Porównywalność wskaźników Smart City – koncepcja oparta na metadanych służąca weryfikacji i harmonizacji porównań między miastami

Kevin Weigel, Phillip Preischhoff
DHBW Heidenheim

Streszczenie

Kluczowe wskaźniki efektywności (KPI) stały się w badaniach nad inteligentnymi miastami podstawowym narzędziem operacjonalizacji celów strategicznych oraz prowadzenia porównań między miastami. Pomimo międzynarodowych inicjatyw normalizacyjnych, takich jak normy ISO 37120, ISO 37122 i ISO 37123, porównywalność w praktyce często pozostaje jedynie domyślnym założeniem. Nie towarzyszy jej systematyczna weryfikacja strukturalnej równoważności konkretnych zbiorów danych. Niniejszy artykuł odpowiada na tę lukę badawczą poprzez opracowanie opartych na metadanych ram porównywalności, w których porównywalność ujmowana jest jako wielowymiarowy warunek niezmienności i operacjonalizowana za pomocą określonych reguł.

Istotą proponowanego podejścia jest ustrukturyzowana karta wskaźnika KPI, która wprost uwzględnia wymiary semantyczne, przestrzenne, czasowe, metodologiczne oraz związane z normalizacją. Na tej podstawie wprowadzono trzypoziomową klasyfikację: C1 – porównywalne, C2 – porównywalne z ograniczeniami oraz C3 – nieporównywalne. Klasyfikacja ta opiera się na udokumentowanych metadanych i zdefiniowanych regułach decyzyjnych. Porównywalność nie jest więc wyprowadzana wyłącznie z nazw wskaźników czy jednostek miary, lecz traktowana jako weryfikowalna właściwość konkretnych par wskaźników KPI.

Zastosowanie modelu zilustrowano na przykładzie studium przypadku obejmującego trzy publicznie dostępne wskaźniki KPI: emisję CO_2 na mieszkańca, liczbę ofiar śmiertelnych wypadków drogowych oraz czynsz netto za metr kwadratowy powierzchni mieszkalnej w Berlinie, Paryżu i Barcelonie. O ile analiza oparta wyłącznie na formalnej zgodności wskaźników klasyfikuje wszystkie trzy miary jako bezpośrednio porównywalne, o tyle studium przypadku pokazuje, że żaden z badanych wskaźników nie spełnia w pełni warunku niezmienności. Jeden wskaźnik KPI sklasyfikowano jako porównywalny z ograniczeniami, a dwa pozostałe jako nieporównywalne. Liczba porównań dopuszczalnych bez zastrzeżeń zmniejsza się tym samym z trzech do zera.

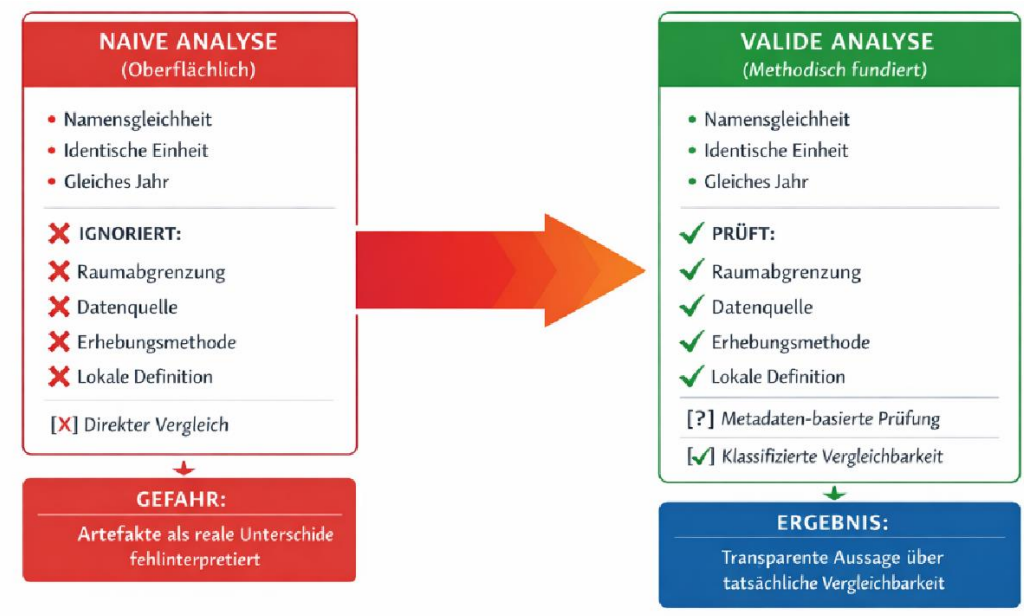
Artykuł stanowi teoretyczne i metodologiczne rozszerzenie istniejących ram wskaźników KPI, ponieważ ujmuje porównywalność jako wyraźny warunek trafności na poziomie danych. Opracowane ramy zwiększają przejrzystość, powtarzalność i rzetelność metodologiczną analiz międzymiastowych. Mogą być również wykorzystywane jako wstępny etap kontroli jakości w rankingach smart city oraz systemach benchmarkingu.

1. Wprowadzenie

W badaniach nad inteligentnymi miastami kluczowe wskaźniki efektywności (KPI) stały się podstawowym narzędziem operacjonalizacji strategicznych celów miast oraz pomiaru postępów. Służą one zarządzaniu wewnętrznemu, rozliczalności zewnętrznej oraz komunikowaniu poziomu rozwoju miejskiego. Szczególnie w złożonych procesach transformacji systemy wskaźników pozwalają w uporządkowany sposób opisywać innowacje technologiczne, zrównoważony rozwój i skutki społeczne (Huovila i in., 2017, s. 114–115). Ramy takie jak CITYkeys mają zapewnić przejrzystą metodologię oraz wspierać porównywalność między miastami poprzez systematyzację wskaźników i ułatwianie wymiany informacji między interesariuszami (Huovila i in., 2017, s. 115).

Wraz z rosnącym wykorzystaniem wskaźników KPI do porównań między miastami na pierwszy plan wysuwa się jednak zasadniczy problem ich wiarygodności. Różnice między miastami są często interpretowane jako rzeczywiste różnice wyników, mimo że nie zawsze sprawdza się, czy porównywane wskaźniki opierają się na strukturalnie równoważnych zbiorach danych. Wiele systemów KPI projektowano jako rozwiązania możliwe do zastosowania w bardzo zróżnicowanych miastach. Taka ogólność ułatwia ich szerokie wykorzystanie, ale jednocześnie nie ujawnia wprost różnic między lokalnymi definicjami,

odniesieniami przestrzennymi i praktykami gromadzenia danych (Huovila i in., 2019, s. 151). Ponadto wiele podejść opartych na KPI koncentruje się silniej na sektorowych lub technologicznych wskaźnikach wyników, podczas gdy wymiary oddziaływania, takie jak jakość życia, zarządzanie czy spójność społeczna, pozostają relatywnie słabiej uwzględnione (Huovila i in., 2019, s. 150; Serrano i in., 2022, s. 1). W analizach międzymiastowych oznacza to, że sama zbieżność nazw lub identyczność jednostek nie wystarcza, aby uznać, że porównywane wartości mierzą ten sam konstrukt w porównywalnych warunkach (Sterner i in., 2024, s. 748; Eurostat, 2021, s. 147). Schematycznie przedstawiono to na rysunku 1.



Rysunek 1: Analiza naiwna a analiza trafna1

Źródło: Opracowanie własne

Problem ten dotyczy nie tyle wyboru priorytetów merytorycznych, ile poprawności wnioskowania (Sterner i in., 2024, s. 748). Porównywalność oznacza stopień, w jakim różnice między wartościami można przypisać rzeczywistym różnicom w badanym konstrukcie, a nie artefaktom wynikającym z odmiennych definicji, granic przestrzennych lub metod pomiaru (Eurostat, 2021, s. 147). Jeśli warunek ten nie jest spełniony, rankingi, benchmarki i korelacje tracą jednoznaczność interpretacyjną. Porównywalność nie jest zatem trywialną cechą wskaźników, lecz metodologicznym warunkiem rzetelnych analiz międzymiastowych (Milfont i Fischer, 2010, s. 112; Sterner i in., 2024, s. 748).

Międzynarodowe inicjatywy normalizacyjne wyznaczają aktualny stan praktyki, zwłaszcza normy ISO 37120:2018, ISO 37122:2019 i ISO 37123:2019. Norma ISO 37120 definiuje wskaźniki służące pomiarowi usług miejskich i jakości życia w kontekście zrównoważonego rozwoju (ISO, 2018), norma ISO 37122 określa wskaźniki służące ocenie inteligentnych miast (ISO, 2019a), natomiast norma ISO 37123 opisuje wskaźniki odporności miejskiej (ISO, 2019b). Normy te ujednolicają terminologię oraz wytyczne pomiarowe i obliczeniowe, a ich celem jest poprawa porównywalności wyników (Huovila i in., 2019, s. 151; ISO, 2018). Jednocześnie porównywalność jest w nich ujmowana przede wszystkim na poziomie definicji wskaźników i ogólnych zaleceń, a nie jako operacyjny problem decyzyjny dotyczący konkretnych zbiorów danych. W przypadku korzystania z heterogenicznych źródeł otwartych danych pozostaje więc pytanie, jak systematycznie identyfikować rozbieżności w zakresie granic przestrzennych, źródeł danych lub metod gromadzenia informacji oraz jak klasyfikować ich wpływ na porównywalność.

W tym kontekście artykuł podejmuje trzy pytania badawcze. Po pierwsze, w jaki sposób można modelować porównywalność miejskich wskaźników KPI jako właściwość konkretnych zbiorów danych możliwą do formalizacji. Po drugie, jak oparte na metadanych i regułach ramy decyzyjne mogą identyfikować oraz przejrzyście klasyfikować odchylenia strukturalne między zbiorami danych KPI. Po trzecie, w jakim stopniu wyraźna weryfikacja porównywalności zmienia wyniki analizy międzymiastowej w porównaniu z założeniem bazowym, które wyprowadza porównywalność z identyczności nazwy wskaźnika, jednostki

miary i roku odniesienia.

Wkład niniejszej pracy ma trzy wymiary. Po pierwsze, porównywalność zostaje ujęta jako cecha relacyjna konkretnych par wskaźników KPI, a nie jako właściwość samego wskaźnika. Po drugie, koncepcję tę zoperacjonalizowano w postaci ustrukturyzowanej karty wskaźnika KPI oraz regułowej logiki decyzyjnej, która dokumentuje odchylenia za pomocą kodów uzasadniających, zwiększając przejrzystość i powtarzalność analizy. Po trzecie, skuteczność modelu pokazano na podstawie studium przypadku, zestawiając bazowe założenie bezpośredniej porównywalności z wynikami weryfikacji opartej na metadanych.

Z teoretycznego punktu widzenia podejście to nawiązuje do zasady niezmienności pomiarowej, zgodnie z którą porównywalność jest warunkiem trafności porównań międzygrupowych (Milfont i Fischer, 2010, s. 111; Sterner i in., 2024, s. 747). W niniejszym artykule nie testuje się jednak ukrytej niezmienności modelu w sensie statystycznym. Przedmiotem analizy jest raczej niezmienność procesów danych, definicji odniesienia oraz ram metodologicznych na poziomie konkretnych zbiorów danych (Eurostat, 2021, s. 13; Sterner i in., 2024, s. 755). W rozdziale 4 opracowano ramy porównywalności oparte na metadanych, w rozdziale 5 oceniono ich działanie na podstawie studium przypadku, w rozdziale 6 omówiono ograniczenia i implikacje, a rozdział 7 zawiera podsumowanie i perspektywy dalszych badań.

2. Stan badań i problem porównywalności

We współczesnych badaniach nad inteligentnymi miastami systemy wskaźników, określane jako kluczowe wskaźniki efektywności (Key Performance Indicators), stały się podstawowym narzędziem ilościowego opisu złożonych procesów transformacji miejskiej oraz wspierania decyzji w zarządzaniu miastami. Wskaźniki KPI umożliwiają uporządkowane rejestrowanie zmian w kluczowych obszarach, takich jak gospodarka, środowisko i jakość życia, a także przekładanie celów strategicznych na mierzalne wielkości operacyjne (Huovila i in., 2019, s. 142). Ramy oparte na KPI porządkują istotne informacje, zwiększają przejrzystość i ułatwiają monitorowanie postępów w czasie.

Ramy KPI pełnią w tym kontekście funkcję narzędzia metodologicznego służącego operacjonalizacji celów strategicznych oraz ustrukturyzowanemu opisowi systemów miejskich. Opracowane przez National Institute of Standards and Technology ramy KPI traktują wskaźniki jako integralną część modelu informacyjnego, który systematycznie odwzorowuje relacje między źródłami danych, procesami transformacji i wynikającymi z nich efektami (Serrano i in., 2022, s. vii). Wskaźniki są więc rozumiane jako elementy uporządkowanego systemu informacyjnego.

Europejskie ramy CITYkeys rozszerzają to podejście, ujmując rozwój inteligentnych miast jako wielowymiarową koncepcję zrównoważonego rozwoju. Oprócz technologicznych wskaźników wyników uwzględniają także długoterminowe wskaźniki oddziaływania, które mają uchwycić zmiany w zakresie zrównoważonego rozwoju i jakości życia (Huovila i in., 2017, s. 113, 119). Ramy KPI wnoszą zatem istotny wkład w systematyczną operacjonalizację strategicznych celów rozwojowych oraz w strukturę oceny wyników miejskich.

W celu poprawy spójności i jasności terminologicznej opracowano międzynarodowe serie norm, takie jak ISO 37120, ISO 37122 i ISO 37123. Norma ISO 37120 definiuje wskaźniki służące pomiarowi usług miejskich i jakości życia w kontekście zrównoważonego rozwoju, norma ISO 37122 określa wskaźniki dla inteligentnych miast, a norma ISO 37123 dotyczy wskaźników służących ocenie odporności miast (ISO, 2018; ISO, 2019a; ISO, 2019b). Normy te zapewniają ujednoliconą terminologię oraz przejrzyste wytyczne pomiarowe i obliczeniowe, a także harmonizują definicje, jednostki miary i logikę obliczeń. Tworzą tym samym formalną podstawę znormalizowanych metod pomiaru. Główne ramy oraz sposób, w jaki odnoszą się one do kwestii porównywalności, przedstawiono w tabeli 1.

Tabela 1: Przegląd głównych ram i standardów dotyczących wskaźników KPI w zakresie inteligentnych miast

Struktura	Typ	Główny punkt zainteresowania	Postępowanie w zakresie porównywalności	Struktura metadanych	Logika podejmowania decyzji oparta na regułach
CITYkeys	Ramy badawcze	Ocena projektów „smart city" i ich wpływu na zrównoważony rozwój	Kwestia porównywalności została uwzględniona na poziomie koncepcyjnym	częściowo dostępne	nie
ISO 37120	międzynarodowy standard	Pomiar usług miejskich i jakości życia	w sposób dorozumiany poprzez ustandaryzowane definicje	częściowo dostępne	nie
Międzynarodowa norma ISO 37122	międzynarodowy standard	Ocena technologii smart city	w sposób dorozumiany poprzez definicje wskaźników	częściowo dostępne	nie
ISO 37123	międzynarodowy standard	Ocena odporności miast	w sposób dorozumiany poprzez standaryzację	częściowo dostępne	nie
Ramy wskaźników KPI NIST	Model informacyjny	Integracja źródeł danych, procesów i skutków	nie zostało wyraźnie poruszone	istnieje w teorii	nie
NGSI-LD / OGC SensorThings	techniczny standard interoperacyjności	Organizacja i wymiana danych miejskich	bez adresata	bardzo wyraźny	nie
Ramy jakości Eurostatu	norma statystyczna	Dokumentacja jakości danych	zwrócone pośrednio	dostępne	nie

Jednocześnie analizy porównawcze wskazują, że znormalizowane systemy wskaźników są opracowywane z myślą o zróżnicowanej grupie miast i z konieczności wykorzystują wskaźniki o charakterze ogólnym (Huovila i in., 2019, s. 151). Takie uogólnienie sprzyja wprawdzie szerokiemu zastosowaniu, ale nie ujawnia wprost różnic strukturalnych między miastami.

Oprócz standaryzacji treści coraz większego znaczenia nabiera techniczna interoperacyjność danych miejskich. Ramy interoperacyjności podkreślają rolę ustrukturyzowanych metadanych w integrowaniu heterogenicznych źródeł danych (Lóscio i in., 2017, b.s.). Standardy techniczne, takie jak NGSI-LD (ETSI, 2021, s. 24) czy OGC SensorThings API (Open Geospatial Consortium, 2021, s. 31, 38), udostępniają uporządkowane modele danych, natomiast ramy statystyczne definiują oznaczenia jakości, na przykład dla wartości zweryfikowanych lub oszacowanych (Eurostat, 2004, s. 7). Słowniki metadanych, takie jak Data Quality Vocabulary, umożliwiają ponadto dokumentowanie cech jakości w formacie nadającym się do odczytu maszynowego (Albertoni i Isaac, 2016; Van den Brink i in., 2019, s. 957).

Różnice w definicjach obszarów stosowanych w statystyce publicznej pokazują ponadto, że jednostki odniesienia mogą różnić się w zależności od tego, czy przyjmuje się granice administracyjne, czy funkcjonalne regiony miejskie (UN-Habitat, 2004, s. 6; Unia Europejska i in., 2021, s. 18; Eurostat, 2004, s. 10). Istniejące podręczniki metodologiczne dokumentują te różnice, ale nie dostarczają operacyjnej procedury klasyfikacji dla konkretnych sytuacji porównawczych.

Podsumowując, istnieje szerokie spektrum ram i standardów służących definiowaniu, standaryzacji oraz technicznej integracji wskaźników dotyczących inteligentnych miast. Ramy KPI przekładają cele strategiczne na poziom operacyjny, normy ISO ujednolicają definicje i metody obliczeniowe, standardy interoperacyjności porządkują przepływy danych oraz metadane, a podręczniki statystyczne dokumentują różnice metodologiczne.

Ani normy ISO (ISO, 2018; ISO, 2019a; ISO, 2019b), ani ramy KPI opracowane przez NIST (Serrano i in., 2022), ani istniejące standardy interoperacyjności i jakości nie zawierają jednak sformalizowanego, opartego na regułach modelu decyzyjnego, który systematycznie identyfikowałby odchylenia strukturalne między konkretnymi zestawami danych KPI i wyraźnie klasyfikował ich wpływ na porównywalność. Porównywalność jest więc zwykle zakładana, ale nie jest operacjonalizowana jako samodzielna, weryfikowalna cecha.

