

ADAM RYBAK^a

**ANALIZA JAKOŚCI DOKUMENTACJI
METODOLOGICZNEJ SONDAŻY:
STUDIUM PRZYPADKU INWENTARYZACJI
30 PROGRAMÓW BADAŃ PORÓWNAWCZYCH
Z CAŁEGO ŚWIATA**

**SURVEY DOCUMENTATION QUALITY: A CASE STUDY
OF THE METHODOLOGICAL INVENTORY COVERING
CHARACTERISTICS OF SURVEYS FROM 30 COMPARATIVE
PROGRAMMES FROM AROUND THE GLOBE**

Methodological documentation serves as the primary source of information about sampling schemes, fieldwork procedures and related sources of potential measurement error in survey research. The quality of such documentation, including its completeness and standardization, is crucial for conducting methodological analyses based on inventories of multiple comparative survey programmes. Individual examples and quantitative summaries of historical development, as well as current methodological reports, serve as a foundation for reflecting on potential future analyses based on large methodological inventories. Therefore, the aim of this paper was to thoroughly analyse the gathered data, identify critical issues, and prepare proposals for a standardized report format suitable for automatic processing. The implementation of these proposals could enhance the capacity to conduct extensive methodological analyses. Given the increasing costs and labour intensity of survey research, both substantial and methodological data should be thoroughly utilized. The analysis is based on a methodological inventory of 10,493 surveys from 30 comparative programmes, spanning all inhabited continents. The paper demonstrates the challenges encountered during data collection and presents completeness indicators for both selected programmes and the complete dataset. The findings reveal that low levels of standardization between analysed methodological reports necessitate a critical evaluation of current practices. This situation impedes the development of future methodological inventories. The paper identifies critical barriers to creating quantitative comparisons of documentation quality and proposes solutions. Additionally, it presents a standardized report format addressing existing issues and potentially facilitating harmonization efforts.

Keywords: methodological documentation; comparative surveys; survey methodology; presenting research results

^a Adam Mickiewicz University, Poznań, Poland /
Uniwersytet im. Adama Mickiewicza w Poznaniu, Polska
adam.rybak@amu.edu.pl, <https://orcid.org/0000-0003-3981-1081>

Podstawowym źródłem informacji o schematach losowania próby i procedurach terenowych badań sondażowych oraz związanych z nimi potencjalnych źródłach błędów pomiaru jest dokumentacja metodologiczna. Jej szeroko pojęta jakość, której istotny element stanowią kompletność i standaryzacja, jest czynnikiem determinującym możliwość prowadzenia badań metodologicznych na dużych zbiorach danych pochodzących z międzykrajowych programów porównawczych. Za punkt wyjścia do poprawy jakości dokumentacji metodologicznej powinno służyć dokładne rozpoznanie dotychczasowego stanu rzeczy. Stanowi to pierwszy cel poniższego badania. Kolejnym celem była analiza problemów obecnych w tym obszarze, która umożliwiła sformułowanie zaleceń dążących do zwiększenia standaryzacji tworzonych raportów i tym samym możliwości lepszego wykorzystania danych pochodzących z badań sondażowych. Analiza zawarta w artykule opiera się na inwentaryzacji danych metodologicznych z 30 międzykrajowych porównawczych programów sondażowych z całego świata. Podsumowuję całość zbioru opartego na 10 493 realizacjach krajowych, poszczególne programy, a także przedstawiam problematyczne praktyki na podstawie konkretnych raportów metodologicznych. Przykłady i zestawienia ukazują rozwój i obecny stan jakości opracowanej dokumentacji. Efektem badania jest krytyczna ocena obecnych praktyk dokumentacyjnych, które ze względu na niewielki poziom standaryzacji pomiędzy projektami bardzo utrudniają praktyczną inwentaryzację informacji na temat przyjętych procedur badawczych. Przedstawiam zestawienia problematycznych praktyk oraz propozycję standaryzacji formy raportów, która rozwiązałaby większość trudności pojawiających się obecnie w procesie harmonizacji danych metodologicznych.

Słowa kluczowe: dokumentacja metodologiczna; międzykrajowe badania porównawcze; metodologia sondażowa; komunikowanie wyników badań

I. WPROWADZENIE

Podstawowym źródłem, z którego korzysta się w przypadku analiz praktyk sondażowych, jest dokumentacja metodologiczna. Badacz ma w pierwszym rzędzie dostęp do tych informacji, które osoby przygotowujące raporty postanowiły w nich zawrzeć. Jeśli nie podejmie dodatkowych wysiłków, jak próba osobistego kontaktu z koordynatorami krajowymi programów sondażowych, przeczesywanie surowych baz danych wynikowych w poszukiwaniu zmiennych metodologicznych czy studia monografii opisujących historię badań – będzie widział tyle, ile chciano mu pokazać. Artykuł ten jest efektem inwentaryzacji praktyk dokumentacyjnych z 30 porównawczych międzykrajowych programów sondażowych z całego świata.

Inwentaryzacja kompletności informacji zawartej w dokumentacji była w przypadku niniejszego badania pochodną inwentaryzacji praktyk sondażowych, co miało szereg konsekwencji. Jedną z nich było to, że dobór zmiennych opierał się na zestawie zadeklarowanych obszarów, które – opierając się na wynikach badań innych metodologów – mogłem określić jako mające wpływ na jakość pomiaru. Z tego powodu zakres danych, które znalazły się w stworzonej przeze mnie bazie kompletności dokumentacji sondaży, jest dosyć szeroki, ale podporządkowany konkretnemu typowi odbiorcy. Nie uwzględnia perspektywy innych grup zainteresowanych użytkowaniem danych, jak dziennikarze,

decydenci polityczni czy badacze zainteresowani np. paradata. Obszary, które zostały wyznaczone jako potencjalnie doniosłe, to liczba i motywowanie ankietów, kontakt z respondentami i ich motywowanie, różne aspekty doboru próby, oraz techniki sondażu.

Poza kilkoma arbitralnie pojętymi decyzjami o sięgnięciu głębiej, zbieranie danych ograniczyłem do informacji umieszczonych w repozytoriach i na stronach internetowych projektów. Przykładami takich podjętych przeze mnie decyzji mogą być: udany kontakt z zespołem Europejskiego Sondażu Społecznego w celu uzgodnienia przelicznika kategorii wywiadów niezrealizowanych na szerzej przyjęty standard (The American Association for Public Opinion Research, 2023); upewnienie się, jak często w Americas Barometer używano tzw. *frequency matching* (Castorena, 2021); czy incydentalne użycie informacji o technice badania zawartych w surowych danych wynikowych w celu ułatwienia kodowania. Nie znaczy to jednak, że inwentaryzacja danych była monotonna, informacje były przedstawione w najróżniejszy sposób, o czym mowa będzie dalej. W przeszłości analizy procedur dokumentacji badań sondażowych skupiały się w dużej mierze na wskazywaniu dobrych praktyk (Mohler i in., 2008, 2010), jednak absorpcja tych rozwiązań przez instytucje prowadzące badania wydaje się powolna.

W niniejszym artykule przyjmuję odwrotne podejście, skupiając się na krytycznej analizie stanu obecnego. Z opisu tego, jak rzeczywiście wyglądają raporty metodologiczne, wyprowadzam propozycje refleksji i ewentualnych ulepszeń. Efektem jest zestawienie zarówno konkretnych przykładów, jak i zbiorczych podsumowań dotyczących jakości dokumentacji w sondażowych programach porównawczych. Przedstawiam także zbiór problemów dotyczących możliwości ilościowych porównań jakości dokumentacji. Brak odpowiedzi na nie może hamować rozwój tego typu analiz.

1. Dostępne rozwiązanie problemu standaryzacji raportów metodologicznych

Dla dużej części problemów zawartych w tej pracy rozwiązanie teoretycznie już istnieje. To, dlaczego dotąd jest implementowane w bardzo ograniczonym zakresie, wymaga jednak zupełnie innego rodzaju studiów. Prawdziwa dostępność, a więc możliwość łatwego gromadzenia, przetwarzania, agregowania danych metodologicznych, wymaga, by nie miały one różnych formatów i form prezentacji graficznej, ale były zestandaryzowane. Idea takiej standaryzacji nie jest niczym nowym – projekt Data Documentation Initiative (DDI), utworzony w 1995 r. (Rasmussen, 2014; Vardigan, 2014; Vardigan i in., 2008), oferuje rozwiązanie, które jest otwarte i rozwinięte do stabilnej wersji 3.3 (Van Wettene, 2022). Opiera się ono na architekturze .xml, co umożliwia zautomatyzowany odczyt maszynowy.

