF, HTML) w spójny korpus. Istotne wyzwanie stanowi również wzbogacenie metadanych. Aby tym wyzwaniom sprostać, należy wykorzystać zróżnicowane narzędzia i metody, odpowiednią infrastrukturę obliczeniową, a także polegać na otwartych bazach danych, co jest zadaniem nietrywialnym, zwłaszcza przy łączeniu danych pochodzących z różnych źródeł.

Opracowany przez nas przepływ pracy opiera się na zasadach FAIR oraz wykorzystaniu mechanizmów Linked Open Data, co gwarantuje możliwość ponownego użycia korpusu w różnych kontekstach badawczych. Proces rozpoczyna się od kwerendy i selekcji danych w oparciu o metadane, a następnie obejmuje ich wzbogacenie za pomocą zasobów takich jak Wikidata i EuroVoc. Wykorzystanie otwartych standardów i narzędzi umożliwia automatyzację i integrację danych, zapewniając ich spójność i wysoką jakość. Dzięki zastosowaniu

modelu 5-star Open Data kolejne etapy procesu – od pobierania dokumentów po wzbogacenie i udostępnianie danych – realizują kryteria otwartości i przygotowują korpus do ponownego wykorzystania. Wzbogacanie metadanych o informacje pochodzące z zewnętrznych źródeł wiedzy oraz zastosowanie wielopoziomowej klasyfikacji tematycznej umożliwia bardziej zaawansowane analizy i zwiększa wartość naukową zasobu. Na koniec warto podkreślić istotny wniosek dotyczący właściwego dla socjologii cyfrowej korzystania z analiz dużych zbiorów danych tekstowych. Wdrożenie takich metod badawczych wymaga tworzenia zespołów interdyscyplinarnych, które łączą kompetencje techniczne, wiedzę z zakresu lingwistyki i metod obliczeniowych, a także dogłębną znajomość badanych obszarów tematycznych. To wyzwanie, ale zarazem ogromna szansa. Z jednej strony współpraca taka wymaga znacznego nakładu pracy i dobrej komunikacji – przede wszystkim umiejętności przekładania różnych perspektyw i języków teoretycznych. Z drugiej – umożliwia prowadzenie autentycznie interdyscyplinarnych badań, które mogą przełamywać ograniczenia charakterystyczne dla poszczególnych dyscyplin i tym samym wносить nowe, innowacyjne treści do dyskursu naukowego.

Bibliografia

- Bandola-Gill, J., Grek, S., Tichenor, M. (2022). *Governing the Sustainable Development Goals: Quantification in Global Public Policy*. Springer Nature. <https://doi.org/10.1007/978-3-031-03938-6>
- Blom-Hansen, J. (2019). Studying Power and Influence in the European Union: Exploiting the Complexity of Post-Lisbon Legislation with EUR-Lex. *European Union Politics*, 20(4), 692–706. <https://doi.org/10.1177/1465116519851181>
- Boswell, C. (2008). The Political Functions of Expert Knowledge: Knowledge and Legitimation in European Union Immigration Policy. *Journal of European Public Policy*, 15(4), 471–488. <https://doi.org/10.1080/13501760801996634>
- Brennan, T. (2017). The Digital-humanities Bust: After a Decade of Investment and Hype, What Has the Field Accomplished? Not Much. *Chronicle of Higher Education*, 64(8). <https://www.chronicle.com/article/the-digital-humanities-bust/>
- Brosz, M., Bryda, G., Siuda, P. (2017). Big Data i CAQDAS a procedury badawcze w polu socjologii jakościowej. *Przegląd Socjologii Jakościowej*, 13(2), 6–23.
- Bruno, I., Jacquot, S., Mandin, L. (2006). Europeanization Through its Instrumentation: Benchmarking, Mainstreaming and the Open Method of Co-ordination... Toolbox or Pandora's Box? *Journal of European Public Policy*, 13(4), 519–536. <https://doi.org/10.1080/13501760600693895>
- Bruno, I. (2009). The "Indefinite Discipline" of Competitiveness Benchmarking as a Neoliberal Technology of Government. *Minerva*, 47(3), 261–280. <https://doi.org/10.1007/s11024-009-9128-0>