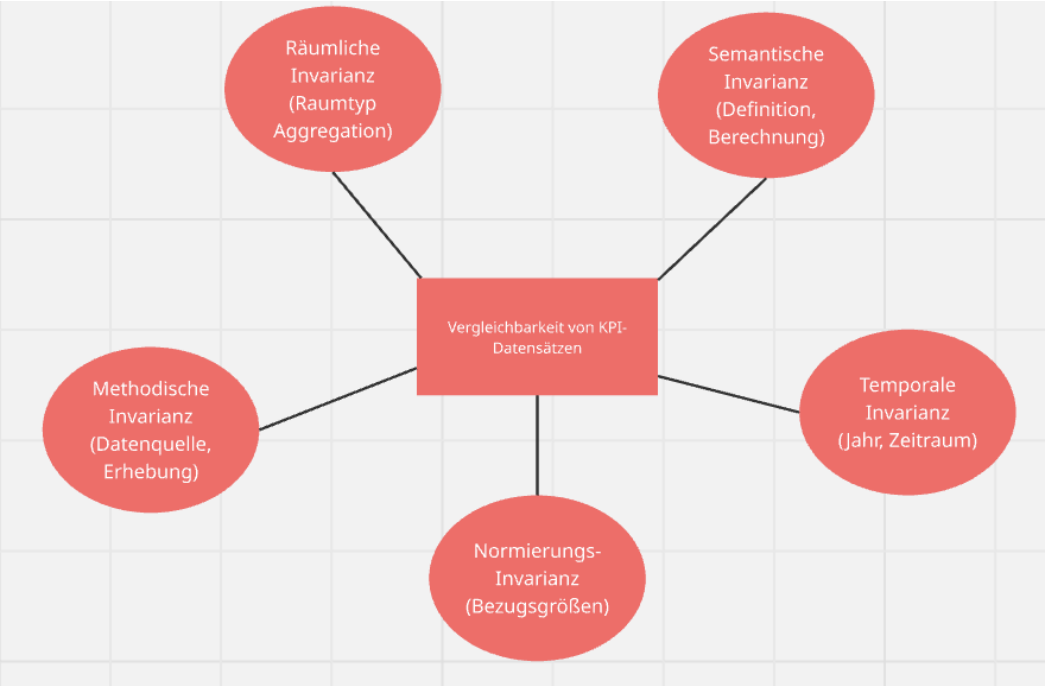
Niniejsza praca ma na celu wypełnienie tej luki badawczej. Nie chodzi w niej o opracowanie nowych wskaźników, lecz o stworzenie opartego na regułach i metadanych systemu porównywalności, który

systematycznie identyfikuje, ocenia i przejrzyście klasyfikuje odchylenia strukturalne między zbiorami danych KPI.

2.1 Porównywalność jako problem metodologiczny

Rzetelność metodologiczna porównań między miastami nie jest w zarządzaniu miejskim cechą oczywistą, lecz warunkiem, który należy wyraźnie sprawdzić i naukowo uzasadnić (Sterner i in., 2024, s. 747). W międzynarodowych inicjatywach monitorujących, na przykład w kontekście Celów Zrównoważonego Rozwoju, porównywalność bywa często przyjmowana domyślnie, bez systematycznej operacjonalizacji strukturalnej równoważności baz danych, na których opierają się analizy (OECD i Komisja Europejska, 2020, s. 10).

Aby porównanie grup było miarodajne, narzędzie pomiarowe musi we wszystkich jednostkach badawczych mierzyć ten sam konstrukt teoretyczny z porównywalną precyzją (Milfont i Fischer, 2010, s. 112). Jeśli warunek ten nie jest spełniony, zaobserwowane różnice tracą jednoznaczność wnioskowania (Sterner i in., 2024, s. 748). Główne aspekty tej problematyki niezmienności przedstawiono schematycznie na rysunku 2.



Rysunek 2: Wymiary porównywalności jako płaszczyzny niezmienności w analizach KPI między miastami

Źródło: Opracowanie własne

Z punktu widzenia teorii pomiaru porównywalność można interpretować jako problem niezmienności. Nie dotyczy ona wyłącznie formalnej zgodności nazw wskaźników, lecz przede wszystkim strukturalnej równoważności modeli pomiarowych, które leżą u podstaw porównywanych danych.

Jedną z głównych przeszkód w formułowaniu wiarygodnych wniosków jest przestrzenna niespójność danych miejskich. Krajowe definicje tego, co statystycznie uznaje się za miasto, często opierają się na granicach administracyjnych, które tylko częściowo odzwierciedlają funkcjonalne powiązania miejskie (Dijkstra i in., 2019, s. 4). Rozbieżność ta wiąże się z problemem modyfikowalnych jednostek obszarowych (Modifiable Areal Unit Problem), zgodnie z którym wybór granic agregacji może systematycznie wpływać na wartości wskaźników (European Union i in., 2021, s. 81). Jeśli miasta porównuje się na podstawie różnych koncepcji przestrzennych, takich jak City Proper i Functional Urban Area, uzyskane wskaźniki odnoszą się do jednostek, które nie są równoważne strukturalnie (Dijkstra i in., 2019, s. 13). Niespójność przestrzenna może być więc rozumiana jako szczególna forma strukturalnej nieinwariancji (Sterner i in., 2024, s. 751).

Problem ten pogłębia fakt, że główne cele zarządzania inteligentnym miastem, takie jak jakość życia czy zrównoważony rozwój, mają charakter konstruktów ukrytych, operacjonalizowanych za pomocą obserwowalnych wskaźników (Sterner i in., 2024, s. 748; Milfont i Fischer, 2010, s. 113). Aby porównania średnich wartości takich konstruktów były miarodajne, konieczna jest niezmienność pomiarowa (Milfont i Fischer, 2010, s. 112; Sterner i in., 2024, s. 747). Jeśli nie jest ona zapewniona, różnic w obserwowanych wartościach wskaźników nie można jednoznacznie interpretować jako różnic w badanym konstrukcie (Milfont i Fischer, 2010; Sterner i in., 2024).

Dodatkowe wyzwania wynikają z architektury istniejących systemów KPI. Zestawy wskaźników są silnie zróżnicowane pod względem celów i priorytetów, przy czym technologiczne wskaźniki wyników bywają eksponowane silniej niż wymiary oddziaływania związane ze zrównoważonym rozwojem (Huovila i in., 2019, s. 141). Ponadto różnice w jakości danych i metodach ich gromadzenia nie zawsze są odpowiednio udokumentowane (Huovila i in., 2019, s. 145). Jeżeli metody szacowania, odmienne definicje lub różne schematy zbierania danych nie są przedstawione w sposób przejrzysty, strukturalna równoważność danych pozostaje niesprawdzona (Eurostat, 2021, s. 125).

Problem porównywalności wynika zatem ze wzajemnej zależności trzech płaszczyzn: odniesienia przestrzennego, struktury pomiaru oraz jakości dokumentacji metodologicznej. Porównywalność jest metodologicznie uzasadniona tylko wtedy, gdy jednostki przestrzenne są funkcjonalnie równoważne (European Union i in., 2021, s. 82), modele pomiarowe wykazują wystarczającą niezmienność (Sterner i in., 2024, s. 748), a metadane umożliwiają rekonstrukcję procesu pomiarowego (Eurostat, 2021, s. 24).

Obecna praktyka polegająca na prowadzeniu benchmarków międzymiastowych wyłącznie na podstawie zbieżności nazw wskaźników często traktuje porównywalność jako założenie domyślne. Z metodologicznego punktu widzenia pozostaje ono jednak niewystarczająco uzasadnione.

Porównywalność staje się więc sama w sobie przedmiotem badania. Wynika z tego potrzeba stworzenia formalnych ram porównywalności opartych na metadanych, które systematycznie identyfikują odchylenia strukturalne i klasyfikują je zgodnie z ustalonymi regułami (Sterner i in., 2024, s. 755; Eurostat, 2021, s. 13). Dopiero taka operacjonalizacja pozwala metodologicznie zabezpieczyć trafność wniosków wyciąganych w analizach międzymiastowych.

3. Ramy porównywalności oparte na metadanych

Opracowane ramy porównywalności mają na celu odejście od domyślnego uznawania porównań między miastami za trafne. Ich zadaniem jest wyraźna, powtarzalna i oparta na metadanych weryfikacja metodologicznej zasadności takich porównań (Eurostat, 2021, s. 152; Sterner i in., 2024, s. 747). Porównywalność jest tu modelowana jako właściwość relacyjna konkretnych par wskaźników KPI. Nie wynika z identyczności nazw wskaźników, lecz ze strukturalnej równoważności udokumentowanych warunków pomiaru (Sterner i in., 2024, s. 751).

Ramy te nawiązują do istniejących podejść w zakresie standaryzacji i interoperacyjności, traktując ustrukturyzowane metadane jako warunek rzetelnego wykorzystania danych (Lóscio i in., 2017, rozdz. 8.2, b.s.; Van den Brink i in., 2019, s. 952; Eurostat, 2021, s. 9). Podczas gdy istniejące standardy harmonizują przede wszystkim definicje, proponowane podejście operacjonalizuje regułową logikę decyzyjną, która służy weryfikacji konkretnych sytuacji porównawczych.

W zarządzaniu miejskim rzetelność metodologiczna porównań między miastami powinna być traktowana jako warunek wymagający weryfikacji, jeśli porównania te mają służyć wyciąganiu wniosków (Sterner i in., 2024, s. 747). W międzynarodowych inicjatywach monitorujących, na przykład w kontekście Celów Zrównoważonego Rozwoju, wskaźniki są wpisywane w logikę porównawczą, natomiast strukturalna równoważność bazowych zbiorów danych nie zawsze jest wyraźnie operacjonalizowana (OECD i Komisja Europejska, 2020, s. 10; UN-Habitat, 2025).

W przypadku analiz międzymiastowych oznacza to, że porównywalności nie można sprowadzać do identycznych nazw wskaźników. Należy ją rozumieć jako równoważność warunków pomiaru i modeli

pomiarowych (Eurostat, 2021, s. 147; Milfont i Fischer, 2010, s. 116; Sterner i in., 2024, s. 748).

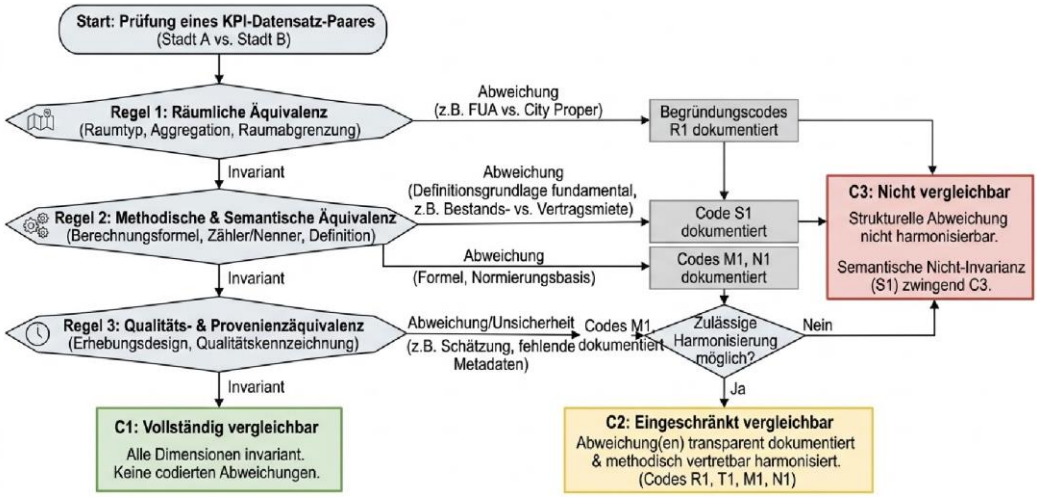
W celu operacjonalizacji proponuje się standardową kartę wskaźnika KPI (zob. tabela 2). Zawiera ona ustrukturyzowane pola metadanych, które powinny być obowiązkowo dokumentowane dla każdego wskaźnika.

Tabela 2 Szablon podstawowy karty charakterystyki KPI – pola metadanych, definicje i przykłady

Feld	Opis / Definicja	Wartość przykładowa
Nazwa wskaźnika KPI	Oficjalna nazwa wskaźnika	Emisja CO_2 na mieszkańca
Odniesienie do normy	Związek ze standardowymi ramami lub systemami wskaźników	ISO 37120
Wersja	Numer wersji wskaźnika lub zbioru danych	Wersja 2022
Typ pomieszczenia	Jednostka odniesienia przestrzennego wskaźnika	Samo miasto
Poziom agregacji	Poziom analizy danych	Miasto
Wyznaczenie przestrzeni	Opis wytyczenia granic geograficznych	Administracyjna granica miasta
Rok odniesienia	Rok zebrania danych	2022
Okres rozliczeniowy	Okres gromadzenia danych	Rok kalendarzowy
Częstotliwość	Częstotliwość gromadzenia danych	corocznie
Wzór obliczeniowy	Matematyczna definicja wskaźnika	Emisje CO_2 na mieszkańca
Licznik	Wartość licznika w obliczeniach	Całkowita emisja CO_2 związana z energią
Mianownik	Mianownik obliczeń	Liczba mieszkańców
Jednostka	Jednostka miary wskaźnika	t CO_2 na mieszkańca
Źródło danych	Instytucja lub zbiór danych	Urząd Statystyczny Berlina i Brandenburgii
Metoda badania	Sposób gromadzenia danych	Inwentaryzacja administracyjna
Metody szacowania	Oznaczenie możliwych metod szacowania	brak
Oznaczenie jakości	Informacje dotyczące jakości danych	zweryfikowane
Poziom dokumentacji	Kompletność metadanych	całkowicie

Struktura ta opiera się na wymaganiach dotyczących jakości i pochodzenia metadanych (Eurostat, 2021, s. 31; Albertoni i Isaac, 2016; Van den Brink i in., 2019, sekcja 3), rozszerzając je jednak o wyraźny wymiar porównywalności.

Porównywalność dwóch zestawów danych KPI jest sprawdzana za pomocą formalnej logiki decyzyjnej. Opiera się ona na sekwencyjnych regułach weryfikacji (zob. rysunek 3):



Rysunek 3: Logika decyzyjna ram porównywalności opartych na metadanych

Źródło: Opracowanie własne

Zasada 1: równoważność przestrzenna. Jeżeli typ jednostki przestrzennej lub logika agregacji różnią się

strukturalnie, para wskaźników jest klasyfikowana jako porównywalna z ograniczeniami albo jako nieporównywalna.

Zasada 2: równoważność metodologiczna. Odchylenia w formule obliczeniowej, liczniku, mianowniku lub podstawie definicyjnej prowadzą do obniżenia poziomu porównywalności.

Zasada 3: równoważność jakości i pochodzenia danych. Różnice w schematach gromadzenia danych lub brak oznaczeń jakości wpływają na klasyfikację zgodnie z udokumentowanymi kodami uzasadnienia.

Wynikiem analizy nie jest decyzja zero-jedynkowa, lecz trzystopniowa klasyfikacja: C1 – w pełni porównywalne, C2 – porównywalne z ograniczeniami, C3 – nieporównywalne.

Klasyfikacja ta jest uzupełniana udokumentowanymi kodami uzasadniającymi, co służy zapewnieniu przejrzystości i powtarzalności oceny.

Ramy dopuszczają wyraźnie określone etapy harmonizacji, o ile są one udokumentowane i możliwe do prześledzenia. Mogą one obejmować na przykład: przeliczenie do wspólnego typu jednostki przestrzennej, standaryzację wielkości odniesienia, ujawnienie zastosowanych metod szacowania.

Harmonizacja nie może jednak maskować braku niezmienności strukturalnej. Każde dostosowanie powinno być zatem opatrzone wyraźną informacją o zakresie i charakterze transformacji.

Ramy te nie służą testowaniu ukrytej niezmienności pomiarowej w ścisłym sensie statystycznym. Stanowią raczej operacyjną procedurę weryfikacji niezmienności procesu przetwarzania danych. Przekładają wymagania teorii pomiaru dotyczące porównywalności na reguły weryfikacyjne oparte na metadanych (Milfont i Fischer, 2010, s. 116; Sterner i in., 2024, s. 755; Eurostat, 2021, s. 147). W ten sposób porównywalność przestaje być domyślnym założeniem i staje się udokumentowaną, uregulowaną właściwością konkretnych zbiorów danych. W dalszej części artykułu sprawdzono, czy zdefiniowane wymiary porównawcze i reguły decyzyjne można zastosować w praktyce do rzeczywistych zbiorów danych.

Karta charakterystyki wskaźników KPI

Na podstawie koncepcyjnego uzasadnienia ram porównywalności opartych na metadanych zdefiniowano operacyjną strukturę karty wskaźnika KPI. W poprzednich częściach omówiono jej uwarunkowania normatywne i interoperacyjnościowe (m.in. DCAT, PROV-O, DQV), natomiast w tym miejscu przedstawiono metodologiczną konkretyzację w postaci ustrukturyzowanego profilu metadanych. Profil ten stanowi podstawę weryfikacji decyzji dotyczących porównywalności. Pola metadanych zdefiniowane w tabeli 2 są punktem wyjścia do oceny porównywalności.

Celem karty wskaźnika KPI jest systematyczne ujęcie cech niezbędnych do oceny niezmienności semantycznej, przestrzennej, czasowej i metodologicznej. Metadane nie są więc traktowane wyłącznie jako narzędzie dokumentacyjne, lecz jako struktura weryfikacyjna istotna dla podejmowania decyzji. Podejście to jest zgodne z podstawowymi zasadami Data on the Web Best Practices, zgodnie z którymi ustrukturyzowane metadane, informacje o pochodzeniu oraz dane dotyczące jakości są warunkiem interoperacyjności, ponownego wykorzystania i rzetelnego użycia danych (W3C, Data on the Web Best Practices, 2017). Również model DCAT podkreśla konieczność stosowania standardowych elementów metadanych, które umożliwiają interoperacyjny opis i wyszukiwanie rozproszonych zbiorów danych (Maali i Erickson, Data Catalog Vocabulary (DCAT), W3C Recommendation, 2014).

Struktura karty wskaźnika KPI

Karta wskaźnika składa się z obowiązkowych pól kontrolnych, które należy wypełnić dla każdej jednostki porównawczej. Brak poszczególnych danych formalnie ogranicza możliwość weryfikacji porównywalności; w takim przypadku przypisywany jest kod niepewności (U). Proponowana karta KPI obejmuje następujące obowiązkowe pola metadanych:

(1) Jednoznaczne określenie wskaźnika KPI. Spójna nazwa służy identyfikacji oraz ewentualnemu przyporządkowaniu do istniejących systemów wskaźników, na przykład serii ISO. Sama zbieżność nazw nie stanowi jednak kryterium porównywalności, lecz punkt wyjścia do dalszej analizy.