Data Documentation Initiative umożliwia także udoskonalenie kontroli jakości w całym procesie przygotowania i realizacji badania, ale dla rozważań w niniejszym artykule interesujący jest jedynie jego potencjał do wy-

tworzenia łatwej w obróbce dokumentacji o stałych, zestandaryzowanych nagłówkach. W praktyce DDI nie stało się dotąd alternatywą dla dotychczasowych form dokumentacji metodologicznej. Tym samym nie można wykorzystać jego potencjału, prowadząc badania skupiające się na inwentaryzacji dużych zbiorów danych sondażowych ukierunkowanych na analizy metodologiczne. Nie potrafię wskazać przyczyny tego stanu rzeczy, w innych analizach wskazuje się np. mentalność izolacji (*silo mentality*) i strach przed zmianą (Fonteneau, 2011).

2. Przykłady obecnych w literaturze sposobów analizowania dokumentacji

Sposób oceny dokumentacji metodologicznej, który przedstawiam w tym artykule, nie jest ani jedynym, ani najlepszym z możliwych podejść. Jego niewątpliwym ograniczeniem jest fakt, że był podporządkowany procesowi inwentaryzacji praktyk badawczych, był więc środkiem, a nie celem.

Zanim omówię dalej własne podejście, przedstawię przykłady alternatyw. Najbardziej odmienne sposoby oceny jakości dokumentacji prezentują badania będące wynikiem projektu Survey Data Recycling ([SDR]; Dubrow i in., 2016; Ślomeczynski i in., 2017, 2022; Ślomeczynski i Tomescu-Dubrow, 2018). Badania te koncentrują się nie na samej zawartości dokumentacji, ale ocenie zgodności pomiędzy dokumentacją a danymi wynikowymi (np. zgodność między opisem zmiennych w kluczach kodowych i danych wynikowych), jak również jakością samych danych wynikowych (np. obecnością duplikatów obserwacji). W tym przypadku jest to związane z celem użycia tych danych ponownie, w formie zharmonizowanej, wzbogaconej o odpowiednie wskaźniki i flagi opracowane na podstawie znalezionych niezgodności i błędów powstałych przy opracowywaniu wyników.

Przykład propozycji sposobu analizowania dokumentacji, który częściowo pokrywa się z przedstawionym przeze mnie, ale jednocześnie proponuje analizę szerszej problematyki, przedstawił Jedinger z zespołem (2018). Wymieniony przez nich zestaw kluczowych pytań, jakie należy sobie zadać przy ocenie dokumentacji metodologicznej, zawiera, poza kwestiami obecnymi w moim badaniu, poszukiwanie w nich odpowiedzi na takie pytania jak: Jaki był cel zbierania danych? Jakie zbierano informacje? Jak dane były edytowane, kodowane i ważone? Jakie podjęto środki w celu ochrony danych osobowych? Takie podejście pozwala na bardziej całościową ocenę przydatności danych metodologicznych i staranności opracowywania wyników badań. Umożliwia także większe skupienie się na aspekcie dokumentacji jako sposobu komunikowania się zbierającego dane z ich potencjalnym użytkownikiem (Tepper i in., 2022). Nie zastosowałem go, gdyż byłoby nadmiarowe dla osiągnięcia celu opisu praktyk i stworzenia modeli metaanaliz reprezentatywności (Ry-

bak, 2023), które były głównym celem projektu będącego źródłem danych do niniejszego artykułu.

Wreszcie najbardziej zbliżony sposób działania zaprezentowano w analizach Kołczyńskiej (2014; Kołczyńska i Schoene, 2018), która jako pierwsza zaprezentowała koncepcję ilościowej oceny kompletności dokumentacji badań sondażowych, opierając się na danych z krajów Europy Południowo-Wschodniej. Badaczka uczestniczyła następnie zarówno w badaniach związanych ze wspomnianym projektem SDR, jak i w inspirowanych jej dokonaniach analizach zespołu skupionego wokół ośrodka poznańskiego (Jabkowski, 2019, 2023; Jabkowski i Kołczyńska, 2020; Jabkowski i in., 2021). Podstawową różnicą między podejściem cytowanych badaczy a moim jest fakt, że autorzy ci podjęli się prób porównania kompletności pomiędzy modułami zmiennych w obrębie programu oraz położyli silny nacisk na porównania pomiędzy programami, ograniczając krytykę dostępnej dokumentacji wyłącznie do aspektu jej kompletności. To podejście wydaje się częściowo problematyczne ze względu na dramatyczną różnicę w formie i standaryzacji dokumentów pomiędzy programami, jak również na różną wagę poszczególnych informacji z dokumentacji dla odbiorcy końcowego. Te fragmenty niniejszego artykułu, które opierają się na podsumowaniach kompletności zmiennych dotyczących informacji zawartych w dokumentacji, kontynuują podejście zaproponowane przez wymienionych badaczy.

Istotnym dla omawianej problematyki wydarzeniem jest przedstawienie przez łączoną komisję American Association for Public Opinion Research [AAPOR] i World Association of Public Opinion Research [WAPOR] raportu dotyczącego jakości sondażowych badań porównawczych (AAPOR/WAPOR, 2021). Przedstawiono w nim rozdział dotyczący dokumentacji, który porusza między innymi kwestie dotyczące zakresu tematów badanych w tym artykule. Jego zakres jest jednak szerszy. Przykładowo, poniższe analizy planowo pomijają kwestie przygotowania, testowania i tłumaczenia kwestionariuszy, które są istotne dla wielu badaczy. W wytycznych raportu wskazano także problemy dostępności różnych rodzajów para- i metadata przydatnych do wtórnych analiz. Z punktu widzenia tego artykułu interesujące są wytyczne dotyczące *study-level-metadata*, czyli informacji o konkretnych sondażach krajowych w ramach projektu.

Tabela 1 zawiera porównanie elementów, których obecność sprawdzałem w swojej analizie z wytycznymi z raportu. Warto zaznaczyć, że informacje o tym, czy sondaż był częścią panelu, były pomocniczo kodowane tylko dla tej niewielkiej części programów, w których panowała bardzo duża dowolność metodologiczna i panele rzeczywiście występowały, nie posiadam więc wiedzy o obecności tej informacji dla całej bazy. Pomiąłem także w całości temat wag statystycznych, gdyż planowane metaanalizy dotyczące reprezentatywności nie miały posługiwać się nimi ze względu na dotychczasowe ustalenia badaczy (Jabkowski i in., 2021).

Tabela 1

Porównanie zaleceń zawartych w Załączniku 6 raportu (AAPOR/WAPOR, 2021)
ze zmiennymi zawartymi w inwentaryzacji

Zalecenia AAPOR/WAPOR	Informacje w bazie o:	Zgodność
Finansowanie		–
Instytucja badawcza		–
Cele badania		–
Przekrojowe/podłużne	(zbierane dla niektórych programów, nieuwzględnione w bazie dokumentacji)	–
Populacja badana	minimalnym wieku	+
	maksymalnym wieku	+
Jednostka obserwacji	(we wszystkich tylko jednostka)	+
Dobór próby	operacje	+
	doborze wewnątrzgrupowym	0
	substytucjach/kwotach	+
Zachęty materialne	zachętach materialnych	+
	listach zapowiednich	0
Technika	technice	+
Kwestionariusz	długości wywiadu	0
Testowanie kwestionariusza		–
Tłumaczenie kwestionariusza		–
Ankieterzy	sposobie wynagradzania ankieterów	0
	liczbie ankieterów	+
Procedury terenowe	minimalnej liczbie prób kontaktu	0
	backcheckach	0
Response rate	liczbie wywiadów	0
	dowolnym RR	0
	umożliwiająca obliczenie RR6 AAPOR	+
Wagi	(nieuwzględnione celowo)	–
Czas i miejsce	(dostępne dla wszystkich)	+

+ pełna zgodność, 0 – zgodność tematyczna, – brak zgodności

Źródło: opracowanie własne.

II. DANE

Uwzględnione sondaże pochodzą ze 179 krajów ze wszystkich zamieszkałych kontynentów (z uwzględnieniem krajów współcześnie nieistniejących, jak Jugosławia, Związek Sowiecki, Czechosłowacja czy NRD, oraz nieuznawanych, jak Tajwan). Zakres danych podsumowuje tabela 2.

Punktem wyjścia selekcji badań do analizy była lista porównawczych sondaży międzykrajowych wykonana i udostępniona w Internecie przez GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS, 2019). Lista ta nie jest już udostępniana przez tę instytucję.