- Copeland, P., Daly, M. (2012). Varieties of Poverty Reduction: Inserting the Poverty and Social Exclusion Target into Europe 2020. *Journal of European Social Policy*, 22(3), 273–287. <https://doi.org/10.1177/0958928712440203>
- de la Porte, C., Pochet, P., Room, B. G. (2001). Social Benchmarking, Policy Making and New Governance in the EU. *Journal of European Social Policy*, 11(4), 291–307. <https://doi.org/10.1177/095892870101100401>
- Diefenbach, D., Wilde, M. D., Alipio, S. (2021). Wikibase as an infrastructure for knowledge graphs: The eu knowledge graph. W: International Semantic Web Conference (s. 631–647). Cham.
- Düro, M. (2009). *Crosswalking EUR-Lex: A Proposal for a Metadata Mapping to Improve Access to EU Documents*. Office for Official Publications of the European Communities.
- Golub, J. (1999). In the Shadow of the Vote? Decision Making in the European Community. *International Organization*, 53(4), 733–764. <https://doi.org/10.1162/002081899551057>
- Golub, J. (2007). Survival Analysis and European Union Decision-making. *European Union Politics*, 8(2), 155–179. <https://doi.org/10.1177/1465116507076428>
- Golub, J. (2023). EUPROPS: A New Dataset on Policymaking in the European Union from 1958 to 2021. *European Union Politics*, 25(1), 197–217. <https://doi.org/10.1177/14651165231202034>
- Gornitzka, Å., Sverdrup, U. (2015). Societal Inclusion in Expert Venues: Participation of Interest Groups and Business in the European Commission Expert Groups. *Politics and Governance*, 3(1), 151–165. <https://doi.org/10.17645/pag.v3i1.130>
- Grimmer, J., Roberts, M. E., Stewart, B. M. (2022). *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.
- Gábos, A., Goedemé, T. (2016). The Europe 2020 Social Inclusion Indicators: Main Conclusions of the ImPROvE Project on Validity, Methodological Robustness and Interrelationships. *ImPROvE Working Paper*, 16(13).
- Häge, F. M. (2007). Committee Decision-making in the Council of the European Union. *European Union Politics*, 8(3), 299–328. <https://doi.org/10.1177/1465116507079539>
- Hertz, R., Leuffen D. (2011). Too big to run? Analysing the impact of enlargement on the speed of EU decision-making. *European Union Politics* 12(2), 193–215.
- Hunt, D. (2021). Corpus Linguistics: Examining Tensions in General Practitioners' Views about Diagnosing and Treating Depression. W: G. Brookes, D. Hunt (red.), *Analysing Health Communication. Discourse Approaches* (s. 133–160). Palgrave Macmillan.
- Häge, F. M. (2008). Who Decides in the Council of the European Union? *Journal of Common Market Studies*, 46(3), 533–558.
- Häge, F. M. (2011). The European Union Policy-making Dataset. *European Union Politics*, 12(3), 455–477. <https://doi.org/10.1111/j.1468-5965.2008.00790.x>
- Janowicz, K., Hitzler, P., Adams, B., Kolas, D., Vardeman II, C. (2014). Five Stars of Linked Data Vocabulary Use. *Semantic Web*, 5(3), 173–176. <https://doi.org/10.3233/SW-140135>