(2) Definicja formalna i logika obliczeniowa. Obejmuje ona matematyczną lub algorytmiczną specyfikację wskaźnika, na przykład iloraz, regułę agregacji lub wagę. Tylko wtedy, gdy podstawowa struktura obliczeniowa jest przejrzyście udokumentowana, można zweryfikować zgodność semantyczną. Międzynarodowe ramy normalizacyjne, takie jak norma ISO 37120, podkreślają, że wskaźniki muszą być precyzyjnie zdefiniowane i opisane w sposób metodologicznie zrozumiały, aby umożliwić porównywalność między miastami (ISO 37120:2018).
(3) Jednostka i poziom skali. Należy podać jednostkę miary, na przykład %, t CO_2 rocznie lub liczbę osób na 100 000 mieszkańców, a także poziom skali. Identyczność jednostek jest warunkiem koniecznym, ale niewystarczającym do zapewnienia porównywalności. Domyślne normalizacje lub ukryte przekształcenia powinny być dokumentowane oddzielnie.
(4) Zasięg przestrzenny. Należy precyzyjnie opisać jednostkę terytorialną, na której opiera się analiza, na przykład granice administracyjne miasta, funkcjonalny obszar miejski lub region metropolitalny. Niespójności przestrzenne uznaje się za jedną z głównych przyczyn zniekształconych porównań między miastami. Znaczenie jasno zdefiniowanych odniesień przestrzennych podkreśla zarówno europejska metodologia dotycząca miast, jak i definicja funkcjonalnych obszarów miejskich opracowana przez OECD i UE, rozróżniająca miasta administracyjne, ośrodki miejskie oraz funkcjonalne obszary miejskie (OECD i Komisja Europejska, 2020, s. 24).
(5) Ramy czasowe. Należy wskazać rok lub okres odniesienia oraz formę agregacji, na przykład średnią roczną albo dzień odniesienia. Niespójności czasowe mogą systematycznie wpływać na kolejność w rankingach, dlatego stanowią odrębne kryterium weryfikacji. Kwestie spójności czasowej i porównywalności są również omawiane w kontekście tworzenia międzynarodowych wskaźników (OECD i Komisja Europejska, 2008, s. 14).
(6) Źródło danych i metoda ich gromadzenia. Należy udokumentować pochodzenie danych, na przykład gminny urząd statystyczny, organ krajowy lub model obliczeniowy, oraz logikę ich pozyskiwania, taką jak dane administracyjne, próba badawcza lub symulacja. Informacje o pochodzeniu danych mają zasadnicze znaczenie dla wyjaśniania różnic metodologicznych; odpowiednie modele są formalizowane w standardzie ontologii pochodzenia PROV-O (Lebo i in., PROV-O: The PROV Ontology, W3C Recommendation, 2013).

Te pola obowiązkowe stanowią minimalną bazę informacyjną potrzebną do przeprowadzenia ustrukturyzowanej analizy porównywalności. Brak części danych ogranicza możliwość weryfikacji. Dodatkowo można uwzględnić następujące metadane, jeśli są dostępne: zasadę normalizacji, na przykład obliczenia na mieszkańca lub normalizację powierzchni, informacje dotyczące niepewności lub jakości, na przykład błąd standardowy albo metody estymacji, zgodnie z Data Quality Vocabulary (Albertoni i Isaac, 2016), odniesienie do ram normatywnych, na przykład wskaźnika ISO lub wytycznych krajowych, uwagi dotyczące ewentualnej harmonizacji.

Dane te poszerzają podstawę oceny, ale nie są bezwzględnie konieczne do przeprowadzenia formalnej klasyfikacji porównywalności.

Karta wskaźnika KPI stanowi uporządkowany poziom wejściowy dla logiki decyzyjnej. Dopiero wyraźne uwzględnienie wymienionych wymiarów pozwala nie tylko zakładać porównywalność, lecz także weryfikować ją na podstawie określonych reguł.

W ten sposób punkt ciężkości przesuwa się z prostego utożsamienia nazw wskaźników ku wielowymiarowej weryfikacji niezmienności opartej na udokumentowanych metadanych. Karta wskaźnika operacjonalizuje więc teoretycznie uzasadnione założenie, że porównywalność jest weryfikowalną właściwością struktur danych, a nie jedynie interpretacją.

Na podstawie ustrukturyzowanych metadanych wyróżnia się pięć analitycznych wymiarów porównywalności:

(1) Porównywalność semantyczna – sprawdzenie, czy definicja, wielkość docelowa i logika obliczeniowa są merytorycznie spójne. Nawet niewielkie odchylenia, takie jak różne wielkości odniesienia lub reguły agregacji, mogą prowadzić do pomiarów nieinwariantnych. Znaczenie precyzyjnych definicji

wskaźników podkreślają również międzynarodowe ramy normalizacyjne, takie jak norma ISO 37120 (ISO 37120:2018).

(2) Porównywalność przestrzenna – ocena, czy granice obszarów są identyczne lub funkcjonalnie równoważne. Różnice między granicami administracyjnymi a funkcjonalnymi obszarami miejskimi uznaje się za strukturalne źródło systematycznych zniekształceń (OECD i Komisja Europejska, 2020, s. 24).

(3) Porównywalność czasowa – analiza identyczności lat odniesienia lub metodologicznie uzasadnionych okresów porównawczych. Kwestie spójności czasowej i harmonizacji danych są omawiane w podręczniku OECD dotyczącym konstruowania wskaźników jako kluczowy wymiar jakości (OECD i Komisja Europejska, 2008, s. 14).

(4) Porównywalność metodologiczna – analiza metod gromadzenia danych i obliczeń. Różnice między pełnymi badaniami administracyjnymi a szacunkami opartymi na modelach mogą powodować efekty systematyczne i negatywnie wpływać na porównywalność.

(5) Porównywalność normalizacji i skalowania – sprawdzenie, czy zastosowano identyczne wielkości odniesienia, na przykład normalizację na mieszkańca, lub czy możliwa jest dopuszczalna transformacja.

Wymiary te stanowią uporządkowaną podstawę oceny i prowadzą do klasyfikacji opisanej poniżej.

Logika decyzyjna i system klasyfikacji

Na podstawie pięciu wymiarów porównawczych wprowadza się trzystopniową klasyfikację, która odzwierciedla stopień porównywalności między dwoma zestawami danych KPI. Celem tej klasyfikacji nie jest ogólne przyjmowanie porównywalności, lecz jej przejrzyste wyprowadzenie z udokumentowanych metadanych.

Parę wskaźników KPI klasyfikuje się jako C1 – porównywalną, jeżeli we wszystkich istotnych wymiarach zachowana jest niezmienność lub jeśli istniejące odchylenia można skompensować za pomocą transformacji w pełni zgodnych z regułami. W takim przypadku bezpośrednie porównanie jest metodologicznie uzasadnione.

Kategorię C2 – porównywalność z ograniczeniami – stosuje się wtedy, gdy w jednym lub kilku wymiarach występują odchylenia, które są jednak przejrzyście udokumentowane i mogą zostać uwzględnione dzięki dopuszczalnej harmonizacji albo ostrożnej interpretacji. Porównanie jest wówczas możliwe, ale nie powinno być prowadzone bez wskazania ograniczeń metodologicznych.

Jako C3 – nieporównywalne – klasyfikuje się pary wskaźników KPI, gdy naruszony jest co najmniej jeden kluczowy wymiar porównawczy i nie ma możliwości metodologicznie uzasadnionej harmonizacji. Dotyczy to zwłaszcza przypadków nieinwariancji semantycznej lub metodologicznej, w których analizowane zbiory danych, mimo podobnych nazw, nie odnoszą się do tej samej koncepcyjnej wielkości docelowej.

W celu zapewnienia przejrzystości stwierdzone rozbieżności są dodatkowo dokumentowane za pomocą standardowych kodów uzasadniających. Kod S1 odnosi się do odmiennej definicji lub wielkości docelowej wskaźnika, R1 – do różnych granic przestrzennych, T1 – do niespójności czasowych, M1 – do różnic metodologicznych, a N1 – do odmiennych podstaw normalizacji lub skalowania. Dzięki temu widoczny jest nie tylko wynik klasyfikacji, lecz także konkretny powód, dla którego porównanie uznano za dopuszczalne, ograniczone lub niedopuszczalne.

Znaczenie poszczególnych kodów zestawiono w tabeli 3.

Tabela 3:3 Kody uzasadnienia ram porównywalności

Kod	Wymiar	Znaczenie	Typowy wpływ na klasyfikację	Możliwa harmonizacja
S1	Porównywalność semantyczna	Różne definicje lub koncepcyjne wartości docelowe wskaźnika	prowadzi bezpośrednio do C3 (nie da się tego porównać)	nie
R1	Porównywalność przestrzenna	Różne definicje obszaru (np. centrum miasta a funkcjonalny obszar miejski)	zazwyczaj C2 (ograniczona porównywalność)	częściowo
T1	Porównywalność w czasie	Różne lata odniesienia lub logiki agregacji	zazwyczaj C2, o ile harmonizacja jest możliwa	częściowo
M1	Porównywalność metodologiczna	Różne metody gromadzenia danych lub obliczania	zazwyczaj C2, o ile nie występuje semantyczna nieinwariancja	częściowo
N1	Porównywalność normalizacji / skalowania	Różne podstawy normalizacji (np. definicja populacji)	zazwyczaj C2, o ile transformacja jest możliwa	częściowo
U1	Niepewność dotycząca metadanych	Brakująca lub niekompletna dokumentacja metadanych	zazwyczaj C2 ze względu na ograniczoną weryfikowalność	nie

Logika decyzyjna ram opiera się na zasadzie minimalnego założenia. Porównywalność nie jest zakładana z góry, lecz potwierdzana tylko wtedy, gdy między dwoma zestawami danych KPI nie stwierdza się istotnej niespójności wymiarowej albo gdy istniejące rozbieżności mogą zostać przejrzyście udokumentowane i zharmonizowane zgodnie z regułami. Pozwala to uniknąć sytuacji, w której formalne podobieństwa, takie jak identyczne nazwy wskaźników lub jednostki miary, są pochopnie interpretowane jako dowód faktycznej porównywalności.

Jednocześnie stwierdzone odchylenie nie oznacza automatycznie konieczności wykluczenia porównania. W określonych warunkach dopuszczalne są przekształcenia, o ile są jednoznaczne, spójne metodologicznie i w pełni udokumentowane. Dotyczy to zwłaszcza czysto matematycznych przeliczeń jednostek, na przykład z kilogramów na tony, wyraźnych normalizacji na mieszkańca opartych na identycznych definicjach populacji, a także agregacji lub dezagregacji, jeśli podstawowe jednostki przestrzenne są ze sobą kompatybilne. Takie dostosowania mogą poprawić porównywalność, ale nie mogą ukrywać różnic strukturalnych między zbiorami danych.

Niedopuszczalne są natomiast szacunki ex post bez udokumentowanej metodologii, domyślne założenia dotyczące brakujących danych oraz działania harmonizacyjne w przypadku zasadniczo odmiennych podstaw definicyjnych. W takich sytuacjach korekta obliczeniowa tworzyłaby jedynie pozór porównywalności, nie zapewniając faktycznego pomiaru tej samej wielkości docelowej. Ramy te opierają się zatem na metodologicznej zasadzie ostrożności: harmonizacja jest uzasadniona tylko tam, gdzie w sposób przejrzysty wzmacnia porównywalność, a nie tam, gdzie maskuje różnice koncepcyjne.

Podejście to jest zgodne z zaleceniami metodologicznymi dotyczącymi przejrzystego tworzenia wskaźników i harmonizacji danych, sformułowanymi w międzynarodowych wytycznych dotyczących opracowywania wskaźników złożonych (OECD i Komisja Europejska, 2008, s. 14). Opracowane ramy łączą trzy elementy: karta wskaźnika KPI zapewnia ustrukturyzowaną bazę metadanych, wymiary porównawcze tworzą analityczną podstawę weryfikacji, a logika decyzyjna przekłada udokumentowane odchylenia na zrozumiałą klasyfikację. Zasady harmonizacji gwarantują przy tym, że dostosowania nie są arbitralne, lecz wynikają z jasno określonych warunków metodologicznych.

W ten sposób ramy porównywalności zostały zoperacjonalizowane pod względem metodologicznym. W kolejnym rozdziale zastosowano je do wybranych zbiorów danych dotyczących miejskich wskaźników KPI, aby sprawdzić, jaki efekt przynosi ocena oparta na metadanych w porównaniu z czysto formalnym założeniem bazowym.

4. Studium przypadku: zastosowanie w Berlinie, Paryżu i Barcelonie

W celu oceny opracowanych ram porównywalności przeprowadzono studium przypadku, w którym zestawiono konwencjonalną, formalną logikę porównawczą z analizą opartą na metadanych. Celem jest pokazanie, czy i w jakim stopniu zmienia się ocena dopuszczalnych porównań między miastami, gdy porównywalność nie jest określana wyłącznie na podstawie nazw wskaźników, jednostek miary i roku odniesienia, lecz systematycznie weryfikowana na podstawie udokumentowanych metadanych (W3C, 2017, sekcje 8.4 i 8.5; ISO 37120:2018).

Punktem odniesienia jest najpierw model bazowy, który definiuje porównywalność wyłącznie poprzez formalną tożsamość wskaźników. W modelu tym para wskaźników KPI jest uznawana za porównywalną, jeśli nazwa wskaźnika, jednostka miary i rok odniesienia są identyczne. Nie uwzględnia się natomiast dalszych wymiarów, takich jak definicja, zasięg przestrzenny, metoda gromadzenia danych czy logika normalizacji. Model bazowy odzwierciedla zatem założenie często obecne w praktykach benchmarkingu: wskaźniki o formalnie identycznych nazwach traktuje się jako bezpośrednio porównywalne.

Badaniem objęto Berlin, Paryż i Barcelonę. Wybór tych miast wynika z ich dużego znaczenia w międzynarodowych porównaniach miejskich oraz z dostępności obszernych zbiorów otwartych danych i statystyk. Ponadto miasta te szczególnie dobrze nadają się do analizy podjętego problemu, ponieważ charakteryzują się różnymi definicjami obszarów odniesienia. Różnica między administracyjnymi granicami miasta (City Proper) a funkcjonalnymi obszarami miejskimi (Functional Urban Area) jest w międzynarodowych podejściach do urbanizacji traktowana jako metodologicznie istotna i nie jest uznawana za równoważną (OECD i Komisja Europejska, 2020, s. 24).

Na potrzeby studium przypadku wybrano trzy wskaźniki: emisje CO_2 związane z energią na mieszkańca, liczbę ofiar śmiertelnych wypadków drogowych na terenie miasta oraz czynsz netto za metr kwadratowy. Wybór ten wynika nie tylko z ich znaczenia politycznego i praktycznego, lecz przede wszystkim z odmiennej struktury niezmienności. Emisje CO_2 umożliwiają sprawdzenie różnych podejść do bilansowania i inwentaryzacji, a tym samym odnoszą się szczególnie do porównywalności semantyczno-metodycznej. Wskaźnik ofiar śmiertelnych wypadków drogowych służy jako przypadek testowy dla pozornie znormalizowanych wskaźników bezpieczeństwa, w których założenia definicyjne, takie jak zasada 30 dni, mogą być dokumentowane w różny sposób. Z kolei wskaźniki czynszowe stanowią szczególnie wyraźny przykład semantycznej nieinwariancji, ponieważ czynsz porównawczy, czynsz rynkowy i czynsz umowny odnoszą się do różnych wielkości docelowych. Przykłady te pokazują, że międzynarodowe ramy standaryzacji wymagają wprawdzie precyzyjnych definicji i przejrzystości metodologicznej, ale konkretna porównywalność poszczególnych zestawów danych wymaga dodatkowej weryfikacji (ISO 37120:2018; OECD i Komisja Europejska, 2008, s. 14).

Analiza opiera się wyłącznie na ogólnodostępnych źródłach pierwotnych. Dla każdego z trzech wskaźników sporządzono kartę KPI zgodnie z ramami opracowanymi w rozdziale 4. Następnie zbiory danych sprawdzono pod kątem pięciu wymiarów porównywalności: semantycznego, przestrzennego, czasowego, metodologicznego oraz związanego z normalizacją i skalowaniem. Luki w metadanych istotne dla kontroli niezmienności oznaczono kodem U1. Klasyfikacja opiera się na wcześniej zdefiniowanej logice decyzyjnej: C1 oznacza zbiory danych w pełni porównywalne i pozbawione zakodowanych odchyleń, C2 – zbiory danych porównywalne z ograniczeniami, a C3 – zbiory nieporównywalne, zwłaszcza w przypadku nieinwariancji semantycznej lub różnic metodologicznych, których nie można zharmonizować. Aby pokazać wartość dodaną ram, dla każdego KPI najpierw przedstawiono klasyfikację bazową, a następnie zestawiono ją z klasyfikacją opartą na metadanych.

W przypadku pierwszego wskaźnika, czyli emisji CO_2 na mieszkańca, Berlin podaje dane dotyczące emisji CO_2 związanych z energią na mieszkańca za rok 2022 (Urząd Statystyczny Berlina i Brandenburgii, 2022a, s. 24). Paryż dla tego samego roku odniesienia raportuje terytorialne emisje gazów cieplarnianych (Ville de Paris, 2025). Zgodnie z modelem bazowym wskaźnik ten wydaje się porównywalny, ponieważ jego nazwa, jednostka miary i rok odniesienia są formalnie zgodne. Analiza oparta na metadanych pokazuje jednak inny obraz. Różnice w metodach bilansowania i inwentaryzacji prowadzą do rozbieżności semantycznej (S1). Dochodzą do tego różnice w odniesieniu przestrzennym, ponieważ administracyjne granice miasta mogą wydawać się podobne, ale funkcjonalne obszary miejskie nie są strukturalnie równoważne (R1). Występują również różnice metodologiczne między systemami inwentaryzacji emisji (M1), a normalizacja na mieszkańca może opierać się na różnych definicjach populacji (N1). Ponieważ występuje semantyczna nieinwariancja, wskaźnik ten zostaje, zgodnie z logiką decyzyjną, sklasyfikowany jako nieporównywalny (C3). Późniejsza harmonizacja nie jest dopuszczalna, ponieważ odmienne logiki bilansowania zmieniają koncepcyjną wielkość docelową wskaźnika.

Drugi wskaźnik dotyczy ofiar śmiertelnych wypadków drogowych na terenie miasta. W Berlinie w 2022 r. odnotowano łącznie 34 ofiary śmiertelne wypadków drogowych (Urząd Statystyczny Berlin-Brandenburgia, 2022b, s. 5). Dla Paryża i Barcelony dostępne są odpowiednie zbiory danych typu open data (Paris Open Data, 2024; Ajuntament de Barcelona, b.d.). W modelu bazowym również ten wskaźnik uznaje się za porównywalny, ponieważ nazwa wskaźnika, jednostka miary i rok odniesienia są zgodne. Analiza ramowa pokazuje jednak, że podstawa definicyjna, zwłaszcza w odniesieniu do zasady 30 dni, nie jest jednoznacznie udokumentowana we wszystkich zbiorach danych. Niepewność tę oznaczono kodem U1. Pod względem przestrzennym dane odnoszą się wprawdzie do obszaru miejskiego, a rok kalendarzowy 2022 jest spójny, jednak występują różnice w procedurach gromadzenia danych i klasyfikacji, które zarejestrowano jako odchylenie metodologiczne (M1). Ponieważ nie ma potwierdzonej semantycznej nieinwariancji, ale porównanie ograniczają różnice metodologiczne i niekompletna dokumentacja metadanych, wskaźnik klasyfikuje się jako porównywalny z ograniczeniami (C2). Porównanie jest więc możliwe, ale wyłącznie przy wyraźnym wskazaniu ograniczeń metodologicznych.