Tabela 2

Liczebności i zakres dat programów badawczych uwzględnionych w analizie

Program	Liczba sondaży	Pierwszy i ostatni rok badań uwzględnionych w bazie
AB – Afrobarometer	205	1999–2021
ABS – Asian Barometer Survey	69	2001–2021
ARAB – Arab Barometer	63	2006–2022
ASES – Asia Europe Survey	18	2001
ASIAB – Asia Barometer	34	2005–2008
CCEB – Candidate Countries Eurobarometer	142	2001–2004
CCS – Civic Culture Study	5	1959
CEEB – Central and Eastern Eurobarometer	120	1990–1997
CNEP – Comparative National Elections Project	61	1990–2021
CONS – Consolidation of Democracy in Central and Eastern Europe	24	1990–2001
CRRC – Caucasus Barometer	30	2006–2020
CSES – Comparative Study of Electoral Systems	209	1996–2020
EASS – East Asia Social Survey	24	2006–2018
EB – Eurobarometer	4 604	1974–2019
ECS – European Communities Study	20	1970–1973
EES – European Election Studies	174	1989–2019
EQLS – European Quality of Life Surveys	125	2003–2016
ESS – European Social Survey	243	2002–2022
EUNB – EU Neighbourhood Barometer	96	2012–2014
EVS – European Values Study	147	1981–2020
GALLUP – Gallup World Poll	2 280	2006–2021
INTUNE – Integrated and United	35	2007–2009
ISSP – International Social Survey Programme	961	1985–2022
ISSP_SOR – Study of Religion of Additional 14 Countries, based on ISSP 2018	14	2018
LAPOP – Latin American Public Opinion Project (Americas Barometer)	275	1973–2021
LITS – Life in Transition Survey	98	2006–2016
MEDW – Making Electoral Democracy Work	10	2010–2016
NBB – New Baltic Barometer	12	1993–2000
NEB/NDB – New Europe Barometer/New Democracies Barometer	94	1991–2007
WVS – World Values Survey	301	1981–2021
Razem	10 493	1959–2022

Źródło: opracowanie własne.

Kryteriami, które decydowały o obecności programu badawczego w stworzonej bazie, były:

1) obecność w zestawieniu programów porównawczych (GESIS, 2019) – dostępne w archiwum autora i w narzędziu Internet Archive Wayback Machine (<https://web.archive.org/>);

2) anglojęzyczna dokumentacja w wolnym dostępie internetowym (z tego powodu wykluczyłem duży program Latinobarometro, który swoją dokumentację publikuje konsekwentnie wyłącznie po hiszpańsku);

3) niekomercyjny charakter (poza GALLUP);

4) bazy danych wyników w wolnym dostępie internetowym (poza GALLUP i ASES);

5) włączyłem tylko projekty badające (niemal) całą populację kraju, to jest nie uwzględniłem tych obejmujących wyłącznie osoby starsze, młodzież, miasta albo wybrany region kraju. (Kilka uwzględnionych i dodatkowo oznaczonych przypadków badań obejmujących tylko wybrany obszar kraju zostało w większości wychwyconych na etapie kontrolnym. Te, w których mniejszościowy obszar kraju był wyłączony z badania na skutek wojny, powstania, albo klęki żywiołowej, kodowałem normalnie);

6) uwzględniłem tylko badania, które dotyczyły (co najmniej w części kwestionariusza) jednostki, a nie, przykładowo, gospodarstwa domowego jako całości;

7) nie uwzględniłem „programów” będących tylko harmonizacją *post factum* wcześniejszych badań.

III. METODOLOGICZNA INWENTARYZACJA SONDAŻY: KRYTYCZNA ANALIZA FORMUŁY RAPORTÓW

1. Jakość samej formy publikacji metodologicznych – przykłady

Dokumentacja metodologiczna sondaży rodzi się wtedy, kiedy one same – w pierwszej połowie XX w. (Mohler i in., 2008), a więc w czasach maszynopisu. Z punktu widzenia jakości samego zapisu powinniśmy się spodziewać wyłącznie postępu – do typografii cyfrowej. Jednak jest wprost przeciwnie, część zgromadzonych przeze mnie kwestionariuszy metodologicznych wypełniona jest na maszynie, ale wiele nadal odręcznie długopisem (CSES1 w 1996 r. w Holandii, na Tajwanie, w Czechach, Norwegii, Ukrainie; CSES2 w 2001 r. Norwegii – rys. S1 [Rybak, 2024b]¹; WVS2 z 1990 w Chinach – rys. S2 [Rybak, 2024b] – ma postać wyrwanej z żółtego notatnika kartki zapisanej ołówkiem, z tej samej fali dla Chile ma postać krzywego wydruku z drukarki igłowej).

Kolejne przykłady trudnych w odczytaniu raportów to np. raport WVS4 z Iraku pokazujący tabelkę, która nie mieści się na kartce i jej część pozostaje dla odbiorcy nieznaną. Kwestionariusz metodologiczny z WVS4 w Ugandzie także wypełniony jest długopisem. Raport LITS1 ma tabele w postaci obrazków tak skompresowanych, że liczby są nieczytelne (zob. rys. S3; Rybak, 2024b). Większość dokumentacji ISSP przedstawiona jest w formie centralnie

¹ Wszystkie ryciny, do których odwołania znajdują się w tekście, dostępne są w Suplemencie zdeponowanym w repozytorium (Rybak, 2024b), dlatego odwołania do nich poprzedzone są literą S.

zaplanowanego raportu dla całej fali, ale już na przykład CSES i WVS mają zazwyczaj osobny kwestionariusz metodologiczny dla każdego kraju, który potrafi być wypełniony, jeśli nie odręcznie, to różną czcionką w różnym kolorze, stosując różne symbole zaznaczenia konkretnej opcji („iksy”, „ptaszki”, „krzyże”, wypełnienie całego pola jednym kolorem, wypełnienie gradientem, zostawienie pola pustego, a zakreślenie na żółto tekstu, który stoi przy polu (tak np. Kirgistan w WVS6, tutaj także adnotacje cyrylicą w anglojęzycznym kwestionariuszu, zob. rys. S4; Rybak, 2024b). Wspomniany ISSP w fali 2020 zrezygnował z jednolitego raportu i także wprowadził pełne kwestionariusze metodologiczne dla krajów – ustawił tam jednak pola tekstowe stałej wielkości, dlatego tekst je wypełniający jest ucinany.

Bardzo wiele dokumentów ma postać skanów: dokumentacja EES99 posiada po bokach ślady otworów po dziurkaczu, wszystkie fale CEEB to krzywo zeskanowane kartki. Nawet kiedy dokument ma formę cyfrowej typografii, to potrafi mieć różny format, od najpopularniejszego PDF, przez doc (np. CNEP10 z Dominikany), arkuszy Excel (lista wszystkich EB albo większości CNEP), po txt (całościowa dokumentacja fal CSES, w odróżnieniu od osobnych kwestionariuszy dla krajów). Na rycinach w Suplemencie (Rybak, 2024b) przedstawiam reprodukcje kilku przykładów.

2. Ujednolicenie treści publikacji metodologicznych – przykłady

Istotne dla jakości dokumentacji są jej kompletność i jednolitość, ściśle ze sobą powiązane. Rozdzielenie dokumentacji sondażu na kilka plików znacznie utrudnia ich zinwentaryzowanie, a jednocześnie może wprowadzać niezgodności w opisie pomiędzy dokumentami. Przykładowo, większość fal EB posiada jeden raport w repozytorium GESIS, drugi, różny, w repozytorium Komisji Europejskiej i jeszcze inny dokument w formacie Excel (GESIS Eurobarometer Data Service, 2023), który zawiera daty prac fieldworku czy wielkości próby wszystkich badań ECS, EB i CEEB. Co więcej, informacje potrafią się różnić między tymi dokumentami.

Wspomniany CSES (Comparative Study of Electoral Systems, 2023) jest częściowo omówiony metodologicznie w plikach PDF „design” dla krajów, a częściowo w pliku txt dla całej fali. Część badań CNEP znajduje swoje omówienie we wspomnianym skrócie Excel (The Ohio State University, 2024), reszta jest natomiast opisana w osobnych dokumentach na tej samej stronie. W dokumentacji EVS nie zgadzają się liczebności zrealizowanej próby pomiędzy raportami metodologicznymi w repozytorium (European Values Study, 2023) a umieszczonym tam również plikiem Excel, zatytułowanym „EVS 1981–2017 Method Overview”, np. w fali 2008 dla Francji, w fali 1999 np. dla Austrii czy Chorwacji. Repozytorium WVS (World Values Survey, 2023) dla wielu krajów zawiera pliki, które nazywają się raportami technicznymi, kwestionariuszami metodologicznymi, *study designs* itd., których liczba dla każdego sondażu jest zmienna, a zawartość częściowo się pokrywa, a częściowo uzupełnia.