- Karlińska A., Rosiński, C., Kubis, M., Hubar, P., Wieczorek, J. (2024). Using Bibliodata LODification to Create Metadata-enriched Literary Corpora in Line with FAIR Principles. W: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (s. 17271–17284). ELRA and ICCL.
- König, T., Luetgert, B., Dannwolf, T. (2006a). *EU-Lex: EU Legislative Databank Based on Celex/Prelex*. University of Mannheim. <http://www.sowi.unimannheim.de/lsp012/08downloads01.html>
- König, T., Luetgert, B., Dannwolf, T. (2006b). Quantifying European Legislative Research: Using Celex and Prelex in EU legislative studies. *European Union Politics*, 7(4), 553–574.
- Kurunmäki, L., Miller, P. (2013). Calculating Failure: The Making of a Calculative Infrastructure for Forgiving and Forecasting Failure. *Business History*, 55(7), 1100–1118. <https://doi.org/10.1080/00076791.2013.838036>
- Kurunmäki, L., Mennicken, A., Miller, P. (2019). Assembling Calculative Infrastructures. W: M. Kornberger i in. (red.), *Thinking Infrastructures*. Emerald Publishing Limited.
- König, T. (2007). Divergence or Convergence? From Ever-growing to Ever-slowng European Legislative Decision-making. *European Journal of Political Research*, 46, 417–444. <https://doi.org/10.1111/j.1475-6765.2007.00705.x>
- Lelie, P., Vanhercke, B. (2013). Inside the Social OMC's Learning Tools: How "Benchmarking Social Europe" Really Worked. *OSE Research Paper*, 10, 2–61. <https://doi.org/10.13140/RG.2.1.2419.7202>
- Louis, M., Maertens, L. (2021). *Why International Organizations Hate Politics: Depoliticizing the World*. Taylor & Francis. <https://doi.org/10.4324/9780429466984>
- Maitre, B., Nolan, B., Whelan, Ch. T. (2013). A Critical Evaluation of the EU 2020 Poverty and Social Exclusion Target: An Analysis of EU-SILC 2009.
- McEnery, T., Wilson, A. (2001). *Corpus Linguistics: An Introduction*. Edinburgh University Press.
- McEnery, T., Xiao, R., Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. Routledge.
- Mehrpouya, A., Samiolo, R. (2019). Indexal Thinking – Reconfiguring Global Topologies for Market-based Intervention. W: M. Kornberger i in. (red.), *Thinking Infrastructures*. Emerald Publishing Limited.
- Menyhárt, B., Cseres-Gergely, Z., Kvedaras, V., Mina, B., Pericoli, F., Zec, S. (2021). *Measuring and Monitoring Absolute Poverty (ABSPO) – Final Report* (JRC Technical Report No. JRC127444; EUR 30924 EN). Publications Office of the European Union. <https://publications.jrc.ec.europa.eu/repository/handle/JRC127444>
- Metz, J. (2013). Expert Groups in the European Union: A Sui Generis Phenomenon? *Policy and Society*, 32(3), 267–278. <https://doi.org/10.1016/j.polsoc.2013.07.007>
- Năstase, A., Radulova, E. (2025). When Consultants Come to Town: How the European Commission Justifies the Involvement of Private Contractors in Online Public Consultations. *International Review of Public Policy*, 7(2). <https://doi.org/10.4000/14mnk>

- Nolan, B., Whelan, Ch. T. (2011). *Poverty and Deprivation in Europe*. University Press.
- Pfiffner, N. (2021). Identifying Patterns in Communication Science: Mapping Knowledge Structures Using Semantic Network Analysis of Keywords. W: E. Segev (red.), *Semantic Network Analysis in Social Sciences*. Sage Publishing.
- Rauh, C. (2021). One Agenda-setter or Many? The Varying Success of Policy Initiatives by Individual Directorates-General of the European Commission 1994–2016. *European Union Politics*, 22(1), 3–24. <https://doi.org/10.1177/1465116520961467>
- Reilly, J., Scheytt, T. (2019). A Calculative Infrastructure in the Making: The Emergence of a Multi-layered Complex for Governing Healthcare. W: M. Kornberger i in. (red.), *Thinking Infrastructures*. Emerald Publishing Limited.
- Salganik, M. J. (2018). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.
- Szpunar, M. (2016). *Kultura cyfrowego narcyzmu*. Wydawnictwa AGH.
- Thedvall, R. (2006). *Eurocrats at Work: Negotiating Transparency in Postnational Employment Policy*. Almqvist & Wiksell International.
- Toshkov, D. (2017). The Impact of the Eastern Enlargement on the Decision-making Capacity of the European Union. *Journal of European Public Policy*, 24(2), 177–196.
- Veltri, G. A. (2020). Digital Sosocial researchResearch. Polity.
- Waagmeester, A., Stupp, G., Burgstaller-Muehlbacher, S., Good, B. M., Griffith, M., Griffith, O. L. (2020). Wikidata as a Knowledge Graph for the Life Sciences. *eLife*, 9. <https://doi.org/10.7554/eLife.52614>
- Wasserman, S, Faust, K. (2012). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Wiedemann, G. (2013). Opening up to Big Data: Computer-assisted Analysis of Textual Data in Social Sciences. *Forum Qualitative Sozialforschung Forum: Qualitative Social Research*, 14(2), 332–357. <https://doi.org/10.17169/fqs-14.2.1949>
- Wiedemann, G. (2015). *Text Mining for Qualitative Data Analysis in the Social Sciences*. Springer Berlin Heidelberg.