Trzeci wskaźnik, czyli czynsz netto bez opłat eksploatacyjnych na metr kwadratowy, szczególnie wyraźnie pokazuje opisywaną problematykę. Berlin podaje lokalny czynsz porównawczy na podstawie tabeli czynszów (Kraj Związkowy Berlin, 2023, s. 5), Paryż korzysta z danych Observatoire des loyers de l'agglomération parisienne (OLAP, 2022), natomiast Barcelona publikuje średnie czynsze umowne (Ajuntament de Barcelona, b.d.). W modelu bazowym wartości te wydają się formalnie porównywalne, ponieważ są podawane jako czynsze za metr kwadratowy. Analiza oparta na metadanych wskazuje jednak na fundamentalną nieinwariancję semantyczną: czynsz porównawczy, czynsz w istniejących umowach i czynsz umowny odnoszą się do różnych pojęciowych wielkości docelowych. Ponadto występują odchylenia czasowe wynikające z różnych dat odniesienia i logiki agregacji (T1), a także różnice metodologiczne w gromadzeniu danych i obliczeniach (M1). Formalnie identyczna jednostka €/m² nie jest w stanie zniwelować tych różnic koncepcyjnych. Ze względu na semantyczną nieinwariancję wskaźnik czynszu zostaje sklasyfikowany jako nieporównywalny (C3). Bezpośrednie zestawienie stawek czynszu między miastami byłoby metodologicznie niedopuszczalne.

Studium przypadku pokazuje zatem, że analiza oparta na metadanych prowadzi do znaczącej rewizji oceny porównywalności. Zgodnie z modelem bazowym wszystkie trzy wskaźniki zostałyby uznane za porównywalne, natomiast analiza ramowa nie wskazuje ani jednej w pełni porównywalnej pary wskaźników KPI. Emisje CO_2 na mieszkańca oraz czynsz netto za metr kwadratowy zostały sklasyfikowane jako nieporównywalne, a wskaźnik ofiar śmiertelnych wypadków drogowych jako porównywalny jedynie w ograniczonym zakresie. Liczba porównań

dopuszczalnych bez zastrzeżeń zmniejsza się więc z trzech do zera. Jednocześnie opracowane ramy nie tylko dokumentują różnice formalne, lecz także zapobiegają tworzeniu metodologicznie problematycznych rankingów: ranking emisji i ranking czynszów byłyby na tej podstawie danych niedopuszczalne, a ranking bezpieczeństwa ruchu drogowego można by interpretować jedynie z zastrzeżeniami.

Macierz porównywalności i efekt empiryczny

Wyniki studium przypadku przedstawiono w tabeli 5. Zestawienie pokazuje, jak wyraźnie zmienia się ocena trzech badanych wskaźników KPI, gdy poza kryteriami formalnymi, takimi jak nazwa wskaźnika, jednostka miary i rok odniesienia, uwzględni się również metadane zdefiniowane w proponowanych ramach. Podczas gdy model bazowy klasyfikuje wszystkie trzy wskaźniki jako porównywalne, analiza oparta na metadanych prowadzi do znacznie bardziej restrykcyjnej oceny.

Tabela 5: Macierz wyników

KPI	Klasyfikacja bazowa	Klasyfikacja frameworków	Kody uzasadnienia	Krótkie uzasadnienie
Emisja CO_2 na mieszkańca	porównywalny	C3: nie da się porównać	Pierwszy rok, pierwszy semestr, pierwszy okres, pierwszy rok	Różne zasady sporządzania bilansów i systemy inwentaryzacyjne
ofiary śmiertelne wypadków drogowych	porównywalny	C2: porównywalne w ograniczonym zakresie	Najpierw, najpierw	Różnice metodologiczne i niekompletna dokumentacja definicji
Czynsz netto za metr kwadratowy bez opłat	porównywalny	C3: nie da się porównać	Pierwszy rok, pierwszy semestr, pierwszy miesiąc	Różne koncepcyjne definicje najmu

Macierz wyników pokazuje, że formalna zgodność wskaźnika nie wystarcza do przeprowadzenia rzetelnego porównania miast. Zgodnie z modelem bazowym wszystkie trzy wskaźniki KPI zostałyby dopuszczone do bezpośredniego porównania między miastami. Zastosowanie opracowanych ram analitycznych pokazuje jednak, że żaden z badanych wskaźników nie jest w pełni porównywalny. Wskaźnik dotyczący ofiar śmiertelnych wypadków drogowych został sklasyfikowany jedynie jako porównywalny z ograniczeniami, natomiast emisje CO_2 na mieszkańca oraz czynsz netto za metr kwadratowy uznano za nieporównywalne.

W ten sposób liczba porównań dopuszczalnych bez zastrzeżeń zmniejsza się z trzech do zera. Wartość dodana modelu jest szczególnie widoczna w tym, że wszystkie trzy wskaźniki zostają ponownie ocenione względem założenia bazowego. Emisje CO_2 zmieniają status z „porównywalnych" na „nieporównywalne", ponieważ stosowane są różne logiki bilansowania, systemy inwentaryzacji, odniesienia przestrzenne oraz podstawy normalizacji. Wskaźnik ofiar śmiertelnych wypadków drogowych pozostaje zasadniczo interpretowalny, jednak ze względu na różnice metodologiczne i niekompletną dokumentację definicji klasyfikowany jest tylko jako porównywalny z ograniczeniami. W przypadku wskaźnika czynszów odmienne podstawy koncepcyjne danych – czynsz porównawczy, czynsz rynkowy i czynsz umowny – prowadzą do jednoznacznej klasyfikacji jako nieporównywalny.

Ponowna ocena ma bezpośredni wpływ na ewentualne rankingi. Przy założeniu scenariusza bazowego można byłoby sporządzić rankingi dla wszystkich trzech wskaźników. Po zastosowaniu ram metodologicznych zarówno ranking emisji, jak i ranking czynszów byłyby jednak metodologicznie niedopuszczalne. Porównanie bezpieczeństwa ruchu drogowego można by interpretować jedynie w ograniczonym zakresie i musiałoby ono wyraźnie uwzględniać udokumentowane niepewności metodologiczne. Ramy zapobiegają zatem powstaniu dwóch

potencjalnie wprowadzających w błąd rankingów, a zarazem pokazują, że nawet pozornie proste wskaźniki wymagają systematycznej weryfikacji porównywalności.

Dyskusja, ograniczenia i implikacje

Studium przypadku pokazuje przede wszystkim, że dostępność i jakość metadanych są kluczowym warunkiem wiarygodnych porównań między miastami. Szczególnie dobrze widać to na przykładzie wskaźnika ofiar śmiertelnych wypadków drogowych. Choć na pierwszy rzut oka wydaje się on ustandaryzowany, jego podstawa definicyjna – na przykład w odniesieniu do zasady 30 dni – nie jest jednoznacznie udokumentowana we wszystkich zestawach danych. Ten brak przejrzystości prowadzi do nadania kodu U1, a tym samym do ograniczonej porównywalności. Bez wyraźnej weryfikacji metadanych wskaźnik ten zostałby w modelu bazowym błędnie potraktowany jako w pełni porównywalny. Porównywalność wskaźników miejskich zależy więc nie tylko od dostępności samych danych, lecz w znacznym stopniu także od jakości ich dokumentacji metodologicznej.

Analiza pokazuje również, że nie każde odchylenie można skorygować poprzez harmonizację. W przypadku wskaźnika czynszów występuje fundamentalna nieinwariancja semantyczna, ponieważ czynsz porównawczy, czynsz w istniejących umowach i czynsz umowny odnoszą się do różnych wielkości docelowych. Przekształcenie matematyczne nie jest w stanie usunąć tych różnic koncepcyjnych. W tym miejscu ramy metodologiczne zapobiegają niedopuszczalnej harmonizacji ex post i jasno wskazują, że porównywalność kończy się tam, gdzie podobnie nazwane wskaźniki nie mierzą tego samego.

Kolejne ograniczenie dotyczy pozostającej swobody interpretacji. Nawet sformalizowany zbiór zasad nie jest w stanie całkowicie zastąpić oceny eksperckiej. Widać to na przykład przy analizie różnic metodologicznych w inwentaryzacjach CO_2. Logika decyzyjna ogranicza subiektywność poprzez wyraźne kodowanie i dokumentowanie odchyleń, ale jej całkowicie nie eliminuje. Jej wartość dodana polega więc nie tyle na usunięciu wszelkiej niepewności, ile na zapewnieniu przejrzystości, zrozumiałości i weryfikowalności decyzji oceniających.

Studium przypadku zwraca wreszcie uwagę na kwestię skalowalności. Zastosowanie modelu do trzech wskaźników KPI jest metodologicznie przejrzyste i możliwe do zrealizowania. W przypadku bardziej rozbudowanych zestawów wskaźników nakład pracy związany z dokumentacją i weryfikacją znacznie by jednak wzrósł. Praktyczna użyteczność podejścia zależy zatem w dłuższej perspektywie od tego, czy metadane będą gromadzone w sposób bardziej znormalizowany i czy będzie można je przynajmniej częściowo analizować automatycznie. Bez ustrukturyzowanych metadanych zastosowanie modelu do dużych zbiorów danych byłoby operacyjnie wymagające. Dlatego badanie pokazuje, że porównywalność należy uwzględniać nie dopiero w momencie analizy, lecz już na etapie udostępniania danych miejskich.

6. Podsumowanie

Niniejszy artykuł wychodzi od fundamentalnego problemu trafności wyników w aktualnych badaniach nad inteligentnymi miastami: praktyki opierania porównań między miastami na naiwnym założeniu, że zbieżność nazw, identyczne jednostki oraz ten sam rok odniesienia są wystarczającym warunkiem porównywalności. Aby zminimalizować ryzyko błędnej interpretacji artefaktów metodologicznych jako rzeczywistych różnic w wynikach, sformułowano trzy pytania badawcze, których rozstrzygnięcie poszerza istniejące ramy KPI o niezbędną weryfikację trafności na poziomie danych.

W odpowiedzi na pierwsze pytanie badawcze, dotyczące formalizowania porównywalności, wykazano, że nie jest ona trywialną ani immanentną cechą samego wskaźnika. Została natomiast ujęta jako cecha relacyjna konkretnych par wskaźników KPI oraz jako wielowymiarowy warunek niezmienności. Trafne

wnioskowanie wymaga więc systematycznego zapewnienia równoważności strukturalnej na poziomie semantycznym, przestrzennym, czasowym, metodologicznym oraz w zakresie normalizacji.

W odniesieniu do drugiego pytania badawczego, dotyczącego opracowania systemu decyzyjnego opartego na metadanych, zaproponowano operacyjne podejście w postaci ustrukturyzowanej karty wskaźnika KPI. Dzięki sekwencyjnej, regułowej logice weryfikacji możliwe jest systematyczne identyfikowanie odchyleń strukturalnych między zestawami danych. Odchylenia te są nie tylko wykrywane, lecz także przejrzyście dokumentowane za pomocą standardowych kodów uzasadniających i klasyfikowane w trzypoziomowym systemie: w pełni porównywalne (C1), porównywalne z ograniczeniami (C2) oraz nieporównywalne (C3).

W ramach trzeciego pytania badawczego analizowano, w jakim stopniu wyraźna weryfikacja porównywalności zmienia wyniki analizy międzymiastowej w porównaniu z naiwnym założeniem bazowym. Studium przypadku oparte na danych z Berlina, Paryża i Barcelony przyniosło jednoznaczny wynik: podczas gdy założenie bazowe klasyfikowało wszystkie trzy badane wskaźniki, czyli emisje CO_2, ofiary śmiertelne wypadków drogowych i czynsz netto bez opłat, jako bezpośrednio porównywalne, opracowane ramy zmniejszyły liczbę porównań dopuszczalnych bez zastrzeżeń z trzech do zera. Jeden wskaźnik KPI został sklasyfikowany jako porównywalny z ograniczeniami (C2), a dwa – ze względu na zasadnicze niespójności semantyczne i metodologiczne – jako nieporównywalne (C3).

Podsumowując, niniejsza praca pokazuje, że rzetelność metodologiczna porównań między miastami musi być wyraźnie weryfikowana i dokumentowana. Opracowane ramy porównywalności wypełniają lukę w istniejących normach i systemach benchmarkingu, ponieważ znacząco zwiększają przejrzystość i powtarzalność analiz oraz zapobiegają tworzeniu rankingów niedopuszczalnych z metodologicznego punktu widzenia. Dalsze prace powinny koncentrować się na automatyzacji weryfikacji metadanych, aby zapewnić skalowalność ram w przypadku dużych zbiorów danych.

Bibliografia

Albertoni, R., & Isaac, A. (2016). *Data quality vocabulary (DQV)* (W3C Working Group Note). World Wide Web Consortium. https://www.w3.org/TR/vocab-dqv/

Ajuntament de Barcelona. (o. J.). *Nombre de víctimes de sinistres de trànsit*. Portal de Dades. https://portaldades.ajuntament.barcelona.cat/

Ajuntament de Barcelona. (o. J.). *Preu mitjà per superfície del lloguer*. Portal de Dades. https://portaldades.ajuntament.barcelona.cat/

Amt für Statistik Berlin-Brandenburg. (2022a). *Energie- und CO_2-Bilanz in Berlin 2022 (E IV 4 - j/22)*. https://download.statistik-berlin-brandenburg.de/

Amt für Statistik Berlin-Brandenburg. (2022b). *Straßenverkehrsunfälle in Berlin 2022 (H I 2 - j/22)*. https://download.statistik-berlin-brandenburg.de/

Dijkstra, L., & Poelman, H. (2014). *A harmonised definition of cities and rural areas: The new degree of urbanisation* (Regional Working Paper WP 01/2014). European Commission, Directorate-General for Regional and Urban Policy.

Dijkstra, L., Poelman, H., & Veneri, P. (2019). *The EU OECD definition of a functional urban area* (OECD Regional Development Working Papers No. 2019/11). OECD Publishing. https://doi.org/10.1787/d58cb34d-en

ETSI. (2021). *Context information management (CIM), NGSI-LD API (ETSI GS CIM 009 V1.5.1)*. European Telecommunications Standards Institute.

European Union, FAO, UN-Habitat, OECD, & The World Bank. (2021). *Applying the degree of urbanisation: A methodological manual to define cities, towns and rural areas for international comparisons* (2021 ed.). Publications Office of the European Union. https://doi.org/10.2785/706535

Eurostat. (2004). *Urban audit methodological handbook* (2004 ed.). Office for Official Publications of the European Communities.

Eurostat. (2017). *Methodological manual on city statistics* (2017 ed.). Publications Office of the European Union. https://doi.org/10.2785/708009

Eurostat. (2021). *European statistical system handbook for quality and metadata reports* (2021 ed.). Publications Office of the European Union.

Huovila, A., Airaksinen, M., Pinto-Seppä, I., Piira, K., Bosch, P., Penttinen, T., Neumann, H.-M., & Kontinakis, N. (2017). CITYkeys smart city performance measurement system. *International Journal for Housing Science and Its Applications, 41*(2), 113–125.

Huovila, A., Bosch, P., & Airaksinen, M. (2019). Comparative analysis of standardized indicators for smart sustainable cities. *Cities, 89*, 141–153. https://doi.org/10.1016/j.cities.2019.01.041

ISO. (2018). *ISO 37120:2018. Sustainable cities and communities — Indicators for city services and quality of life*. International Organization for Standardization.

ISO. (2019a). *ISO 37122:2019. Sustainable cities and communities — Indicators for smart cities*. International Organization for Standardization.

ISO. (2019b). *ISO 37123:2019. Sustainable cities and communities — Indicators for resilient cities*. International Organization for Standardization.

Land Berlin. (2023). *Berliner Mietspiegel 2023 (Stichtag 01.09.2022)*. Senatsverwaltung für Stadtentwicklung, Bauen und Wohnen. https://www.berlin.de/

Lebo, T., Sahoo, S., & McGuinness, D. (Hrsg.). (2013). *PROV-O: The PROV ontology* (W3C Recommendation). World Wide Web Consortium. https://www.w3.org/TR/prov-o/

Lóscio, B. F., Burle, C., & Calegari, N. (2017). *Data on the web best practices* (W3C Recommendation). World Wide Web Consortium. https://www.w3.org/TR/dwbp/

Maali, F., & Erickson, J. (2014). *Data catalog vocabulary (DCAT)* (W3C Recommendation). World Wide Web Consortium. https://www.w3.org/TR/vocab-dcat/

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research, 3*(1), 111–121.

Observatoire des loyers de l'agglomération parisienne (OLAP). (2022). *Rapport Paris 2022*. https://www.observatoire-des-loyers.fr/

OECD i Komisja Europejska. (2008). Handbook on constructing composite indicators: Methodology and user guide. OECD Publishing. https://doi.org/10.1787/9789264043459-en https://doi.org/10.1787/9789264043459-en

OECD i Komisja Europejska. (2020). Cities in the world: A new perspective on urbanisation

Open Geospatial Consortium. (2021). *OGC SensorThings API Part 1: Sensing Version 1.1* (Implementation Standard OGC 18-088). https://docs.ogc.org/is/18-088/18-088.html

Paris Open Data. (2024). *Accidentologie: base victimes*. Ville de Paris. https://opendata.paris.fr/

Paris Open Data. (2026). *Inventaire des émissions de gaz à effet de serre du territoire parisien*. Ville de Paris. https://opendata.paris.fr/

Serrano, M., Griffor, E., Wollman, D., Dunaway, M., Burns, M., Rhee, S., & Greer, C. (2022). *Smart cities and communities: A key performance indicators framework* (NIST Special Publication 1900-206). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1900-206

Sterner, P., Pargent, F., Deffner, D., & Goretzko, D. (2024). *A causal framework for the comparability of latent variables*. Manuscript submitted for publication.

UN-Habitat. (2004). *Urban indicators guidelines: Monitoring the Habitat Agenda and the Millennium Development Goals*. United Nations Human Settlements Programme.

UN-Habitat. (2025). *SDG indicator metadata: Indicator 11.3.1: Ratio of land consumption rate to population growth rate*. United Nations Human Settlements Programme.

Van den Brink, L., Barnaghi, P., Tandy, J., Atemezing, G., Atkinson, R., Cochrane, B., Fathy, Y., García-Castro, R., Haller, A., Harth, A., Janowicz, K., Kolozali, Ş., van Leeuwen, B., Lefrançois, M., Lieberman, J., Perego, A., Le-Phuoc, D., Roberts, B., Taylor, K., & Troncy, R.