Wreszcie problem jednolitości i kompletności łączy się z problemem trwałej archiwizacji. Strona i jednocześnie repozytorium AB do lipca 2022 r. udostępniały manuale – a więc jedynie deklaracje odnośnie do charakterystyk ba-

danian, fiszki techniczne dla danych krajów, które zawierały tylko podstawowe charakterystyki, oraz (w osobnym pliku) szczegółowe wskaźniki kompletności, odmów, kontaktów i kooperacji, które jednak nie dawały się przeliczyć na liczby naturalne, gdyż były przybliżeniami ułamków. Po przebudowie strony duża część tych dokumentów zniknęła z repozytorium.

Te wyliczenia nie mają za zadanie przedstawienia kompletnego zbioru wszystkich błędów i nieścisłości, ale pokazanie możliwie realistycznie, w jakim stanie znajduje się dokumentacja najważniejszych światowych projektów i jak wygląda praca naukowa z jej wykorzystaniem.

3. Pytania przed analizą ilościową

Należy się także zastanowić, czy w opisanej sytuacji „pomiar” jakości dokumentacji jest w ogóle możliwy, a jeśli tak, to jakie problemy stawia przed nami takie przedsięwzięcie. Pojawiają się następujące pytania:

a) Jak traktować różnice pomiędzy projektami, które dopuszczają pewną różnorodność, np. w kwestii techniki sondażu? Czy dokument metodologiczny, który stwierdza, że w każdym badanym kraju przeprowadzono wywiad kwestionariuszowy, jest istotnie lepszym źródłem informacji niż taki, który opisuje stosowane w różnych krajach różne techniki, ale koordynator w jednym z tych krajów zaniedbał podanie odpowiedniej informacji?

Uwagi: O ile w obecnej sytuacji informacje podawane zbiorczo (nie tylko w sytuacji jednej techniki dla programu, ale też zbiorczych tabel, jak w ISSP) ułatwiają kodowanie manualne, o tyle w sytuacji upowszechnienia standardu łatwego do kodowania automatycznego zalecana byłaby osobna tabela dla każdej realizacji krajowej.

b) Czy program sondażowy, który w tzw. manualu opisuje intencję zastosowania jakiegoś rozwiązania należy traktować tak samo, jak program, który w raporcie z pracy terenowej zaświadcza, że to rozwiązanie zastosował?

Uwagi: Literatura zaleca (AAPOR/WAPOR, 2021) prowadzenie dokumentacji na bieżąco, odbiorcę końcowego interesuje zazwyczaj stan finalny i on przede wszystkim powinien być opisany.

c) Uwzględnienie których charakterystyk w dokumentacji jest konieczne? Przykładowo, czy jeśli cały program sondażowy nie stosuje zachęt materialnych dla respondentów i z racji tego nie wspomina o tym dokumentacja, to jest ona gorszej jakości? Czy konieczne jest udokumentowanie liczby pakietów kwestionariuszy przesłanych respondentom pocztą (dostępne np. w części raportów ISSP)? Czy liczba ankietów pracujących w terenie jest niezbędną informacją?

Uwagi: Możliwa jest inspiracja z innej dyscypliny, sugerująca uwzględnianie wszystkich względnie często spotykanych rozwiązań. Przy pisaniu wyników diagnozy medycznej sugeruje się zawrzeć informacje o wszystkich obszarach, które składały się na badanie, nie tylko o tym, co jest anormalne (Raffle i Gray, 2007). Użycie schematów (*templates*) diagnozy, które uzupełnia

się w trakcie badania, poprawia jakość dokumentacji (Mohammad i in., 2021; Rose i in., 2001). Na tej samej zasadzie dokumentacja sondażowa powinna zawierać komplet informacji o przeprowadzonym badaniu, nie tylko o metodach użytych, lecz także tych, których nie użyto (zachęty materialne, listy, substytucje itd.).

d) Jak traktować różnice pomiędzy programami, które opisują falę w jednej publikacji, takimi, które mają osobną publikację dla każdego kraju, a takimi, które mają kilka publikacji zbiorczych, z których każda zawiera tylko część informacji o interesujących nas zmiennych?

Uwagi: Ze względu zarówno na archiwizację, dokumentowanie źródeł, jak i potencjał do stosowania algorytmów automatycznego zbierania danych wskazane jest stosowanie jednej publikacji dla jednej fali.

e) Jak rozróżniać stopień czytelności dokumentów metodologicznych, jeśli wiemy, że część z nich jest wypełniona długopisem, zawiera rozpixselowane tabele albo wstawki w innych językach?

Uwagi: Czytelność pisma odręcznego i grafik w bardzo niskiej rozdzielczości jest subiektywna.

f) Jak traktować różnicę pomiędzy programem, którego dokumentacja nie zawiera jakiejś informacji, a takim, która zawiera, ale w dwóch sprzecznych wersjach? Albo jest podana, ale w oczywisty sposób błędnie (casus CSES 3 i 4 w Tajlandii, z której dokumentacji wynikało jednoznacznie, że wskaźnik realizacji został nieprawidłowo obliczony)?

Uwagi: Przy braku innych przesłanek w badaniach praktyk stosowałem zapisywanie losowej wartości. Przy czystej analizie dokumentacji należałoby uznać to za brak danych.

g) Po ustaleniu katalogu niezbędnych do wyszczególnienia charakterystyk należy przejść do przypisania każdej z nich odpowiedniej wagi, wszak wskaźnik realizacji jest zazwyczaj bardziej istotną informacją o jakości badania niż średnia długość wypełniania kwestionariusza.

Uwagi: W dotychczasowej analizie (Rybak, 2023) związków pomiędzy charakterystykami badania a reprezentatywnością stosowałem podstawowy zestaw wymaganych zmiennych: technika, operat, dobór wewnątrzgrupowy, kwoty/substytucje, *response rate*. Inni odbiorcy mogą definiować swoje wymagania minimalne zupełnie inaczej, producent dokumentacji powinien wziąć pod uwagę różnych użytkowników i zdefiniować ich potrzeby. To samo podejście można zastosować przy analizie.

4. Analizowane zmienne i ich kompletność

Tabela 3 zawiera opis i podsumowanie kompletności zakodowanych zmiennych metodologicznych dla wszystkich sondaży łącznie. Co zaskakujące, najkompletniej udokumentowaną charakterystyką sondażu była technika. Po

niej następne w kolejności były podstawowe informacje o doborze próby, jak operat, zakres wieku badanej populacji czy użycie substytucji i kwot demograficznych. Bardziej szczegółowa, ale nie mniej istotna informacja o sposobie doboru wewnątrzgrupowego była obecna nie w około 90%, jak poprzednie, ale w niecałej 1/4 raportów. Informacja dosyć niewielkiej wartości o samej tylko liczbie wywiadów była obecna także w około 90% raportów, ale już uważane zazwyczaj za najbardziej podstawowy wskaźnik jakości badania współczynniki realizacji zostały zawarte w około 15% dokumentów.

Informacja o często analizowanych przez metodologów kwestiach użycia listów zapowiednich i zachętach materialnych przeznaczonych dla badanych znajduje się w niecałych 15% raportów. W tym wypadku jednak krytyczna analiza może spotkać się z odpowiedzią, że być może autorzy dokumentacji nie czuli się zobowiązani do wspomnienia tych charakterystyk sondażu, które nie były obecne w ich badaniu. Poruszyłem ten temat w sekcji III.3 artykułu. Nie można w ten sposób podsumować jednak wspólnych dla wszystkich badań przeprowadzonych techniką wywiadu kwestionariuszowego problemów liczby i sposobu wynagradzania ankietów, które mogą mieć wpływ na jakość zebranych danych, a były opisane w kolejno około 5% i 10% analizowanych raportów.

Nie przedstawiam obliczeń kompletności dla kategorii tematycznych złożonych z kilku zmiennych. Wynika to z faktu, że, przykładowo, dla badacza chcącego ocenić reprezentatywność badań informacja o zastosowaniu doboru kwotowego na poziomie gospodarstwa domowego będzie bardziej istotna niż to, czy operat był imienny czy adresowy. Natomiast dla osoby chcącej skorzystać z uzyskanych danych substancjalnych zakres wieku populacji, z której losowana jest próba, może być ważniejszy niż wszystkie inne charakterystyki procesu doboru. Problemem jest więc nadanie poszczególnym zmiennym w obrębie kategorii wag.