(2019). Best practices for publishing, retrieving, and using spatial data on the web. *Semantic Web, 10*(6), 951–980. https://doi.org/90336 10.3233/SW-1

Ville de Paris. (2025). *Les émissions de gaz à effet de serre de Paris*. https://www.paris.fr/

World Wide Web Consortium. (2017). *Data on the web best practices*. https://www.w3.org/TR/dwbp/

Vergleichbarkeit von Smart City Kennzahlen, ein metadatenbasiertes Konzept zur Prüfung und Harmonisierung von Cross-City Vergleichen

Kevin Weigel, Phillip Preischhoff
DHBW Heidenheim,

Abstract

Key Performance Indicators, KPIs, haben sich in der Smart City Forschung als zentrales Instrument zur Operationalisierung strategischer Zielsetzungen und zur Durchführung von Cross-City Vergleichen etabliert. Trotz internationaler Standardisierungsinitiativen wie ISO 37120, 37122 und 37123 bleibt Vergleichbarkeit in der Praxis häufig eine implizite Annahme, ohne dass die strukturelle Äquivalenz konkreter Datensätze systematisch geprüft wird. Der vorliegende Beitrag adressiert diese Forschungslücke durch die Entwicklung eines metadatenbasierten Vergleichbarkeitsrahmens, der Vergleichbarkeit als mehrdimensionale Invarianzbedingung konzeptualisiert und regelgebunden operationalisiert.

Kern des Ansatzes ist ein strukturierter KPI-Steckbrief, der semantische, räumliche, temporale, methodische sowie normierungsbezogene Dimensionen explizit erfasst. Auf dieser Grundlage wird eine dreistufige Klassifikation eingeführt, C1 vergleichbar, C2 eingeschränkt vergleichbar, C3 nicht vergleichbar, die auf dokumentierten Metadaten und definierten Entscheidungsregeln basiert. Vergleichbarkeit wird damit nicht aus Indikatornamen oder Maßeinheiten abgeleitet, sondern als prüfbare Eigenschaft konkreter KPI-Paare modelliert.

Die Anwendung des Rahmens wird anhand einer Fallstudie demonstriert mit drei öffentlich zugänglichen KPIs, CO_2 Emissionen pro Einwohner, Verkehrstote, Nettokaltmiete pro Quadratmeter, für die Städte Berlin, Paris und Barcelona. Während eine naive indikatorenbasierte Betrachtung alle drei Kennzahlen als unmittelbar vergleichbar klassifiziert, zeigt die Fallstudie, dass keiner der untersuchten Indikatoren vollständig invariant ist. Ein KPI wird als eingeschränkt vergleichbar eingestuft, zwei KPIs als nicht vergleichbar. Die Anzahl uneingeschränkt zulässiger Vergleiche reduziert sich damit von drei auf null.

Der Beitrag leistet eine theoretische und methodische Erweiterung bestehender KPI-Frameworks, indem Vergleichbarkeit als explizite Validitätsbedingung auf Datenebene modelliert wird. Der entwickelte Rahmen erhöht Transparenz, Reproduzierbarkeit und methodische Robustheit urbaner Cross-City Analysen und kann als vorgelagerte Qualitätssicherung in Smart City Rankings und Benchmarking Systemen eingesetzt werden.

1. Einleitung

In der Smart City Forschung haben sich Key Performance Indicators, KPIs, als zentrales Instrument etabliert, um strategische Zielsetzungen von Städten zu operationalisieren und Fortschritte messbar zu machen. KPIs dienen der internen Steuerung, der externen Rechenschaft sowie der Kommunikation urbaner Entwicklungsstände. Insbesondere in komplexen Transformationsprozessen werden Kennzahlensysteme genutzt, um technologische Innovation, nachhaltige Entwicklung und gesellschaftliche Wirkungen strukturiert abzubilden (Huovila et al., 2017, S. 114-115). Frameworks wie CITYkeys verfolgen dabei das Ziel, eine transparente Methodik bereitzustellen und Vergleichbarkeit zwischen Städten zu fördern, indem sie Indikatoren systematisieren und den Austausch zwischen Akteurinnen und Akteuren erleichtern (Huovila et al., 2017, S. 115).

Mit der zunehmenden Nutzung von KPIs für Cross-City Vergleiche tritt jedoch ein grundlegendes Validitätsproblem in den Vordergrund. Unterschiede zwischen Städten werden häufig unmittelbar als Leistungsunterschiede interpretiert, obwohl nicht systematisch geprüft wird, ob die zugrunde liegenden Kennzahlen auf strukturell äquivalenten Datensätzen beruhen. Zahlreiche KPI-Systeme sind als Kompromiss für heterogene Städte konzipiert. Dadurch entstehen Interpretationsspielräume, die zwar breite Anwendbarkeit ermöglichen, jedoch die strukturellen Differenzen zwischen lokalen Definitionen, Raumbezügen und Erhebungspraktiken nicht explizit abbilden (Huovila et al., 2019, S. 151). Zusätzlich fokussieren viele KPI-Ansätze stärker auf sektorspezifische oder technologische Output Größen, während Wirkungsdimensionen wie Lebensqualität, Governance oder soziale Kohäsion vergleichsweise

unterrepräsentiert bleiben (Huovila et al., 2019, S. 150; Serrano et al., 2022, S. 1). Für Cross-City Analysen bedeutet dies: Namensgleichheit oder identische Einheiten sind keine hinreichende Bedingung dafür, dass die verglichenen Werte dasselbe Konstrukt unter vergleichbaren Bedingungen messen (Sterner et al., 2024, S. 748; Eurostat, 2021, S. 147)). Dies wird schematisch in Abbildung 1 dargestellt.

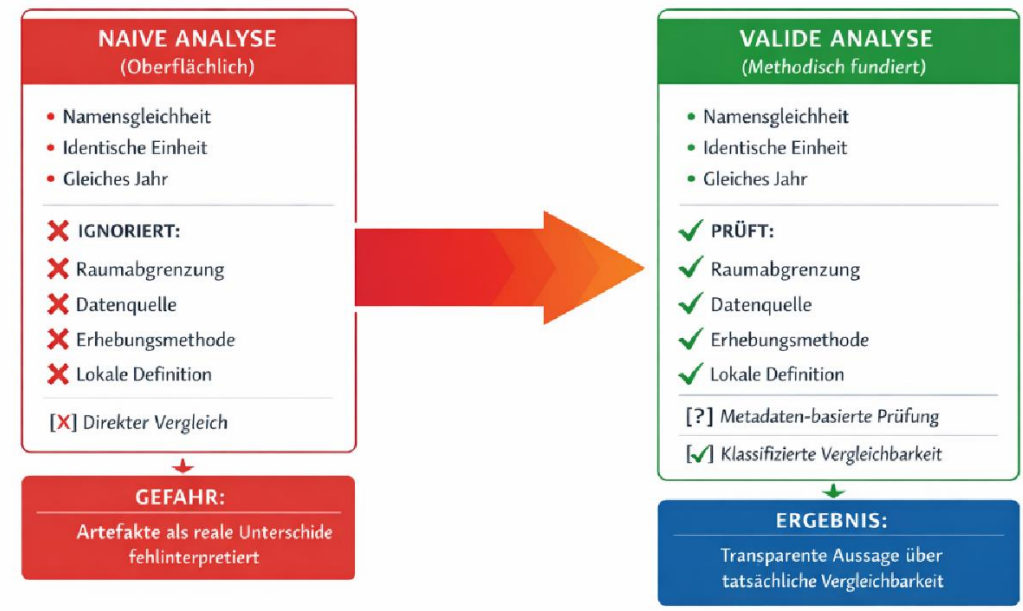


Abbildung 2: naive vs. valide Analyse
Quelle: Eigene Darstellung

Diese Problematik ist nicht primär eine Frage inhaltlicher Schwerpunktsetzung, sondern eine Frage der Inferenz (Sterner et al., 2024, S. 748). Vergleichbarkeit beschreibt das Ausmaß, in dem Unterschiede zwischen Werten auf reale Divergenzen im zugrunde liegenden Konstrukt zurückgeführt werden können und nicht auf Artefakte abweichender Definitionen, Raumabgrenzungen oder Methoden (Eurostat, 2021, S. 147). Wenn diese Voraussetzung verletzt ist, verlieren Rangreihen, Benchmarks und Korrelationen ihre interpretative Eindeutigkeit. Vergleichbarkeit ist damit keine triviale Eigenschaft von Kennzahlen, sondern eine methodische Voraussetzung belastbarer Cross-City Analysen. (Milfont & Fischer, 2010, S. 112; Sterner et al., 2024, S. 748).

Internationale Standardisierungsinitiativen markieren den Stand der Praxis, insbesondere die Normen ISO 37120:2018, ISO 37122:2019 und ISO 37123:2019. ISO 37120 definiert Indikatoren zur Messung städtischer Dienstleistungen und der Lebensqualität im Kontext nachhaltiger Entwicklung (ISO, 2018), ISO 37122 spezifiziert Kennzahlen zur Bewertung von Smart Cities (ISO, 2019a), und ISO 37123 beschreibt Indikatoren zur Erfassung urbaner Resilienz (ISO, 2019b). Diese Normen harmonisieren Terminologien sowie Mess- und Berechnungsvorgaben und zielen darauf, die Vergleichbarkeit von Ergebnissen zu verbessern (Huovila et al., 2019, S. 151; ISO, 2018). Gleichzeitig adressieren sie Vergleichbarkeit primär auf der Ebene von Indikatordefinitionen und Vorgaben, nicht als operatives Entscheidungsproblem konkreter Datensätze. Bei der Nutzung heterogener Open Data Quellen bleibt deshalb offen, wie Abweichungen in räumlicher Abgrenzung, Datenquelle oder Erhebungsmethodik systematisch identifiziert und in ihrer Auswirkung auf die Vergleichbarkeit klassifiziert werden können.

Vor diesem Hintergrund adressiert die vorliegende Arbeit drei Forschungsfragen. Erstens: Wie lässt sich Vergleichbarkeit zwischen städtischen KPIs als formalisierbare Eigenschaft konkreter Datensätze modellieren. Zweitens: Wie kann ein metadatenbasierter, regelgebundener Entscheidungsrahmen strukturelle Abweichungen zwischen KPI-Datensätzen identifizieren und transparent klassifizieren. Drittens: In welchem Ausmaß verändert eine explizite Vergleichbarkeitsprüfung die Ergebnisse einer Cross-City Analyse im Vergleich zu einem indikatorbasierten Baseline-Annahme, die Vergleichbarkeit aus Namensgleichheit, Einheit und Referenzjahr ableitet.

Der Beitrag dieser Arbeit ist dreifach. Erstens wird Vergleichbarkeit als relationale Eigenschaft konkreter KPI-Paare konzeptualisiert, nicht als Eigenschaft eines Indikators an sich. Zweitens wird dieses Konzept in Form eines strukturierten KPI-Steckbriefs und einer regelbasierten Entscheidungslogik operationalisiert, die Abweichungen über Begründungscodes dokumentiert und damit Transparenz sowie Reproduzierbarkeit erhöht. Drittens wird die Wirkung des Rahmens anhand einer Fallstudie demonstriert, indem die Baseline-Annahme einer unmittelbaren Vergleichbarkeit den Ergebnissen einer metadatenbasierten Prüfung gegenübergestellt wird.

Theoretisch knüpft der Ansatz an das Prinzip der Messinvarianz an, das Vergleichbarkeit als Voraussetzung valider Gruppenvergleiche versteht (Milfont & Fischer, 2010, S. 111; Sterner et al., 2024, S. 747). Operativ wird jedoch keine latente Modellinvarianz im statistischen Sinn getestet. Gegenstand der Analyse ist vielmehr die Invarianz von Datenprozessen, Referenzdefinitionen und methodischen Rahmenbedingungen auf Ebene konkreter Datensätze (Eurostat, 2021, S. 13; Sterner et al., 2024, S. 755). Kapitel 4 entwickelt den metadatenbasierten Vergleichbarkeitsrahmen, Kapitel 5 evaluiert dessen Wirkung anhand einer Fallstudie, Kapitel 6 diskutiert Limitationen und Implikationen, Kapitel 7 schließt mit Fazit und Ausblick.

2. Stand der Forschung und Problemstellung der Vergleichbarkeit

In der aktuellen Smart-City-Forschung haben sich Kennzahlensysteme, sogenannte Key Performance Indicators, als zentrales Instrument etabliert, um komplexe Transformationsprozesse urbaner Räume zu quantifizieren und Entscheidungsprozesse im urbanen Management zu unterstützen. KPIs ermöglichen die strukturierte Erfassung von Entwicklungen in zentralen Handlungsfeldern wie Wirtschaft, Umwelt und Lebensqualität und überführen strategische Zielsetzungen in operationalisierbare, messbare Größen (Huovila et al., 2019, S. 142). KPI-basierte Frameworks systematisieren relevante Informationen, erhöhen Transparenz und unterstützen die Nachvollziehbarkeit von Fortschritten über die Zeit.

Ein KPI-Framework fungiert dabei als methodisches Instrument zur Operationalisierung strategischer Ziele und zur strukturierten Beschreibung urbaner Systeme. Der vom National Institute of Standards and Technology entwickelte KPI-Rahmen versteht Kennzahlen als integralen Bestandteil eines Informationsmodells, das Beziehungen zwischen Datenquellen, Transformationsprozessen und resultierenden Wirkungen systematisch abbildet (Serrano et al., 2022, S. vii). Kennzahlen werden somit als Elemente eines strukturierten Informationssystems konzipiert.

Das europäische CITYkeys-Framework erweitert diesen Ansatz, indem es Smart-City-Entwicklung als mehrdimensionales Nachhaltigkeitskonzept definiert. Neben technologischen Output-Indikatoren werden auch langfristige Impact-Indikatoren berücksichtigt, die Veränderungen in nachhaltiger Entwicklung und Lebensqualität erfassen sollen (Huovila et al., 2017, S. 113, 119). KPI-Frameworks leisten damit einen zentralen Beitrag zur systematischen Operationalisierung strategischer Entwicklungsziele und zur Strukturierung urbaner Leistungsbewertung.

Zur Verbesserung von Konsistenz und terminologischer Klarheit wurden internationale Normenreihen wie ISO 37120, ISO 37122 und ISO 37123 entwickelt. ISO 37120 definiert Indikatoren zur Messung städtischer Dienstleistungen und der Lebensqualität im Kontext nachhaltiger Entwicklung, ISO 37122 spezifiziert Indikatoren für Smart Cities, und ISO 37123 adressiert Indikatoren zur Bewertung urbaner Resilienz (ISO, 2018; ISO, 2019a; ISO, 2019b). Die Normen stellen einheitliche Terminologien sowie transparente Mess- und Berechnungsvorgaben bereit und harmonisieren Definitionen, Einheiten und Berechnungslogiken. Sie schaffen damit eine formale Grundlage für standardisierte Messansätze. Die zentralen Frameworks und deren Umgang mit Vergleichbarkeit sind in Tabelle 1 zusammengefasst.

Tabelle 4: Überblick über zentrale Smart-City-KPI-Frameworks und Standards

Framework	Typ	Hauptfokus	Umgang mit Vergleichbarkeit	Metadatenstruktur	Regelbasierte Entscheidungs-logik
CITYkeys	Forschungsframework	Bewertung von Smart-City-Projekten und Nachhaltigkeitswirkungen	Vergleichbarkeit konzeptionell adressiert	teilweise vorhanden	nein
ISO 37120	internationaler Standard	Messung urbaner Dienstleistungen und	implizit über standardisierte	teilweise vorhanden	nein

Framework	Typ	Hauptfokus	Umgang mit Vergleichbarkeit	Metadatenstruktur	Regelbasierte Entscheidungs-logik
		Lebensqualität	Definitionen		
ISO 37122	internationaler Standard	Bewertung von Smart-City-Technologien	implizit über Indikatordefinitionen	teilweise vorhanden	nein
ISO 37123	internationaler Standard	Bewertung urbaner Resilienz	implizit über Standardisierung	teilweise vorhanden	nein
NIST KPI Framework	Informationsmodell	Integration von Datenquellen, Prozessen und Wirkungen	nicht explizit adressiert	konzeptionell vorhanden	nein
NGSI-LD / OGC SensorThings	technischer Interoperabilitätsstandard	Strukturierung und Austausch urbaner Daten	nicht adressiert	stark ausgeprägt	nein
Eurostat Qualitätsrahmen	statistischer Standard	Dokumentation von Datenqualität	indirekt adressiert	vorhanden	nein

Gleichzeitig weisen vergleichende Analysen darauf hin, dass standardisierte Indikatorensysteme für eine heterogene Gruppe von Städten konzipiert sind und notwendigerweise generische Indikatoren verwenden (Huovila et al., 2019, S. 151). Diese Generalisierung unterstützt zwar die breite Anwendbarkeit, modelliert jedoch strukturelle Unterschiede zwischen Städten nicht explizit.

Neben der inhaltlichen Standardisierung gewinnt die technische Interoperabilität urbaner Daten an Bedeutung. Interoperabilitätsrahmen betonen die Rolle strukturierter Metadaten für die Integration heterogener Datenquellen (Lóscio et al., 2017, O. S.). Technische Standards wie NGSI-LD (ETSI, 2021, S. 24) oder die OGC SensorThings API (Open Geospatial Consortium, 2021, S. 31, 38) stellen strukturierte Datenmodelle bereit, während statistische Rahmenwerke Qualitätskennzeichnungen definieren, etwa zur Kennzeichnung validierter oder geschätzter Werte (Eurostat, 2004, S. 7). Metadatenvokabulare wie das Data Quality Vocabulary ermöglichen zudem eine maschinenlesbare Dokumentation von Qualitätsmerkmalen (Albertoni & Isaac, 2016; Van den Brink et al., 2019, S. 957).

Unterschiedliche Raumdefinitionen in der amtlichen Statistik verdeutlichen darüber hinaus, dass Referenzeinheiten zwischen administrativen Grenzen und funktionalen Stadtregionen variieren können (UN-Habitat, 2004, S. 6; European Union et al., 2021, S. 18; Eurostat, 2004, S. 10). Bestehende methodische Handbücher dokumentieren diese Unterschiede, ohne jedoch ein operatives Klassifikationsverfahren für konkrete Vergleichssituationen bereitzustellen.

Zusammenfassend existiert somit ein umfangreiches Spektrum an Frameworks und Standards zur Definition, Standardisierung und technischen Integration von Smart-City-Kennzahlen. KPI-Frameworks operationalisieren strategische Ziele, ISO-Normen harmonisieren Definitionen und Berechnungsmethoden, Interoperabilitätsstandards strukturieren Datenflüsse sowie Metadaten, und statistische Handbücher dokumentieren methodische Unterschiede.

Weder die ISO-Normen (ISO, 2018; ISO, 2019a; ISO, 2019b) noch der NIST-KPI-Rahmen (Serrano et al., 2022) oder bestehende Interoperabilitäts- und Qualitätsstandards enthalten jedoch ein formalisiertes, regelbasiertes Entscheidungsmodell, das strukturelle Abweichungen zwischen konkreten KPI-Datensätzen systematisch identifiziert und deren Auswirkungen auf die Vergleichbarkeit explizit klassifiziert. Vergleichbarkeit wird damit implizit vorausgesetzt, jedoch nicht als eigenständige, prüfbare Eigenschaft operationalisiert.

An dieser Forschungslücke setzt die vorliegende Arbeit an. Ziel ist nicht die Entwicklung neuer Indikatoren, sondern die Konzeption eines regelbasierten, metadatenintegrierenden Vergleichbarkeitsrahmens, der strukturelle Abweichungen zwischen KPI-Datensätzen systematisch identifiziert, bewertet und transparent klassifiziert.

2.1 Vergleichbarkeit als methodisches Problem

Die methodische Belastbarkeit von Cross-City-Vergleichen stellt in der urbanen Governance keine

gegebene Eigenschaft dar, sondern eine zentrale Voraussetzung, die explizit geprüft und wissenschaftlich begründet werden muss (Sterner et al., 2024, S. 747). In internationalen Monitoring-Initiativen, etwa im Kontext der Sustainable Development Goals, wird Vergleichbarkeit häufig implizit vorausgesetzt, ohne die strukturelle Äquivalenz der zugrunde liegenden Datenbestände systematisch zu operationalisieren (OECD & European Commission, 2020, S. 10).

Ein valider Gruppenvergleich setzt voraus, dass ein Messinstrument in allen Untersuchungseinheiten dasselbe theoretische Konstrukt mit vergleichbarer Präzision erfasst (Milfont & Fischer, 2010, S. 112). Wird diese Voraussetzung verletzt, verlieren beobachtete Differenzen ihre inferenzielle Eindeutigkeit (Sterner et al., 2024, S. 748). Die zentralen Dimensionen dieser Invarianzproblematik sind schematisch dargestellt (siehe Abbildung 2).



Abbildung 3: Vergleichbarkeitsdimensionen als Invarianzebenen in Cross-City-KPI-Analysen

Quelle: Eigene Darstellung

Aus messtheoretischer Perspektive lässt sich Vergleichbarkeit daher als Invarianzproblem interpretieren. Sie betrifft nicht allein die formale Gleichheit von Indikatorbezeichnungen, sondern die strukturelle Äquivalenz der zugrunde liegenden Messmodelle.

Ein zentrales Hindernis für belastbare Inferenzschlüsse liegt in der räumlichen Inkonsistenz urbaner Daten. Nationale Definitionen dessen, was statistisch als Stadt gilt, beruhen häufig auf administrativen Grenzen, die funktionale Verflechtungsräume nur unvollständig abbilden (Dijkstra et al., 2019, S. 4). Diese Varianz steht im Zusammenhang mit dem Modifiable Areal Unit Problem, wonach die Wahl der Aggregationsgrenzen Kennzahlen systematisch beeinflussen kann (European Union et al., 2021, S. 81). Werden Städte auf Basis unterschiedlicher Raumkonzepte, etwa City Proper und Functional Urban Areas, verglichen, beruhen die resultierenden Kennzahlen auf strukturell nicht äquivalenten Einheiten (Dijkstra et al., 2019, S. 13). Räumliche Inkonsistenz kann daher als spezifische Form struktureller Nicht-Invarianz verstanden werden (Sterner et al., 2024, S. 751).

Die Problematik verschärft sich dadurch, dass zentrale Zielgrößen der Smart-City-Governance, etwa Lebensqualität oder Nachhaltigkeit, latente Konstrukte darstellen, die über beobachtbare Indikatoren operationalisiert werden (Sterner et al., 2024, S. 748; Milfont & Fischer, 2010, S. 113). Für valide Mittelwertvergleiche latenter Konstrukte ist Messinvarianz erforderlich (Milfont & Fischer, 2010, S. 112; Sterner et al., 2024, S. 747). Liegt keine hinreichende Invarianz vor, sind Unterschiede in beobachteten Indikatorwerten nicht eindeutig als Unterschiede im zugrunde liegenden Konstrukt interpretierbar (Milfont

& Fischer, 2010; Sterner et al., 2024).

Zusätzliche Herausforderungen ergeben sich aus der Architektur bestehender KPI-Systeme. Indikatorensätze weisen eine erhebliche Heterogenität hinsichtlich Zielausrichtung und Schwerpunktsetzung auf, wobei technologische Output-Indikatoren teilweise stärker gewichtet werden als Wirkungsdimensionen nachhaltiger Entwicklung (Huovila et al., 2019, S. 141). Zudem werden Unterschiede in Datenqualität und Erhebungsmethoden nicht immer hinreichend dokumentiert (Huovila et al., 2019, S. 145). Werden Schätzverfahren, abweichende Definitionen oder unterschiedliche Erhebungsdesigns nicht transparent ausgewiesen, bleibt die strukturelle Äquivalenz der Daten ungeprüft (Eurostat, 2021, S. 125).

Der Problemraum der Vergleichbarkeit ergibt sich damit aus der Interdependenz dreier Ebenen: der räumlichen Referenz, der messtheoretischen Struktur und der methodischen Dokumentationsqualität. Vergleichbarkeit ist nur dann methodisch vertretbar, wenn erstens die räumlichen Einheiten funktional äquivalent definiert sind (European Union et al., 2021, S. 82), zweitens Messmodelle hinreichende Invarianz aufweisen (Sterner et al., 2024, S. 748) und drittens Metadaten die Rekonstruktion des Messprozesses ermöglichen (Eurostat, 2021, S. 24).

Die gegenwärtige Praxis, Cross-City-Benchmarks auf Basis bloßer Namensgleichheit von Indikatoren durchzuführen, behandelt Vergleichbarkeit häufig als implizite Annahme. Methodisch bleibt sie damit ungesichert.

Vergleichbarkeit wird damit selbst zum Untersuchungsgegenstand. Daraus folgt die Notwendigkeit eines formalen, metadatenbasierten Vergleichbarkeitsrahmens, der strukturelle Abweichungen systematisch identifiziert und regelgebunden klassifiziert (Sterner et al., 2024, S. 755; Eurostat, 2021, S. 13). Erst durch eine solche Operationalisierung kann die Validität von Inferenzschlüssen in Cross-City-Analysen methodisch abgesichert werden.

3. Metadatenbasierter Vergleichbarkeitsrahmen

Der entwickelte Vergleichbarkeitsrahmen verfolgt das Ziel, Cross-City-Vergleiche nicht implizit als valide anzunehmen, sondern ihre methodische Tragfähigkeit explizit, reproduzierbar und metadatenbasiert zu prüfen (Eurostat, 2021, S. 152; Sterner et al., 2024, S. 747) Vergleichbarkeit wird dabei als relationale Eigenschaft konkreter KPI-Paare modelliert. Sie ergibt sich nicht aus der Namensgleichheit von Indikatoren, sondern aus der strukturellen Äquivalenz dokumentierter Messbedingungen (Sterner et al., 2024, S. 751).

Der Rahmen knüpft an bestehende Standardisierungs- und Interoperabilitätsansätze an, indem er strukturierte Metadaten als Voraussetzung belastbarer Datennutzung behandelt (Lóscio et al., 2017, Abschn. 8.2, O. S.; Van den Brink et al., 2019, S. 952; Eurostat, 2021, S. 9). Während bestehende Standards Definitionen harmonisieren, operationalisiert der vorliegende Ansatz eine regelbasierte Entscheidungslogik zur Prüfung konkreter Vergleichssituationen.

Vergleichbarkeit als relationale Eigenschaft

Die methodische Belastbarkeit von Cross-City-Vergleichen ist in der urbanen Governance als zu prüfende Voraussetzung zu behandeln, wenn Vergleiche als Grundlage für Inferenzschlüsse dienen sollen (Sterner et al., 2024, S. 747). In internationalen Monitoring-Initiativen, etwa im Kontext der Sustainable Development Goals, werden Indikatoren in Vergleichslogiken eingebettet, während die strukturelle Äquivalenz der zugrunde liegenden Datenbestände nicht immer explizit operationalisiert wird (OECD & European Commission, 2020, S. 10; UN-Habitat, 2025).

Für Cross-City-Analysen folgt daraus, dass Vergleichbarkeit nicht allein als Frage identischer Indikatorbezeichnungen, sondern als Frage der Äquivalenz von Messbedingungen und Messmodellen verstanden werden muss (Eurostat, 2021, S. 147; Milfont & Fischer, 2010, S. 116; Sterner et al., 2024, S.748).

Core Template des Vergleichbarkeitsrahmens

Zur Operationalisierung wird ein standardisierter KPI-Steckbrief definiert (siehe Tabelle 2). Dieser enthält strukturierte Metadatenfelder, die für jeden Indikator verpflichtend dokumentiert werden müssen.

Tabelle 5 KPI-Steckbrief Core Template - Metadatenfelder, Definition und Beispiel

Feld	Beschreibung / Definition	Beispielwert
KPI-Name	Offizielle Bezeichnung des Indikators	CO_2-Emissionen pro Einwohner
Normreferenz	Bezug zu Standardframeworks oder Indikatorensystemen	ISO 37120
Version	Versionsstand des Indikators oder Datensatzes	Version 2022
Raumtyp	Räumliche Referenzeinheit des Indikators	City Proper
Aggregationsebene	Ebene der Datenauswertung	Stadt
Raumabgrenzung	Beschreibung der geographischen Abgrenzung	Administrative Stadtgrenze
Referenzjahr	Jahr der Datenerhebung	2022
Bezugszeitraum	Zeitraum der Datenerhebung	Kalenderjahr
Periodizität	Häufigkeit der Datenerhebung	jährlich
Berechnungsformel	Mathematische Definition des Indikators	CO_2-Emissionen / Einwohner
Zähler	Zähler der Berechnung	Gesamte energiebedingte CO_2-Emissionen
Nenner	Nenner der Berechnung	Einwohnerzahl
Einheit	Maßeinheit des Indikators	t CO_2 pro Einwohner
Datenquelle	Institution oder Datensatz	Amt für Statistik Berlin-Brandenburg
Erhebungsmethode	Art der Datenerhebung	administratives Inventar
Schätzverfahren	Kennzeichnung möglicher Schätzmethoden	keine
Qualitätskennzeichnung	Angabe zur Datenqualität	validiert
Dokumentationsgrad	Vollständigkeit der Metadaten	vollständig

Diese Struktur orientiert sich an Anforderungen an Metadatenqualität und Provenienz (Eurostat, 2021, S. 31; Albertoni & Isaac, 2016; Van den Brink et al., 2019, Abschnitt 3), erweitert diese jedoch um eine explizite Vergleichbarkeitsdimension.

Regelbasierte Entscheidungslogik

Die Vergleichbarkeit zweier KPI-Datensätze wird durch eine formale Entscheidungslogik geprüft. Diese basiert auf sequenziellen Prüfregeln (siehe Abbildung 3):

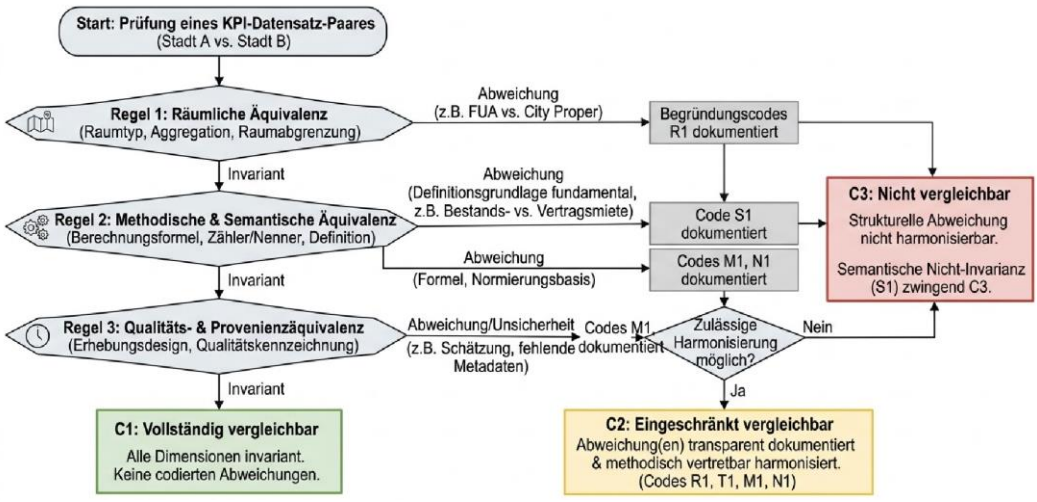


Abbildung 4: Entscheidungslogik des metadatenbasierten Vergleichbarkeitsrahmens

Quelle: Eigene Darstellung

Regel 1: Räumliche Äquivalenz Wenn Raumtyp oder Aggregationslogik strukturell divergieren, erfolgt Klassifikation als eingeschränkt vergleichbar oder nicht vergleichbar.

Regel 2: Methodische Äquivalenz Abweichungen in Berechnungsformel, Zähler, Nenner oder Definitionsgrundlage führen zu einer strukturellen Abwertung der Vergleichbarkeit.

Regel 3: Qualitäts- und Provenienzäquivalenz Unterschiedliche Erhebungsdesigns oder fehlende

Qualitätskennzeichnungen beeinflussen die Klassifikation gemäß dokumentierter Begründungscodes.

Das Ergebnis ist keine binäre Entscheidung, sondern eine dreistufige Klassifikation: **C1:** Vollständig vergleichbar, **C2:** Eingeschränkt vergleichbar, **C3:** Nicht vergleichbar

Diese Klassifikation wird mit dokumentierten Begründungscodes hinterlegt, um Transparenz und Reproduzierbarkeit sicherzustellen.

Der Rahmen erlaubt explizit definierte Harmonisierungsschritte, sofern diese dokumentiert und nachvollziehbar sind. Dazu zählen etwa:

- Umrechnung auf identische Raumtypen
- Standardisierung von Bezugsgrößen
- Offenlegung von Schätzverfahren

Harmonisierung darf jedoch keine strukturelle Nicht-Invarianz verdecken. Jede Anpassung ist daher mit einem expliziten Transformationshinweis zu kennzeichnen.

Der Rahmen versteht sich nicht als Test latenter Messinvarianz im strengen statistischen Sinne, sondern als operative Prüfung von Datenprozess-Invarianz. Er übersetzt messtheoretische Anforderungen an Vergleichbarkeit in metadatenbasierte Prüfregeln (Milfont & Fischer, 2010, S. 116; Sterner et al., 2024, S. 755; Eurostat, 2021, S. 147). Vergleichbarkeit wird damit von einer impliziten Annahme zu einer dokumentierten, regelgebundenen Eigenschaft konkreter Datensätze. Die nachfolgende Fallstudie prüft, ob die definierten Vergleichsdimensionen und Entscheidungsregeln in realen Datensätzen praktikabel anwendbar sind.

KPI-Steckbrief Framework

Aufbauend auf der konzeptionellen Herleitung eines metadatenbasierten Vergleichbarkeitsrahmens wird im Folgenden die operative Struktur des KPI-Steckbriefs definiert. Während die vorhergehenden Abschnitte die normative und interoperabilitätsbezogene Einbettung (u. a. DCAT, PROV-O, DQV) begründet haben, erfolgt hier die methodische Konkretisierung in Form eines strukturierten Metadatenprofils, das als Prüfgrundlage für Vergleichbarkeitsentscheidungen dient. Die in Tabelle 2 definierten Metadatenfelder bilden die Grundlage der Vergleichbarkeitsprüfung.

Ziel des KPI-Steckbriefs ist es, jene Merkmale systematisch zu erfassen, die für die Beurteilung semantischer, räumlicher, zeitlicher und methodischer Invarianz erforderlich sind. Metadaten werden damit nicht lediglich als Dokumentationsinstrument verstanden, sondern als entscheidungsrelevante Prüfstruktur. Dieser Zugang entspricht den Grundprinzipien der Data on the Web Best Practices, wonach strukturierte Metadaten, Angaben zur Provenienz sowie Qualitätsinformationen zentrale Voraussetzungen für Interoperabilität, Wiederverwendbarkeit und belastbare Datennutzung darstellen (W3C, Data on the Web Best Practices, 2017). Ebenso betont das DCAT-Modell die Notwendigkeit standardisierter Metadatenelemente zur interoperablen Beschreibung und Auffindbarkeit verteilter Datensätze (Maali & Erickson, Data Catalog Vocabulary (DCAT), W3C Recommendation, 2014).

Struktur des KPI-Steckbriefs

Der Steckbrief besteht aus verpflichtenden Prüffeldern, die für jede Vergleichseinheit vollständig auszufüllen sind. Fehlen einzelne Angaben, ist die Vergleichbarkeitsprüfung formal eingeschränkt; in diesem Fall wird ein Unsicherheitscode (U) vergeben. Der vorgeschlagene KPI-Steckbrief umfasst folgende verpflichtende Metadatenfelder:

(1) Eindeutige Bezeichnung des KPI Eine konsistente Benennung dient der Identifikation sowie der möglichen Zuordnung zu bestehenden Indikatorensystemen (z. B. ISO-Reihen). Die bloße Namensgleichheit stellt jedoch kein Vergleichbarkeitskriterium dar, sondern lediglich einen Referenzpunkt für die weitergehende Prüfung.

(2) Formale Definition und Berechnungslogik Die mathematische oder algorithmische Spezifikation des Indikators (z. B. Quotient, Aggregationsregel, Gewichtung). Nur wenn die zugrunde liegende

Berechnungsstruktur transparent dokumentiert ist, kann semantische Gleichheit überprüft werden. Internationale Standardisierungsrahmen wie ISO 37120 betonen ausdrücklich, dass Indikatoren präzise definiert und methodisch nachvollziehbar beschrieben sein müssen, um Vergleichbarkeit zwischen Städten zu ermöglichen (ISO 37120:2018).

(3) Einheit und Skalenniveau Angabe der Maßeinheit (z. B. %, t CO_2 pro Jahr, Personen pro 100 000 Einwohner) sowie des Skalenniveaus. Identische Einheiten sind notwendige, jedoch keine hinreichende Bedingung für Vergleichbarkeit. Implizite Normalisierungen oder verdeckte Transformationen werden gesondert dokumentiert.

(4) Räumliche Abdeckung Präzise Beschreibung der zugrunde liegenden Gebietseinheit (administrative Stadtgrenze, funktionaler Urbanraum, Metropolregion etc.). Räumliche Inkonsistenzen gelten als zentrale Ursache verzerrter Stadtvergleiche. Die Relevanz klar definierter Raumbezüge wird in der europäischen Städte-Methodik sowie in der OECD-EU-Definition funktionaler Urbanräume hervorgehoben, die zwischen administrativen Städten, Urban Centres und Functional Urban Areas differenzieren (OECD & European Commission, 2020, S. 24).