Tabela 3

Opis zmiennych uwzględnionych w analizie i podsumowanie ich dostępności w bazie

Nazwa	Opis: Czy jest dostępna informacja o...	Liczebności	Procent sondaży w bazie z dostępnymi informacjami
mod_info	technice	Nie: 137 Tak: 10 356	0,987
loage_info	minimalnym wieku	Nie: 432 Tak: 10 061	0,959
upage_info	maksymalnym wieku	Nie: 450 Tak: 10 043	0,957
fra_info	operacie	Nie: 630 Tak: 9 863	0,94

Tab. 3 *cd.*

Nazwa	Opis: Czy jest dostępna informacja o...	Liczebności	Procent sondaży w bazie z dostępnymi informacjami
int_info	liczbie wywiadów	Nie: 997 Tak: 9 496	0,905
subquota_info	substytucjach/kwotach	Nie: 1631 Tak: 8 862	0,845
cont_info	minimalnej liczbie prób kontaktu	Nie: 6362 Tak: 4 131	0,394
long_info	długości wywiadu	Nie: 7209 Tak: 3 284	0,313
hhselection_info	doborze wewnątrzspołowym	Nie: 7584 Tak: 2 909	0,277
check_info	backcheckach	Nie: 8535 Tak: 1 958	0,187
outcome_info	umożliwiająca obliczenie RR6 wg AAPOR	Nie: 8843 Tak: 1 650	0,157
rr_info	dowolnym RR	Nie: 9014 Tak: 1 479	0,141
lett_info	listach zapowiednich	Nie: 9056 Tak: 1 437	0,137
inc_info	zachętach materialnych	Nie: 9165 Tak: 1 328	0,127
paid_info	sposobie wynagradzania ankierów	Nie: 9410 Tak: 1 083	0,103
work_info	liczbie ankierów	Nie: 9887 Tak: 606	0,058

Źródło: opracowanie własne.

W tabeli 4 przedstawiam średnie wartości kompletności dla zmiennych w poszczególnych obszarach – z wyraźnym zastrzeżeniem, że dane te należy potraktować krytycznie. Sekcje III.1 i III.2 artykułu pokazują, że różnice w sposobach dokumentowania badań w różnych programach sondażowych są często tak wielkie, że ciężko przyłożyć do nich jedną miarę. Dodatkowo różne programy były realizowane w różnych latach. Okres, w którym była tworzona dokumentacja, ma istotny wpływ na jej kształt.

Można zauważyć, że z programów o zasięgu globalnym najkompletniejszą (pod względem założonych kryteriów) dokumentację posiadają CSES, ISSP i (będący jedną falą) ISSP_SOR. Wśród projektów obejmujących, z niewielkimi wyjątkami, kontynent europejski za najlepiej udokumentowane należy uznać EQLS, ESS i EVS. Dodatkowo ESS jest najlepiej udokumentowanym programem ze wszystkich zbadanych.

Tabela 4

Kompletność dokumentacji metodologicznej według programu

Programy	mod	loage	upage	fra	int	sub quota	cont	long	hselect	check	outcome	rr	lett	inc	paid	work
AB	1	1	1	1	0,4	0,7	0,3	1	0,3	0,7	0	0,6	0,3	0,3	0	0
ABS	1	0,2	0,2	0,3	0,3	0,2	0,1	0,9	0,2	1	0	0,1	0	0	0	0,2
ARAB	1	1	1	1	1	0	0	1	0,9	1	0	0,5	0	0	0	0
ASES	1	1	1	1	0	0,1	0,2	0	0,9	0,1	0	0	0	0	0	0
ASIAB	1	1	1	1	1	0,1	0,1	0	0,7	0,1	0,1	0	0	0	0	0
CCEB	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
CCS	1	1	1	1	1	1	0,2	0	1	0	0,8	1	0	0	0	0
CEEB	1	1	1	0	1	1	0	0	0,1	0	0	0	0	0	0	0
CNEP	1	0,2	0,2	0,8	1	0,1	0	0	0,6	0	0	0,8	0	0	0	0,1
CONS	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0
CRRC	1	0,8	0,8	1	1	0,1	0,7	0,4	1	0,6	0,1	0,8	0,2	0	0	0,3
CSSES	1	0,9	0,9	0,9	0,9	0,7	0,8	0	0,9	0,6	0,7	0,7	0,9	0,9	0	0,3
EASS	1	0,8	0,8	0,8	1	0	0	0	0	0	0,5	1	0	0	0	0
EB	1	1	1	1	0,9	1	0	0	0,1	0	0	0	0	0	0	0
ECS	1	1	1	0,5	1	0,3	0	0	0,2	0	0	0	0	0	0	0
EES	1	1	1	0,3	0,9	0,1	0,6	0,4	0,4	0,2	0,2	0,4	0	0	0	0,6
EQLS	1	1	1	0,8	1	1	1	0,8	0,8	0,8	1	0,7	1	0,3	0	1
ESS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
EUNB	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0
EVS	1	1	1	0,9	1	0,7	0,7	1	0,6	0,7	0,7	0,7	0,5	0,5	0,7	0
GALLUP	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
INTUNE	1	1	1	1	1	0,1	0,8	0	0,9	0	0	0,9	0	0	0	0
ISSP	1	0,9	0,9	0,9	0,9	0,8	0,7	0	0,9	0,8	0,8	0,4	0,7	0,6	0,6	0
ISSP_SOR	0,9	0,9	0,9	0,9	0,9	0,9	0,6	0	0,9	0,9	0,9	0	0,9	0,9	0,9	0
LAPOP	1	0,6	0,6	0,9	1	0,6	0,1	0	0,6	0,3	0	0,1	0	0,1	0	0
LITS	1	1	1	1	1	0,4	1	0	1	0	0,3	0,7	0,4	0,4	0	0
MEDW	0,7	0,7	0,3	0	0,7	0	0,7	0	0,7	0	0,7	0,7	0,7	0,7	0	0
NBB	1	1	1	1	1	0,2	0,4	0	1	0,3	0,5	0,9	0	0	0	0
NEB/NDB	0,5	0,4	0,5	0,5	0,5	0	0	0	0	0,4	0	0	0	0	0	0
WVS	0,9	0,9	0,9	0,8	0,9	0,7	0,6	0,1	0,8	0,6	0,4	0,2	0,1	0	0,5	0,2

Źródło: opracowanie własne.

Tabela 5 zawiera wskaźniki średniej kompletności dokumentacji sondaży z czterech wybranych dekad. Dane te przynoszą wniosek umiarkowanie pozytywny: wskaźniki kompletności rosną w kolejnych okresach. Jest to zgodne z wcześniejszymi ustaleniami badaczy (Jabkowski, 2023; Kołczyńska i Schoene, 2018). Największy wzrost dostępnych informacji dotyczy opisu operatu losowania i informacji o substytucjach i kwotach, a więc ogólnie tematu doboru próby. Jest to ważne, gdyż taka informacja umożliwia wstępną ocenę reprezentatywności danych wynikowych dla badanej populacji.

Udostępniam zbiór wszystkich 10 493 sondaży z przypisanymi im zmiennymi zawierającymi informację o kompletności wszystkich raportów w poszczególnych obszarach (Rybak, 2024a). Można go wykorzystać do dowolnych porównań pomiędzy programami bądź okresami, należy jednak pamiętać o wymienionych w niniejszym artykule problemach z tym związanych. Badacze mogą także użyć tej bazy w celu selekcji do własnych analiz sondaży o dokumentacji wysokiej jakości. Należy jednak pamiętać, że kompletność raportów metodologicznych nie musi się wiązać z reprezentatywnością badania (Jabkowski, 2023).