(5) Zeitlicher Referenzrahmen Angabe des Bezugsjahres bzw. Zeitraums sowie der Aggregationsform (z. B. Jahresmittel, Stichtag). Temporale Inkonsistenzen können Rangfolgen systematisch beeinflussen und sind daher eigenständiges Prüfkriterium. Fragen der zeitlichen Kohärenz und Vergleichbarkeit werden auch im Kontext internationaler Indikatorenkonstruktion thematisiert (OECD & European Commission, 2008, S. 14).

(6) Datenquelle und Erhebungsmethode Dokumentation der Herkunft (z. B. kommunale Statistikstelle, nationale Behörde, Modellrechnung) sowie der Erhebungslogik (administrative Daten, Stichprobe, Simulation). Provenienzangaben sind essenziell, um methodische Unterschiede nachvollziehbar zu machen; entsprechende Modellierungen werden im Provenance-Ontology-Standard formalisiert (Lebo et al., PROV-O: The PROV Ontology, W3C Recommendation, 2013).

Diese Pflichtfelder bilden die minimale Informationsbasis für eine strukturierte Vergleichbarkeitsprüfung. Fehlen einzelne dieser Angaben, ist die Prüfbarkeit eingeschränkt. Ergänzend können folgende Metadaten aufgenommen werden, sofern verfügbar:

- Normalisierungsregel (z. B. Pro-Kopf-Berechnung, Flächennormierung)
- Unsicherheits- oder Qualitätsangaben (z. B. Standardfehler, Schätzverfahren), im Sinne des Data Quality Vocabulary (Albertoni & Isaac, 2016)
- Referenz auf Standardrahmen (z. B. ISO-Indikator, nationale Leitlinie)
- Hinweise zur potenziellen Harmonisierung

Diese Angaben erweitern die Bewertungsgrundlage, sind jedoch nicht zwingend erforderlich, um eine formale Vergleichbarkeitsklassifikation vorzunehmen.

Der KPI-Steckbrief stellt die strukturierte Eingangsebene für die nachfolgend entwickelte Entscheidungslogik dar. Erst durch die explizite Erfassung der genannten Dimensionen wird es möglich, Vergleichbarkeit nicht implizit anzunehmen, sondern regelbasiert zu prüfen.

Damit verschiebt sich der Fokus von einer rein semantischen Gleichsetzung von Indikatornamen hin zu einer mehrdimensionalen Invarianzprüfung, die auf dokumentierten Metadaten beruht. Der Steckbrief operationalisiert somit die zuvor theoretisch begründete Annahme, dass Vergleichbarkeit eine prüfbare Eigenschaft von Datenstrukturen ist und keine bloße Interpretation.

Auf Grundlage der strukturierten Metadaten werden fünf analytische Vergleichsdimensionen unterschieden:

(1) Semantische Vergleichbarkeit Prüfung, ob Definition, Zielgröße und Berechnungslogik inhaltlich übereinstimmen. Bereits geringfügige Abweichungen (z. B. unterschiedliche Bezugsgrößen oder Aggregationsregeln) können zu nicht-invarianten Messungen führen. Die Bedeutung präziser Indikatordefinitionen wird auch in internationalen Standardisierungsrahmen wie ISO 37120 hervorgehoben (ISO 37120:2018).

(2) Räumliche Vergleichbarkeit Bewertung, ob identische oder funktional äquivalente Raumabgrenzungen vorliegen. Unterschiede zwischen administrativen Grenzen und funktionalen Urbanräumen gelten als strukturelle Quelle systematischer Verzerrung (OECD & European Commission, 2020, S. 24).

(3) Temporale Vergleichbarkeit Prüfung identischer Referenzjahre oder methodisch vertretbarer Zeiträume. Fragen der zeitlichen Kohärenz und Datenharmonisierung werden im OECD-Handbook zur Konstruktion von Indikatoren als zentrale Qualitätsdimension diskutiert (OECD & European Commission, 2008, S. 14).

(4) Methodische Vergleichbarkeit Analyse der Erhebungs- und Berechnungsmethoden. Unterschiede zwischen administrativen Vollerhebungen und modellbasierten Schätzungen können systematische Effekte erzeugen und die Vergleichbarkeit beeinträchtigen.

(5) Normierungs- und Skalierungsvergleichbarkeit Prüfung, ob identische Bezugsgrößen (z. B. Pro-Kopf-Normierung) verwendet wurden oder ob eine zulässige Transformation möglich ist.

Diese Dimensionen bilden die strukturierte Bewertungsgrundlage für die nachfolgende Klassifikation.

Entscheidungslogik und Klassifikationssystem

Auf Grundlage der fünf Vergleichsdimensionen wird eine dreistufige Klassifikation eingeführt, die den Grad der Vergleichbarkeit zwischen zwei KPI-Datensätzen abbildet. Ziel dieser Klassifikation ist es, Vergleichbarkeit nicht pauschal anzunehmen, sondern nachvollziehbar aus den dokumentierten Metadaten abzuleiten.

Ein KPI-Paar wird als **C1 – vergleichbar** eingestuft, wenn in allen relevanten Dimensionen Invarianz vorliegt oder wenn bestehende Abweichungen durch vollständig regelkonforme Transformationen ausgeglichen werden können. In diesem Fall ist ein direkter Vergleich methodisch vertretbar.

Die Kategorie **C2 – eingeschränkt vergleichbar** wird vergeben, wenn eine oder mehrere Dimensionen Abweichungen aufweisen, diese jedoch transparent dokumentiert sind und durch zulässige Harmonisierung oder durch eine entsprechend vorsichtige Interpretation berücksichtigt werden können. Ein Vergleich ist in diesem Fall möglich, darf jedoch nicht ohne Hinweis auf die methodischen Einschränkungen erfolgen.

Als **C3 – nicht vergleichbar** werden KPI-Paare klassifiziert, wenn mindestens eine zentrale Vergleichsdimension verletzt ist und keine methodisch vertretbare Harmonisierung möglich erscheint. Dies betrifft insbesondere Fälle semantischer oder methodischer Nicht-Invarianz, in denen die betrachteten Datensätze trotz ähnlicher Bezeichnung nicht dieselbe konzeptionelle Zielgröße erfassen.

Zur Sicherstellung der Transparenz werden festgestellte Abweichungen zusätzlich durch standardisierte Begründungscodes dokumentiert. Der Code **S1** verweist auf eine abweichende Definition oder Zielgröße des Indikators, **R1** auf unterschiedliche räumliche Abgrenzungen, **T1** auf zeitliche Inkonsistenzen, **M1** auf methodische Unterschiede und **N1** auf abweichende Normierungs- oder Skalierungsgrundlagen. Auf diese Weise wird nicht nur das Ergebnis der Klassifikation sichtbar, sondern auch der konkrete Grund, aus dem ein Vergleich als uneingeschränkt, eingeschränkt oder nicht zulässig bewertet wird.

Die Bedeutungen der einzelnen Codes sind zusammengefasst (siehe Tabelle 3)

Tabelle 3:6 Begründungscodes des Vergleichbarkeitsrahmens

Code	Dimension	Bedeutung	Typische Auswirkung auf Klassifikation	Harmonisierung möglich
S1	Semantische Vergleichbarkeit	Unterschiedliche Definition oder konzeptionelle Zielgröße des Indikators	führt direkt zu C3 (nicht vergleichbar)	nein
R1	Räumliche Vergleichbarkeit	Unterschiedliche Raumabgrenzung (z. B. City Proper vs Functional Urban Area)	typischerweise C2 (eingeschränkt vergleichbar)	teilweise
T1	Temporale Vergleichbarkeit	Unterschiedliche Referenzjahre oder Aggregationslogiken	typischerweise C2, sofern Harmonisierung möglich	teilweise
M1	Methodische Vergleichbarkeit	Unterschiedliche Erhebungs- oder Berechnungsmethoden	typischerweise C2, sofern keine semantische Nicht-Invarianz vorliegt	teilweise
N1	Normierungs- / Skalierungsvergleichbarkeit	Unterschiedliche Normierungsbasis (z. B. Bevölkerungsdefinition)	typischerweise C2, sofern Transformation möglich	teilweise
U1	Metadatenunsicherheit	Fehlende oder unvollständige Metadokumentation	typischerweise C2 aufgrund eingeschränkter Prüfbarkeit	nein

Die Entscheidungslogik des Frameworks folgt dem Prinzip der minimalen Annahme. Vergleichbarkeit wird nicht vorausgesetzt, sondern nur dann bestätigt, wenn zwischen zwei KPI-Datensätzen keine relevante dimensionale Inkonsistenz festgestellt wird oder wenn bestehende Abweichungen transparent dokumentiert und regelkonform harmonisiert werden können. Damit wird vermieden, dass formale Ähnlichkeiten — etwa identische Indikatornamen oder Maßeinheiten — vorschnell als Beleg für tatsächliche Vergleichbarkeit interpretiert werden.

Gleichzeitig bedeutet eine festgestellte Abweichung nicht automatisch, dass ein Vergleich ausgeschlossen werden muss. Unter bestimmten Bedingungen sind Transformationen zulässig, sofern sie eindeutig nachvollziehbar, methodisch konsistent und vollständig dokumentiert sind. Dazu zählen insbesondere rein mathematische Einheitenkonversionen, etwa von Kilogramm in Tonnen, explizite Pro-Kopf-Normierungen auf Grundlage identischer Bevölkerungsdefinitionen sowie Aggregationen oder Disaggregationen, sofern die zugrunde liegenden räumlichen Referenzeinheiten miteinander kompatibel sind. Solche Anpassungen können die Vergleichbarkeit verbessern, dürfen jedoch nicht dazu führen, strukturelle Unterschiede der Datensätze zu verdecken.

Nicht zulässig sind hingegen nachträgliche Schätzungen ohne dokumentierte Methodik, implizite Annahmen über fehlende Daten oder Harmonisierungsschritte bei fundamental unterschiedlichen Definitionsgrundlagen. Gerade in solchen Fällen würde eine rechnerische Anpassung lediglich den Eindruck von Vergleichbarkeit erzeugen, ohne dass tatsächlich dieselbe Zielgröße gemessen wird. Das Framework folgt damit einem methodischen Vorsichtsprinzip: Harmonisierung ist nur dort vertretbar, wo sie die Vergleichbarkeit nachvollziehbar herstellt, nicht aber dort, wo sie konzeptionelle Unterschiede überdeckt.

Dieses Vorgehen entspricht den methodischen Empfehlungen zur transparenten Indikatorenkonstruktion und Datenharmonisierung, wie sie in internationalen Leitfäden zur Erstellung zusammengesetzter Indikatoren formuliert werden (OECD & European Commission, 2008, S. 14). Der entwickelte Rahmen verbindet damit drei Elemente: Der KPI-Steckbrief stellt die strukturierte Metadatenbasis bereit, die Vergleichsdimensionen bilden die analytische Prüfgrundlage, und die Entscheidungslogik übersetzt dokumentierte Abweichungen in eine nachvollziehbare Klassifikation. Die Harmonisierungsvorschriften sichern dabei, dass Anpassungen nicht beliebig erfolgen, sondern an klar definierte methodische Bedingungen gebunden bleiben.

Damit ist der Vergleichbarkeitsrahmen methodisch operationalisiert. Im folgenden Kapitel wird er auf ausgewählte städtische KPI-Datensätze angewendet, um zu prüfen, welche Wirkung die metadatenbasierte Bewertung gegenüber einer rein formalen Baseline-Annahme entfaltet.

4. Fallstudie: Anwendung auf Berlin, Paris und Barcelona

Zur Evaluation des entwickelten Vergleichbarkeitsrahmens wird eine Fallstudie durchgeführt, in der eine konventionelle, formale Vergleichslogik mit der metadatenbasierten Prüfung des vorgeschlagenen Frameworks gegenübergestellt wird. Damit soll gezeigt werden, ob und in welchem Umfang sich die Einschätzung zulässiger Cross-City-Vergleiche verändert, wenn Vergleichbarkeit nicht nur über Indikatornamen, Maßeinheit und Referenzjahr bestimmt, sondern anhand dokumentierter Metadaten systematisch geprüft wird (W3C, 2017, Abschn. 8.4 und 8.5; ISO 37120:2018).

Als Referenz dient zunächst ein Baseline-Modell, das Vergleichbarkeit ausschließlich über formale Indikatoridentität definiert. Ein KPI-Paar gilt in diesem Modell dann als vergleichbar, wenn der Indikatorname, die Maßeinheit und das Referenzjahr identisch sind. Weitere Dimensionen wie Definition, Raumabgrenzung, Erhebungsmethode oder Normierungslogik werden in dieser Perspektive nicht berücksichtigt. Das Baseline-Modell bildet damit jene Annahme ab, die vielen Benchmarking-Praktiken implizit zugrunde liegt: Formal gleich bezeichnete Kennzahlen werden als unmittelbar vergleichbar behandelt.

Untersucht werden die Städte Berlin, Paris und Barcelona. Diese Auswahl ist zum einen durch ihre hohe Relevanz in internationalen Stadtvergleichen begründet, zum anderen durch die Verfügbarkeit umfangreicher Open-Data- und Statistikangebote. Zudem eignen sich die drei Städte besonders für die Fragestellung der Arbeit, weil sie unterschiedliche räumliche Referenzdefinitionen aufweisen. Der Unterschied zwischen administrativen Stadtgrenzen („City Proper“) und funktionalen Stadträumen („Functional Urban Area“) wird in internationalen Urbanisierungsansätzen ausdrücklich als methodisch relevant und nicht ohne Weiteres äquivalent behandelt (OECD & European Commission, 2020, S. 24).

Für die Fallstudie werden drei Indikatoren ausgewählt: energiebedingte CO_2-Emissionen pro Einwohner, Verkehrstote im Stadtgebiet sowie Nettokaltmiete pro Quadratmeter. Die Auswahl erfolgt nicht allein aufgrund ihrer politischen und praktischen Relevanz, sondern vor allem aufgrund ihrer unterschiedlichen Invarianzstruktur. CO_2-Emissionen eignen sich zur Prüfung abweichender Bilanzierungsansätze und Inventarsysteme und adressieren damit insbesondere die semantisch-methodische Vergleichbarkeit. Der Indikator Verkehrstote dient als Testfall für scheinbar standardisierte Sicherheitskennzahlen, bei denen dennoch Definitionsannahmen, etwa die 30-Tage-Regel, unterschiedlich dokumentiert sein können. Mietindikatoren wiederum bilden einen besonders aufschlussreichen Fall semantischer Nicht-Invarianz, da Vergleichsmieten, Bestandsmieten und Vertragsmieten unterschiedliche konzeptionelle Zielgrößen erfassen. Gerade an diesen Beispielen lässt sich zeigen, dass internationale Standardisierungsrahmen zwar präzise Definitionen und methodische Transparenz einfordern, die konkrete Vergleichbarkeit einzelner Datensätze jedoch zusätzlich geprüft werden muss (ISO 37120:2018; OECD & European Commission, 2008, S. 14).

Die Analyse stützt sich ausschließlich auf öffentlich zugängliche Primärquellen. Für jeden der drei Indikatoren wird ein KPI-Steckbrief gemäß dem in Kapitel 4 entwickelten Rahmen erstellt. Anschließend werden die Datensätze entlang der fünf Vergleichsdimensionen geprüft: semantische, räumliche, temporale, methodische sowie normierungs- und skalierungsbezogene Vergleichbarkeit. Metadatenlücken, die für die Invarianzprüfung relevant sind, werden mit dem Code U1 gekennzeichnet. Die Klassifikation folgt der zuvor definierten Entscheidungslogik: C1

steht für vollständig vergleichbare Datensätze ohne codierte Abweichung, C2 für eingeschränkt vergleichbare Datensätze mit dokumentierten, aber methodisch vertretbaren Abweichungen, und C3 für nicht vergleichbare Datensätze, insbesondere bei semantischer Nicht-Invarianz oder nicht harmonisierbaren methodischen Unterschieden. Um den Mehrwert des Frameworks sichtbar zu machen, wird für jeden KPI zunächst die Baseline-Klassifikation dargestellt und anschließend mit der metadatenbasierten Framework-Klassifikation verglichen.

Beim ersten Indikator, den CO_2-Emissionen pro Einwohner, weist Berlin für das Jahr 2022 energiebedingte CO_2-Emissionen pro Einwohner aus (Amt für Statistik Berlin-Brandenburg, 2022a, S. 24). Paris berichtet für dasselbe Referenzjahr territoriale Treibhausgasemissionen (Ville de Paris, 2025). Nach dem Baseline-Modell erscheint der Indikator zunächst vergleichbar, da Bezeichnung, Einheit und Referenzjahr formal übereinstimmen. Die metadatenbasierte Prüfung zeigt jedoch ein anderes Bild. Unterschiedliche Bilanzierungs- und Inventarmethoden führen zu einer semantischen Abweichung (S1). Hinzu kommen Unterschiede im räumlichen Bezug, da administrative Stadtgrenzen zwar formal ähnlich erscheinen, funktionale Urbanräume jedoch strukturell nicht äquivalent sind (R1). Darüber hinaus liegen methodische Unterschiede zwischen den Emissionsinventarsystemen vor (M1), und auch die Pro-Kopf-Normierung kann auf unterschiedlichen Bevölkerungsdefinitionen beruhen (N1). Da eine semantische Nicht-Invarianz vorliegt, wird der Indikator gemäß der Entscheidungslogik als nicht vergleichbar klassifiziert (C3). Eine nachträgliche Harmonisierung ist nicht zulässig, weil unterschiedliche Bilanzierungslogiken die konzeptionelle Zielgröße des Indikators verändern.

Der zweite Indikator betrifft Verkehrstote im Stadtgebiet. Für Berlin werden im Jahr 2022 insgesamt 34 Verkehrstote ausgewiesen (Amt für Statistik Berlin-Brandenburg, 2022b, S. 5). Für Paris und Barcelona liegen entsprechende Open-Data-Datensätze vor (Paris Open Data, 2024; Ajuntament de Barcelona, o. J.). Im Baseline-Modell gilt auch dieser Indikator als vergleichbar, da Indikatorname, Einheit und Referenzjahr übereinstimmen. Die Framework-Prüfung zeigt jedoch, dass die definitorische Grundlage, insbesondere mit Blick auf die 30-Tage-Regel, nicht in allen Datensätzen eindeutig dokumentiert ist. Diese Unsicherheit wird mit U1 codiert. Räumlich beziehen sich die Daten zwar jeweils auf das Stadtgebiet, und auch das Kalenderjahr 2022 ist konsistent. Gleichzeitig bestehen jedoch Unterschiede in den Erhebungs- und Klassifikationsverfahren, die als methodische Abweichung (M1) erfasst werden. Da keine gesicherte semantische Nicht-Invarianz vorliegt, der Vergleich aber durch methodische Unterschiede und unvollständige Metadokumentation eingeschränkt ist, wird der Indikator als eingeschränkt vergleichbar klassifiziert (C2). Ein Vergleich ist somit möglich, aber nur unter ausdrücklicher Dokumentation der methodischen Einschränkungen.