Tabela 5

Kompletność dokumentacji metodologicznej według dekady i kontynentu
(dla kontynentu dane od roku 1990–)

Zmienne	Dekada				Kontynent					
	1980– 1989	1990– 1999	2000– 2009	2010– 2019	Afryka	Azja	Europa	Amer. Łac.– Karaiby	Ameryka Północna	Oceania
mod	0,96	0,96	0,98	1	0,99	0,99	0,99	0,99	0,95	0,99
loage	0,93	0,9	0,96	0,98	0,99	0,93	0,98	0,84	0,79	0,91
upage	0,93	0,89	0,96	0,97	0,99	0,93	0,98	0,84	0,79	0,91
fra	0,83	0,81	0,95	0,97	0,99	0,93	0,94	0,96	0,88	0,98
int	0,93	0,92	0,88	0,91	0,83	0,89	0,91	0,98	0,97	0,93
sub quota	0,5	0,74	0,83	0,92	0,84	0,77	0,89	0,82	0,83	0,87
cont	0	0,08	0,39	0,47	0,76	0,68	0,27	0,63	0,6	0,8
long	0,07	0,06	0,24	0,41	0,92	0,59	0,17	0,45	0,28	0,32
hhselect	0,04	0,24	0,32	0,28	0,22	0,32	0,27	0,38	0,59	0,68
check	0	0,15	0,22	0,18	0,25	0,25	0,16	0,24	0,41	0,48
outcome	0	0,15	0,22	0,14	0,04	0,18	0,17	0,13	0,4	0,56
rr	0,01	0,12	0,15	0,15	0,17	0,16	0,14	0,09	0,29	0,28
lett	0	0,03	0,19	0,15	0,11	0,14	0,14	0,11	0,33	0,47
inc	0	0,03	0,16	0,14	0,11	0,13	0,13	0,14	0,34	0,44
paid	0	0,04	0,15	0,1	0,05	0,14	0,1	0,12	0,18	0,13
work	0	0,03	0,06	0,07	0,01	0,04	0,07	0,03	0,02	0,15

Źródło: opracowanie własne.

Kołczyńska i Schoene (2018), analizując dokumentację 1721 sondaży z programów porównawczych, doszli do wniosku, że najlepiej udokumentowane badania pochodzą z Europy, a najgorzej z Ameryki Łacińskiej i Bliskiego Wschodu. Dane przedstawione w tabeli 5, ograniczone do lat 1990+ nie potwierdzają tej diagnozy (dane dla całej bazy dają identyczne wnioski). Dokumentacja w sondażach europejskich wykazuje najmniejszą kompletność, biorąc pod uwagę założone kryteria. Najlepiej udokumentowane sondaże były przeprowadzane w Oceanii i Ameryce Północnej. Zły wynik sondaży europejskich może wynikać z tego, że przeprowadza się tam najwięcej sondaży, z których duża część (jak np. bardzo liczebny EB) posiada dokumentację niskiej jakości. Należy jednak raz jeszcze przypomnieć o tym, że dla odbiorcy końcowego nie wszystkie informacje mają jednakową wagę.

5. Analiza dokumentacji na poziomie programu

Niniejsze rozważania można uzupełnić o ilościowe i jakościowe spojrzenie na cztery projekty: ISSP, AB, CSES i ESS. W dwóch pierwszych widać bowiem wyraźny spadek jakości dokumentacji w ostatnich falach, który nie jest aż tak łatwy do uchwycenia na podstawie samych podsumowań ilościowych, natomiast jest od razu widoczny po spojrzeniu na konkretne raporty metodologiczne. Z kolei dwa kolejne to przykłady o wiele stabilniejszej tendencji w konstruowaniu dokumentacji, choć w swojej formie zupełnie różne.

To, co można zauważyć w przypadku dokumentacji ISSP (tab. 6), to fakt, że od roku 2009 raporty metodologiczne zaczynają uwzględniać *response rate* oraz że ogólna kompletność raportów poprawia się od 2012. Duży spadek kompletności w okolicach 2009 r. wynika z tego, że kraje, które zgodnie z opisem z repozytorium brały udział w fali, nie były uwzględnione w głównym raporcie. Nie znaczy to, że ich raporty nie mogą być ukryte w ramach innej publikacji, dociekliwość koderów jest jednak ograniczona. W dwóch ostatnich falach widzimy jedynie zniknięcie informacji o substytucjach, kwotach i *response rate*. Kiedy ogląda się same raporty, wrażenie przez nie wywierane jest zdecydowanie inne.

Zupełnie nieprzydatny do zautomatyzowanej inwentaryzacji, ale bardzo czytelny i przejrzysty format raportu, który stosowany był konsekwentnie do roku 2017, w kolejnej fali ustępuje miejsca tabelom o podobnym układzie, ale pozbawionym wszelkich udogodnień wizualnych. Dramatycznie zwiększa to potencjał błędów przy ręcznym tworzeniu inwentaryzacji. W dwóch ostatnich falach natomiast otrzymujemy, wspomniany wcześniej, a popularny wśród innych projektów format surowego kwestionariusza metodologicznego wypełnianego przez pracownika krajowej instytucji badawczej. Kodowanie charakterystyk z takiego formatu zajmuje zdecydowanie więcej czasu, a ze względu na brak realnej standaryzacji odpowiedzi na pytania kwestionariuszowe zrzuca obowiązek ich uzgodnienia na badacza. W tym przypadku występuje też wspomniany wcześniej problem techniczny: okna odpowiedzi udzielonych w aplikacji komputerowej przy tworzeniu wydruku do pliku PDF nie rozszerzyły się do rozmiaru tekstu. Tym samym część odpowiedzi jest ucięta. (Poszczególne formy raportów są przedstawione na rysunkach S5, S6 i S7; Rybak, 2024b).

Tabela 6

Kompletność dokumentacji fal według programu (ISSP tylko 2005–2020)

Programy i ich fale	mod	loage	upage	fra	int	sub qu- ota	cont	long	hselect	check	outcome	rr	lett	inc	paid	work
ISSP05	0,97	0,97	0,97	0,97	0,97	0,97	0,94	0	0,97	0,97	0,97	0	0,97	0,97	0,69	0
06	1	0,97	0,97	1	1	0,97	0,97	0	1	0,97	1	0	0,97	0,97	0,71	0
07	1	0,94	0,94	1	1	0,94	0,89	0	1	0,94	0,97	0	0,94	0,94	0,78	0
08	1	0,95	0,95	0,98	1	0,9	0,83	0	0,95	0,93	1	0	0,85	0,93	0,75	0
09	0,76	0,76	0,76	0,76	0,78	0,76	0,76	0	0,73	0,76	0,76	0,76	0,76	0,76	0,63	0
10	0,91	0,88	0,88	0,94	0,94	0,85	0,82	0	0,94	0,76	0,94	0,76	0,79	0,76	0,61	0
11	0,88	0,84	0,84	0,84	0,88	0,81	0,81	0	0,84	0,81	0,88	0,78	0,81	0,81	0,59	0
12	1	1	1	1	1	1	0,95	0	1	1	1	1	1	1	0,76	0
13	1	1	1	1	1	1	0,97	0	1	1	1	1	0,97	1	0,79	0
14	1	1	1	1	1	0,97	1	0	0,97	1	0,97	1	1	1	0,74	0
15	1	1	1	1	1	1	1	0	0,97	1	0,92	0,92	1	1	0,78	0
16	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
17	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
18	1	1	1	1	1	1	0,97	0	1	1	1	1	1	1	1	0
19	1	1	1	1	1	0,07	1	0	1	1	1	0	0,93	0,97	0,93	0
20	1	1	1	1	1	0,07	1	0	1	0,93	1	0	1	0,93	0,93	0
AB1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0
2	1	1	1	1	0	1	0	1	0	0	0	0,94	0	0	0	0
3	1	1	1	1	0	1	0	1	0	0	0	0,94	0	0	0	0
4	1	1	1	1	1	1	0	1	0	0	0	1	0	0	0	0

Tab. 6 *cd.*

Programy i ich fale	mod	loage	upage	tra	int	sub qu- ota	cont	long	hselect	check	outcome	rr	lett	inc	paid	work
5	1	1	1	1	0,97	1	1	1	1	1	0	1	1	1	0	0
6	1	1	1	1	0,97	1	1	1	1	1	0	0,97	1	1	0	0
7	1	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0
8	1	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0
CSES1	0,9	0,78	0,8	0,76	0,73	0	0,71	0	0,76	0,02	0,54	0,61	0,83	0,83	0	0,15
2	0,98	0,9	0,9	0,98	0,88	0,29	0,83	0	0,98	0,05	0,66	0,76	0,9	0,9	0	0,22
3	0,96	0,94	0,94	0,94	0,88	0,96	0,86	0	0,94	0,9	0,78	0,62	0,94	0,94	0	0,34
4	1	1	1	0,98	0,85	1	0,89	0	0,98	0,91	0,81	0,68	1	1	0	0,36
5	1	1	1	1	0,93	0,97	0,8	0	1	0,93	0,9	0,7	1	1	0	0,3
ESS1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,95	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	0,97	1	1	0,97	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	0,95	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	0	1	1	1	1	0	0,9	1	1	1	1

Źródło: opracowanie własne.