Der dritte Indikator, die Nettokaltmiete pro Quadratmeter, zeigt die Problematik besonders deutlich. Berlin berichtet die ortsübliche Vergleichsmiete auf Grundlage des Mietspiegels (Land Berlin, 2023, S. 5), Paris verwendet Daten des Observatoire des loyers de l'agglomération parisienne (OLAP, 2022), während Barcelona durchschnittliche Vertragsmieten veröffentlicht (Ajuntament de Barcelona, o. J.). Im Baseline-Modell erscheinen diese Werte formal vergleichbar, da sie als Mietpreise pro Quadratmeter ausgewiesen werden. Die metadatenbasierte Prüfung zeigt jedoch eine fundamentale semantische Nicht-Invarianz: Vergleichsmiete, Bestandsmiete und Vertragsmiete erfassen unterschiedliche konzeptionelle Zielgrößen. Zusätzlich bestehen temporale Abweichungen durch unterschiedliche Stichtage und Aggregationslogiken (T1) sowie methodische Unterschiede in Datenerhebung und Berechnung (M1). Die formal identische Einheit €/m² kann diese konzeptionellen Unterschiede nicht ausgleichen. Aufgrund der semantischen Nicht-Invarianz wird der Mietindikator daher als nicht vergleichbar eingestuft (C3). Ein direktes Ranking der Mietpreise zwischen den Städten wäre methodisch nicht zulässig.

Die Fallstudie zeigt damit, dass die metadatenbasierte Prüfung zu einer erheblichen Neubewertung der Vergleichbarkeit führt. Während nach dem Baseline-Modell alle drei Indikatoren als vergleichbar gelten würden, ergibt die Framework-Prüfung kein einziges vollständig vergleichbares KPI-Paar. Die CO_2-Emissionen pro Einwohner und die Nettokaltmiete pro Quadratmeter werden als nicht vergleichbar klassifiziert, während der Indikator Verkehrstote lediglich eingeschränkt vergleichbar ist. Die Zahl uneingeschränkt zulässiger Vergleiche reduziert sich damit von drei auf null. Zugleich wird deutlich, dass der entwickelte Rahmen nicht nur formale Unterschiede dokumentiert, sondern methodisch problematische Rangreihen verhindert: Ein Emissionsranking und ein Mietranking wären auf dieser Datengrundlage unzulässig, ein Verkehrssicherheitsranking wäre nur eingeschränkt interpretierbar.

Vergleichbarkeitsmatrix und empirische Wirkung

Die Ergebnisse der Fallstudie werden in Tabelle 5 zusammenfassend dargestellt. Die Gegenüberstellung zeigt, wie deutlich sich die Bewertung der drei untersuchten KPIs verändert, sobald nicht nur formale Kriterien wie Indikatorname, Einheit und Referenzjahr berücksichtigt werden, sondern zusätzlich die im Framework definierten Metadaten geprüft werden. Während das Baseline-Modell alle drei Kennzahlen als vergleichbar einstuft, führt die metadatenbasierte Analyse zu einer deutlich restriktiveren Bewertung.

Tabelle 5: Ergebnis Matrix

KPI	Baseline-Klassifikation	Framework-Klassifikation	Begründungscodes	Kurze Begründung
CO_2-Emissionen pro Einwohner	vergleichbar	C3: nicht vergleichbar	S1, R1, M1, N1	Unterschiedliche Bilanzierungslogiken und Inventarsysteme
Verkehrstote	vergleichbar	C2: eingeschränkt vergleichbar	M1, U1	Methodische Unterschiede und unvollständige Definitionsdokumentation
Nettokaltmiete pro Quadratmeter	vergleichbar	C3: nicht vergleichbar	S1, T1, M1	Unterschiedliche konzeptionelle Mietdefinitionen

Die Ergebnismatrix macht sichtbar, dass die formale Übereinstimmung eines Indikators nicht ausreicht, um einen belastbaren Städtevergleich zu begründen. Nach dem Baseline-Modell wären alle drei KPIs für einen direkten Cross-City-Vergleich zugelassen worden. Die Anwendung des entwickelten Frameworks zeigt jedoch, dass keiner der untersuchten Indikatoren vollständig vergleichbar ist. Der Indikator zu den Verkehrstoten wird lediglich als eingeschränkt vergleichbar eingestuft, während die CO_2-Emissionen pro Einwohner und die Nettokaltmiete pro Quadratmeter als nicht vergleichbar klassifiziert werden.

Damit reduziert sich die Zahl uneingeschränkt zulässiger Vergleiche von drei auf null. Besonders deutlich wird der Mehrwert des Frameworks daran, dass alle drei Indikatoren gegenüber der Baseline-Annahme neu bewertet werden. Die CO_2-Emissionen wechseln von „vergleichbar“ zu „nicht vergleichbar“, weil unterschiedliche Bilanzierungslogiken, Inventarsysteme, Raumbezüge und Normierungsgrundlagen vorliegen. Der Indikator Verkehrstote bleibt zwar grundsätzlich interpretierbar, wird jedoch aufgrund methodischer Unterschiede und unvollständiger Definitionsdokumentation nur als eingeschränkt vergleichbar klassifiziert. Beim Mietindikator führt die unterschiedliche konzeptionelle Grundlage der Daten — Vergleichsmiete, Bestandsmiete und Vertragsmiete — zu einer eindeutigen Einstufung als nicht vergleichbar.

Diese Neubewertung hat unmittelbare Konsequenzen für mögliche Rangreihen. Unter der Baseline-Annahme hätten für alle drei Indikatoren Rankings erstellt werden können. Nach Anwendung des Frameworks wären jedoch sowohl ein Emissionsranking als auch ein Mietranking methodisch unzulässig. Ein Vergleich der Verkehrssicherheit wäre nur eingeschränkt interpretierbar und müsste die dokumentierten methodischen Unsicherheiten ausdrücklich ausweisen. Das Framework verhindert damit zwei potenziell irreführende Rangfolgen und zeigt zugleich, dass auch scheinbar einfache Kennzahlen einer systematischen Vergleichbarkeitsprüfung bedürfen.

Diskussion, Limitationen und Implikationen

Die Fallstudie verdeutlicht zunächst, dass die Verfügbarkeit und Qualität von Metadaten eine zentrale Voraussetzung für belastbare Cross-City-Vergleiche darstellen. Besonders deutlich wird dies beim Indikator „Verkehrstote". Obwohl dieser Indikator auf den ersten Blick standardisiert erscheint, ist die definitorische Grundlage — etwa mit Blick auf die 30-Tage-Regel — nicht in allen Datensätzen eindeutig dokumentiert. Diese fehlende Transparenz führt zur Codierung U1 und damit zu einer eingeschränkten Vergleichbarkeit. Ohne die explizite Prüfung der Metadaten wäre der Indikator im Baseline-Modell fälschlicherweise als vollständig vergleichbar behandelt worden. Die Vergleichbarkeit urbaner Kennzahlen hängt somit nicht nur von der Verfügbarkeit der Daten selbst, sondern wesentlich von der Qualität ihrer methodischen Dokumentation ab.

Zugleich zeigt die Analyse, dass nicht jede Abweichung durch Harmonisierung ausgeglichen werden kann. Beim Mietindikator liegt eine fundamentale semantische Nicht-Invarianz vor, da Vergleichsmiete, Bestandsmiete und Vertragsmiete unterschiedliche Zielgrößen erfassen. Eine rechnerische Transformation kann diese konzeptionellen Unterschiede nicht beheben. Das Framework verhindert an dieser Stelle eine methodisch unzulässige Ex-post-Harmonisierung und macht deutlich, dass Vergleichbarkeit dort endet, wo unterschiedliche Indikatoren zwar ähnlich benannt sind, aber nicht dasselbe messen.

Eine weitere Limitation betrifft die verbleibenden Interpretationsspielräume. Auch ein formalisiertes Regelwerk kann fachliche Einschätzungen nicht vollständig ersetzen. Dies zeigt sich beispielsweise bei der Bewertung methodischer Unterschiede in CO_2-Inventaren. Die Entscheidungslogik reduziert Subjektivität, indem sie Abweichungen explizit codiert und dokumentiert, eliminiert sie jedoch nicht vollständig. Ihr Mehrwert besteht daher weniger darin, jede Unsicherheit aufzulösen, sondern vielmehr darin, Bewertungsentscheidungen transparent, nachvollziehbar und überprüfbar zu machen.

Schließlich verweist die Fallstudie auf Fragen der Skalierbarkeit. Die Anwendung des Frameworks auf drei KPIs ist gut nachvollziehbar und methodisch handhabbar. Bei umfangreicheren Indikatorensets würde der Dokumentations- und Prüfaufwand jedoch deutlich steigen. Die praktische Nutzbarkeit des Ansatzes hängt daher langfristig davon ab, ob Metadaten stärker standardisiert erfasst und zumindest teilweise automatisiert ausgewertet werden können. Ohne strukturierte Metadaten wäre die Anwendung des Frameworks bei großen Datensätzen operativ anspruchsvoll. Gerade deshalb zeigt die Studie, dass Vergleichbarkeit nicht erst im Moment der Analyse, sondern bereits bei der Bereitstellung urbaner Daten mitgedacht werden muss.

6. Fazit

Der vorliegende Beitrag nahm seinen Ausgangspunkt in einem grundlegenden Validitätsproblem aktueller Smart-City-Forschung: der Praxis, Cross-City-Vergleiche auf der naiven Annahme aufzubauen, dass Namensgleichheit, identische Einheiten und dasselbe Referenzjahr eine hinreichende Bedingung für Vergleichbarkeit darstellen. Um die Gefahr zu minimieren, methodische Artefakte als reale Leistungsunterschiede fehlzuinterpretieren, wurden drei zentrale Forschungsfragen adressiert, deren

Beantwortung bestehende KPI-Frameworks um eine essenzielle Validitätsprüfung auf Datenebene erweitert.

Zur Beantwortung der ersten Forschungsfrage, wie sich Vergleichbarkeit als formalisierbare Eigenschaft modellieren lässt, wurde gezeigt, dass Vergleichbarkeit keine triviale, inhärente Eigenschaft eines Indikators an sich ist. Stattdessen wurde sie erfolgreich als relationale Eigenschaft konkreter KPI-Paare und als mehrdimensionale Invarianzbedingung konzeptualisiert. Eine valide Inferenz setzt demnach voraus, dass die strukturelle Äquivalenz auf semantischer, räumlicher, temporaler, methodischer und normierungsbezogener Ebene systematisch gesichert ist.

Hinsichtlich der zweiten Forschungsfrage: der Entwicklung eines metadatenbasierten Entscheidungsrahmens, wurde ein operativer Ansatz in Form eines strukturierten KPI-Steckbriefs entwickelt. Durch eine sequenzielle, regelbasierte Prüflogik lassen sich strukturelle Abweichungen zwischen Datensätzen systematisch identifizieren. Abweichungen werden nicht nur erkannt, sondern über standardisierte Begründungscodes transparent dokumentiert und in ein dreistufiges Klassifikationssystem überführt: vollständig vergleichbar (C1), eingeschränkt vergleichbar (C2) und nicht vergleichbar (C3).

Die dritte Forschungsfrage untersuchte das Ausmaß, in dem diese explizite Prüfung die Ergebnisse einer Cross-City-Analyse gegenüber einer naiven Baseline-Annahme verändert. Die durchgeführte Fallstudie mit Daten aus Berlin, Paris und Barcelona lieferte hierzu ein eindeutiges Ergebnis: Während die Baseline-Annahme alle drei untersuchten Indikatoren (CO_2-Emissionen, Verkehrstote, Nettokaltmiete) als unmittelbar vergleichbar einstufte, reduzierte der entwickelte Rahmen die Anzahl uneingeschränkt zulässiger Vergleiche von drei auf null. Ein KPI wurde als eingeschränkt vergleichbar (C2), zwei aufgrund fundamentaler semantischer und methodischer Inkonsistenzen als nicht vergleichbar (C3) klassifiziert.

Zusammenfassend belegt die Arbeit, dass die methodische Belastbarkeit von Stadtvergleichen explizit geprüft und dokumentiert werden muss. Der entwickelte Vergleichbarkeitsrahmen schließt eine Lücke in bestehenden Normen und Benchmarking-Systemen, indem er Transparenz und Reproduzierbarkeit signifikant erhöht und methodisch unzulässige Rangreihen verhindert. Zukünftige Weiterentwicklungen sollten sich auf die Automatisierung der Metadatenprüfung fokussieren, um die Skalierbarkeit des Frameworks bei großen Datensätzen sicherzustellen.

Literaturverzeichnis

Albertoni, R., & Isaac, A. (2016). *Data quality vocabulary (DQV)* (W3C Working Group Note). World Wide Web Consortium. https://www.w3.org/TR/vocab-dqv/

Ajuntament de Barcelona. (o. J.). *Nombre de víctimes de sinistres de trànsit*. Portal de Dades. https://portaldades.ajuntament.barcelona.cat/

Ajuntament de Barcelona. (o. J.). *Preu mitjà per superfície del lloguer*. Portal de Dades. https://portaldades.ajuntament.barcelona.cat/

Amt für Statistik Berlin-Brandenburg. (2022a). *Energie- und CO_2-Bilanz in Berlin 2022 (E IV 4 - j/22)*. https://download.statistik-berlin-brandenburg.de/

Amt für Statistik Berlin-Brandenburg. (2022b). *Straßenverkehrsunfälle in Berlin 2022 (H I 2 - j/22)*. https://download.statistik-berlin-brandenburg.de/

Dijkstra, L., & Poelman, H. (2014). *A harmonised definition of cities and rural areas: The new degree of urbanisation* (Regional Working Paper WP 01/2014). European Commission, Directorate-General for Regional and Urban Policy.

Dijkstra, L., Poelman, H., & Veneri, P. (2019). *The EU OECD definition of a functional urban area* (OECD Regional Development Working Papers No. 2019/11). OECD Publishing. https://doi.org/10.1787/d58cb34d-en

ETSI. (2021). *Context information management (CIM), NGSI-LD API (ETSI GS CIM 009 V1.5.1)*. European Telecommunications Standards Institute.

European Union, FAO, UN-Habitat, OECD, & The World Bank. (2021). *Applying the degree of urbanisation: A methodological manual to define cities, towns and rural areas for international comparisons* (2021 ed.). Publications Office of the European Union. https://doi.org/10.2785/706535

Eurostat. (2004). *Urban audit methodological handbook* (2004 ed.). Office for Official Publications of the European Communities.

Eurostat. (2017). *Methodological manual on city statistics* (2017 ed.). Publications Office of the European Union. https://doi.org/10.2785/708009

Eurostat. (2021). *European statistical system handbook for quality and metadata reports* (2021 ed.). Publications Office of the European Union.

Huovila, A., Airaksinen, M., Pinto-Seppä, I., Piira, K., Bosch, P., Penttinen, T., Neumann, H.-M., & Kontinakis, N. (2017). CITYkeys smart city performance measurement system. *International Journal for Housing Science and Its Applications, 41*(2), 113–125.

Huovila, A., Bosch, P., & Airaksinen, M. (2019). Comparative analysis of standardized indicators for smart sustainable cities. *Cities, 89*, 141–153. https://doi.org/10.1016/j.cities.2019.01.041

ISO. (2018). *ISO 37120:2018. Sustainable cities and communities — Indicators for city services and quality of life*. International Organization for Standardization.

ISO. (2019a). *ISO 37122:2019. Sustainable cities and communities — Indicators for smart cities*. International Organization for Standardization.

ISO. (2019b). *ISO 37123:2019. Sustainable cities and communities — Indicators for resilient cities*. International Organization for Standardization.

Land Berlin. (2023). *Berliner Mietspiegel 2023 (Stichtag 01.09.2022)*. Senatsverwaltung für Stadtentwicklung, Bauen und Wohnen. https://www.berlin.de/

Lebo, T., Sahoo, S., & McGuinness, D. (Hrsg.). (2013). *PROV-O: The PROV ontology* (W3C Recommendation). World Wide Web Consortium. https://www.w3.org/TR/prov-o/

Lóscio, B. F., Burle, C., & Calegari, N. (2017). *Data on the web best practices* (W3C Recommendation). World Wide Web Consortium. https://www.w3.org/TR/dwbp/

Maali, F., & Erickson, J. (2014). *Data catalog vocabulary (DCAT)* (W3C Recommendation). World Wide Web Consortium. https://www.w3.org/TR/vocab-dcat/

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research, 3*(1), 111–121.

Observatoire des loyers de l'agglomération parisienne (OLAP). (2022). *Rapport Paris 2022*. https://www.observatoire-des-loyers.fr/

OECD & European Commission. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing. https://doi.org/10.1787/9789264043459-en

OECD & European Commission. (2020). *Cities in the world: A new perspective on urbanisation*

Open Geospatial Consortium. (2021). *OGC SensorThings API Part 1: Sensing Version 1.1* (Implementation Standard OGC 18-088). https://docs.ogc.org/is/18-088/18-088.html

Paris Open Data. (2024). *Accidentologie: base victimes*. Ville de Paris. https://opendata.paris.fr/

Paris Open Data. (2026). *Inventaire des émissions de gaz à effet de serre du territoire parisien*. Ville de Paris. https://opendata.paris.fr/

Serrano, M., Griffor, E., Wollman, D., Dunaway, M., Burns, M., Rhee, S., & Greer, C. (2022). *Smart cities and communities: A key performance indicators framework* (NIST Special Publication 1900-206). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1900-206

Sterner, P., Pargent, F., Deffner, D., & Goretzko, D. (2024). *A causal framework for the comparability of latent variables*. Manuscript submitted for publication.

UN-Habitat. (2004). *Urban indicators guidelines: Monitoring the Habitat Agenda and the Millennium Development Goals*. United Nations Human Settlements Programme.

UN-Habitat. (2025). *SDG indicator metadata: Indicator 11.3.1: Ratio of land consumption rate to population growth rate*. United Nations Human Settlements Programme.

Van den Brink, L., Barnaghi, P., Tandy, J., Atemezing, G., Atkinson, R., Cochrane, B., Fathy, Y., García-Castro, R., Haller, A., Harth, A., Janowicz, K., Kolozali, Ş., van Leeuwen, B., Lefrançois, M., Lieberman, J., Perego, A., Le-Phuoc, D., Roberts, B., Taylor, K., & Troncy, R. (2019). Best practices for publishing, retrieving, and using spatial data on the web. *Semantic Web, 10*(6), 951–980. https://doi.org/90336 10.3233/SW-1

Ville de Paris. (2025). *Les émissions de gaz à effet de serre de Paris*. https://www.paris.fr/

World Wide Web Consortium. (2017). *Data on the web best practices*. https://www.w3.org/TR/dwbp/