W przypadku AB sytuacja jest jeszcze bardziej skomplikowana, ponieważ w trakcie inwentaryzacji doszło do przebudowy strony internetowej tego projektu, w trakcie której zniknął dostęp do części dokumentacji. Udało się ją jednak wcześniej zarchiwizować, dlatego w tabeli 6 można zauważyć, że fale 5 i 6 były zdecydowanie lepiej udokumentowane metodologicznie niż wcześniejsze i późniejsze.

W falach 5 i 6 stosowano zestandaryzowany format raportu, który choć nie bardzo rozbudowany, pozwala łatwo wydobyć jednolity zestaw informacji o realizacjach w poszczególnych krajach. Pokazuje to rysunek S8 (Rybak, 2024b). Dodatkowe informacje dla tych fal znajdowały się w manualach (jednak dokument taki powstaje przed realizacją jako zestaw zaleceń, wątpliwe jest więc traktowanie go jako opisu realnych działań terenowych) oraz w tabelach umieszczonych w skróście Excel, zawierających obliczone już wskaźniki realizacji.

W przypadku CSES można zauważyć, przedstawiony w tabeli 6, stabilny wzrost dokładności opisu badania w kolejnych falach. Wynika on przede wszystkim z rozbudowy kwestionariuszy metodologicznych (rys. S9; Rybak, 2024b) wypełnianych przez instytucje z krajów uczestniczących w programie. Taki format był też stosowany np. w niemal całym WVS. Ma on pewne przewagi nad innymi formatami: po pierwsze – pozwala na dosyć dokładny opis charakterystyk sondażu przez instytucję krajową, a po drugie – wymaga najmniej pracy od koordynatora programu, który musi jedynie przygotować kwestionariusz. Z drugiej strony brak zaangażowania koordynatora w opracowanie zebranych danych metodologicznych powoduje, że cierpi standaryzacja danych. Mówiąc wprost, osoby wypełniające kwestionariusze, po pierwsze, różnie rozumieją pytania, a po drugie, mają dużą tendencję do opuszczania niektórych z nich. Z tej i innych wspomnianych powyżej (jak graficzne różnice w udzielaniu odpowiedzi) przyczyn raczej nie ma tu możliwości zastosowania żadnych procedur zautomatyzowanego gromadzenia danych metodologicznych.

Ostatnim przykładem jest ESS – program, który bez wahania można uznać za modelowy dla sondażowych badań porównawczych. Dokumentacja metodologiczna jest tu niezwykle szczegółowa, jak pokazuje tabela 6 (zob. też rys. S10; Rybak, 2024b). W zestawieniu ilościowym praktycznie nie widać żadnych istotnych różnic w jakości dokumentacji. Pewne braki w fali 10 wynikają z tego, że w momencie tworzenia inwentaryzacji dostępna była jedynie część danych. Nie oznacza to jednak, że różnice nie występują – widoczne są jednak na poziomie tak szczegółowym, że niemożliwym do uchwycenia przez stosowany schemat inwentaryzacji. W fali 9 zmienia się także sposób kodowania kategorii osób biorących (niebiorących) udział w badaniu. Nowy podział kategorii utrudniał jasne przełożenie ich na kategorie AAPOR, konieczny był więc kontakt z instytucjami ESS. Można tu zauważyć pewną niekonsekwencję, gdyż w przypadku inwentaryzacji innego programu taka sytuacja mogłaby hipotetycznie spowodować zakodowanie tego jako braku danych. Moją motywacją była tu jednak duża ilość bardzo dobrej jakości danych metodologicznych, których dostarcza ESS. Sposób konstrukcji dokumentacji ESS, gdyby tylko zastosowano format bardziej przyjazny niż pliki PDF, byłby potencjalnie najbardziej przystosowanym do łatwej automatyzacji zbierania danych metodologicznych.

IV. WNIOSKI

Podstawowym ograniczeniem niniejszej analizy jest fakt, że była ona środkiem do osiągnięcia innego celu. Sposób doboru kodowanych zmiennych przynajmniej częściowo podporządkowałem zamiarowi opisu konkretnych charakterystyk i stworzenia statystycznych modeli badających związki pomiędzy nimi a miarami reprezentatywności. Warto poszerzyć analizowane zmienne, przykładowo, o kwestie związane z odbiorem informacji przez jej użytkowników (a więc takie jak cele badania i związek pomiędzy nimi a jego charakterystykami).

Obecnie gromadzenie danych metodologicznych wymaga ręcznej pracy wielu koderów nieuchronnie podatnych na wprowadzanie błędów do tworzonych przez siebie inwentaryzacji, które to błędy nakładają się na te powstałe w trakcie tworzenia samej dokumentacji. Wprowadzenie takiego jak DDI lub podobnego rozwiązania uczyniłoby proces gromadzenia danych metodologicznych o wiele szybszym, mniej pracochłonnym i bardziej odpornym na błędy. Jeśli jednak DDI sprawia szczegółowe techniczne problemy w wypadku jakiegoś projektu, to już wprowadzenie jednolitego, zestandaryzowanego dla wszystkich fal skoroszytu, w którym z łatwością można rozpoznawać tekst, tabele, sumować dane i używać rozmaitych makr, dałoby takiej dokumentacji istotną przewagę nad plikami PDF, które na takie operacje nie pozwalają albo pozwalają w niewielkim tylko stopniu.

Potencjalnie owocnym podejściem mogłoby być przygotowanie jednego, czytelnego i atrakcyjnego wizualnie standardu tabeli o stałych, zestandaryzowanych nagłówkach, która zastąpiłaby obecne często dosyć dowolnie konstruowane kwestionariusze metodologiczne. Wzorem może tutaj być tabela kategorii osób, które nie uczestniczyły w badaniu, przygotowana przez AAPOR (2023), tabela ta została dosyć powszechnie przyjęta, nawet jeśli poszczególne projekty stosują równolegle inne klasyfikacje. Najlepiej byłoby jednak, gdyby taka tabela była powielana w formie lepiej nadającej się do automatycznej obróbki niż pliki PDF, na przykład w arkuszu kalkulacyjnym. W przypadku upowszechnienia dokumentacji w takiej formie należałoby spodziewać się także stworzenia łatwych w obsłudze dla wszystkich kategorii użytkowników końcowych aplikacji i skryptów open-source do ich gromadzenia i obróbki.

Pozytywne wyniki dotyczące wzrostu kompletności dokumentacji metodologicznej badań sondażowych potwierdzają wcześniejsze ustalenia badaczy. Wzrost ten następuje jednak bardzo powoli. W obecnym stanie rzeczy ilościowe porównywanie dokumentacji pomiędzy programami sondażowymi wciąż wydają się problematyczne ze względu na drastyczne różnice w ich formie i czytelności, jak też liberalne podejście do wypracowanych dotąd standardów. Nie zmienia to faktu, że krytyczny namysł nad stanem dokumentacji badawczej jest istotny dla procesu dalszego jej udoskonalania. Motywowanie zespołów badawczych do poprawy jakości ich własnych raportów, a także koordynowania tych wysiłków z całym środowiskiem byłoby najcenniejszym osiągnięciem tego artykułu.

- Author contributions / Indywidualny wkład autora (CRediT):** Adam Rybak – 100% (Conceptualization / Konceptualizacja; Data Curation / Zarządzanie danymi; Formal analysis / Formalna analiza; Funding acquisition / Pozyskanie funduszy; Investigation / Przeprowadzenie badań; Methodology / Metodologia; Project administration / Administracja projektu; Validation / Walidacja; Visualization / Wizualizacja; Writing – original draft / Pisanie – pierwszy szkic; Writing – review & editing / Pisanie – recenzja i edycja).
- Conflict of interest / Konflikt interesów:** The author declares no conflict of interest. / Autor nie zgłosił konfliktu interesów.
- Funding / Finansowanie:** National Science Centre (Poland), grant Preludium 17 no. 2019/33/N/HS6/00322. / Narodowe Centrum Nauki, Preludium 17, umowa nr 2019/33/N/HS6/00322.
- The use of AI tools / Wykorzystanie narzędzi AI:** The author declares no use of AI tools. / Autor oświadczył, że nie korzystał z narzędzi AI.
- Data availability / Dostępność danych:** A full data set of the completeness of the information contained in the documentation for each national implementation is available in the OSF repository. / Pełna baza danych zawierająca informacje o kompletności zawartych w dokumentacji informacji dla poszczególnych realizacji krajowych dostępna w repozytorium OSF. <https://doi.org/10.17605/OSF.IO/85CK9>
- Acknowledgement / Podziękowania:** I would like to thank Prof. Piotr Jabkowski for his support in conducting my research. / Dziękuję prof. Piotrowi Jabkowskiemu za wsparcie w realizacji moich badań.
- Supplementary material / Dodatkowe materiały:** Rybak, A. (2024a), Baza danych – Kompletność dokumentacji sondaży [Baza Danych], OSF, <https://doi.org/10.17605/OSF.IO/A5WBN>. Wszystkie ryciny, do których odwołania znajdują się w tekście: Rybak, A. (2024b), *Ryciny do artykułu* [plik PDF], OSF, <https://doi.org/10.17605/OSF.IO/9VPDM>. Spis wszystkich repozytoriów i stron internetowych, z których pobrane zostały informacje na temat charakterystyk sondaży: Rybak, A. (2024c), *Spis odnośników do baz danych* [plik PDF], OSF, <https://doi.org/10.17605/OSF.IO/DG2VF>

References / Bibliografia

- AAPOR/WAPOR. (2021). *AAPOR/WAPOR task force report on quality in comparative surveys*. American Association for Public Opinion Research. (2023). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (wyd. 10). AAPOR.
- Castorena, O. (2021). *Methodological note #007: Survey weights in AmericasBarometer data*. Vanderbilt University.
- Comparative Study of Electoral Systems. (2023). *CSES download data and documentation*. <https://cses.org/data-download/download-data-documentation/>
- Dubrow, J. K., Jenkins, J. C., Kolczyńska, M., Oleksiyenko, O., Powalko, P., Słomczyński, K. M., Tomescu-Dubrow, I., Wymułek, I., i Zieliński, M. W. (red.). (2016). *Democratic values and protest behavior: Harmonization of data from international survey projects*. IFIS PAN Publishers.
- European Values Study. (2023). *European values study: Methodology, data, documentation*. <https://europeanvaluesstudy.eu/methodology-data-documentation/>
- Fonteneau, F. (2011). *Experiences and benefits of implementing the DDI in data management*. UNESCAP Expert Group Meeting.
- GESIS. (2019). *Overview of comparative surveys worldwide*. <https://web.archive.org/web/20191211072538/https://www.gesis.org/angebot/daten-analysieren/weitere-sekundaerdaten/uebersichten/overview-of-comparative-surveys-worldwide>

- GESIS Eurobarometer Data Service. (2023). *Eurobarometer fieldwork dates and institutes, sample sizes, and geographic units by participating countries*. https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/eurobarometer/eb_standard/eb_countries-over-time.xlsx
- Jabkowski, P. (2019). Zróżnicowanie praktyk sondażowych w międzykrajowych badaniach porównawczych na podstawie archiwizacji 1537 surveyów zrealizowanych w krajach europejskich w latach 1981–2017. *Przegląd Socjologiczny*, 68(1), 9–38. <https://doi.org/10.26485/PS/2019/68.1/1>
- Jabkowski, P. (2023). Increase in the quality of methodological documentation of cross-national pan-European multi-wave surveys over the last 40 years – a research note. *International Journal of Social Research Methodology*, 26(6), 817–824. <https://doi.org/10.1080/13645579.2022.2097394>
- Jabkowski, P., Cichocki, P., i Kołczyńska, M. (2021). Multi-Project Assessments of sample quality in cross-national surveys: The role of weights in applying external and internal measures of sample bias. *Journal of Survey Statistics and Methodology*, 11(2), 1–24. <https://doi.org/10.1093/jssam/smab027>
- Jabkowski, P., i Kołczyńska, M. (2020). Sampling and fieldwork practices in Europe: Analysis of methodological documentation from 1,537 surveys in five cross-national projects, 1981–2017. *Methodology*, 16(3), 186–207. <https://doi.org/10.5964/meth.2795>
- Jedinger, A., Watteler, O., i Förster, A. (2018). Improving the quality of survey data documentation: A total survey error perspective. *Data*, 3(4), 1–10. <https://doi.org/10.3390/data3040045>
- Kołczyńska, M. (2014). Representation of Southeast European countries in international survey projects: Assessing data quality. *Ask: Research and Methods*, 23(1), 57–78.
- Kołczyńska, M., i Schoene, M. (2018). Survey data harmonization and the quality of data documentation in cross-national surveys. W: *Advances in comparative survey methods* (s. 963–984). <https://doi.org/10.1002/9781118884997.ch44>
- Mohammad, L. M., Vidalis, B. M., Fatemi, L., Coffman, R. R., Qeadan, F., i Kimmell, K. T. (2021). Use of standardized history and physical examination for neurosurgical patients improves clinical documentation and reimbursement. *World Neurosurgery*, 148, e667–e673. <https://doi.org/10.1016/j.wneu.2021.01.056>
- Mohler, P. Ph., Hansen, S. E., Pennell, B.-E., Thomas, W., Wackerow, J., i Hubbard, F. (2010). A survey process quality perspective on documentation. W: J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B.-E. Pennell, i T. W. Smith (red.), *Survey methods in multinational, multiregional, and multicultural contexts* (s. 299–314). John Wiley & Sons.
- Mohler, P. Ph., Pennell, B.-E., i Hubbard, F. (2008). Survey documentation: Toward professional knowledge management in sample surveys. W: E. D. de Leeuw, J. J. Hox, i D. A. Dillman (red.), *International handbook of survey methodology* (s. 403–420). L. Erlbaum Associates.
- Ohio State University. (2024). *Comparative national elections project: Publicly available surveys*. <https://u.osu.edu/cnep/surveys/surveys-through-2012/>
- Raffle, A. E., i Gray, J. A. M. (2007). *Screening: Evidence and practice*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199214495.001.0001>
- Rasmussen, K. B. (2014). Social science metadata and the foundations of the DDI. *IASSIST Quarterly*, 37(1–4), 28. <https://doi.org/10.29173/iq499>
- Rose, E. A., Deshikachar, A. M., Schwartz, K. L., i Severson, R. K. (2001). Use of a template to improve documentation and coding. *Family Medicine*, 33(7), 516–521.
- Rybak, A. (2023). Survey mode and nonresponse bias: A meta-analysis based on the data from the international social survey programme waves 1996–2018 and the European social survey rounds 1 to 9. *PLOS ONE*, 18(3). <https://doi.org/10.1371/journal.pone.0283092>
- Rybak, A. (2024a). Baza danych – Kompletność dokumentacji sondaży [Baza Danych]. OSF. <https://doi.org/10.17605/OSF.IO/A5WBN>
- Rybak, A. (2024b). Ryciny do artykułu [plik PDF]. OSF. <https://doi.org/10.17605/OSF.IO/9VPDM>
- Rybak, A. (2024c). Spis odnośników do baz danych [plik PDF]. OSF. <https://doi.org/10.17605/OSF.IO/DG2VF>

- Slomczynski, K. M., Powalko, P., i Krauze, T. (2017). Non-unique records in international survey projects: The need for extending data quality control. *Survey Research Methods*, 11(1), 1–16. <https://doi.org/10.18148/srm/2017.v11i1.6557>
- Slomczynski, K. M., i Tomescu-Dubrow, I. (2018). Basic principles of survey data recycling. W: *Advances in Comparative Survey Methods* (s. 937–962). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118884997.ch43>
- Slomczynski, K. M., Tomescu-Dubrow, I., i Wysmulek, I. (2022). Survey data quality in analyzing harmonized indicators of protest behavior: A survey data recycling approach. *American Behavioral Scientist*, 66(4), 412–433. <https://doi.org/10.1177/00027642211021623>
- Tepper, S. J., Spruill, M. K., Premachandra, B., i Lewis, N. A. (2022). Surveys as conversations between makers and takers: A conversational framework for assessing and responding to community needs. *Analyses of Social Issues and Public Policy*, 22(3), 857–875. <https://doi.org/10.1111/asap.12326>
- Van Wettene, N. (2022, 12 stycznia). *How to implement the metadata standard „Data Documentation Initiative (DDI)”?* <https://doi.org/10.5281/zenodo.5841664>
- Vardigan, M. (2014). DDI Timeline. *IASSIST Quarterly*, 37(1–4), 51. <https://doi.org/10.29173/iq502>
- Vardigan, M., Heus, P., i Thomas, W. (2008). Data documentation initiative: Toward a standard for the social sciences. *International Journal of Digital Curation*, 3(1), 107–113. <https://doi.org/10.2218/ijdc.v3i1.45>
- World Values Survey. (2023). *World values survey: Documentation for download*. <https://www.worldvaluessurvey.org/WVSContents.jsp>

