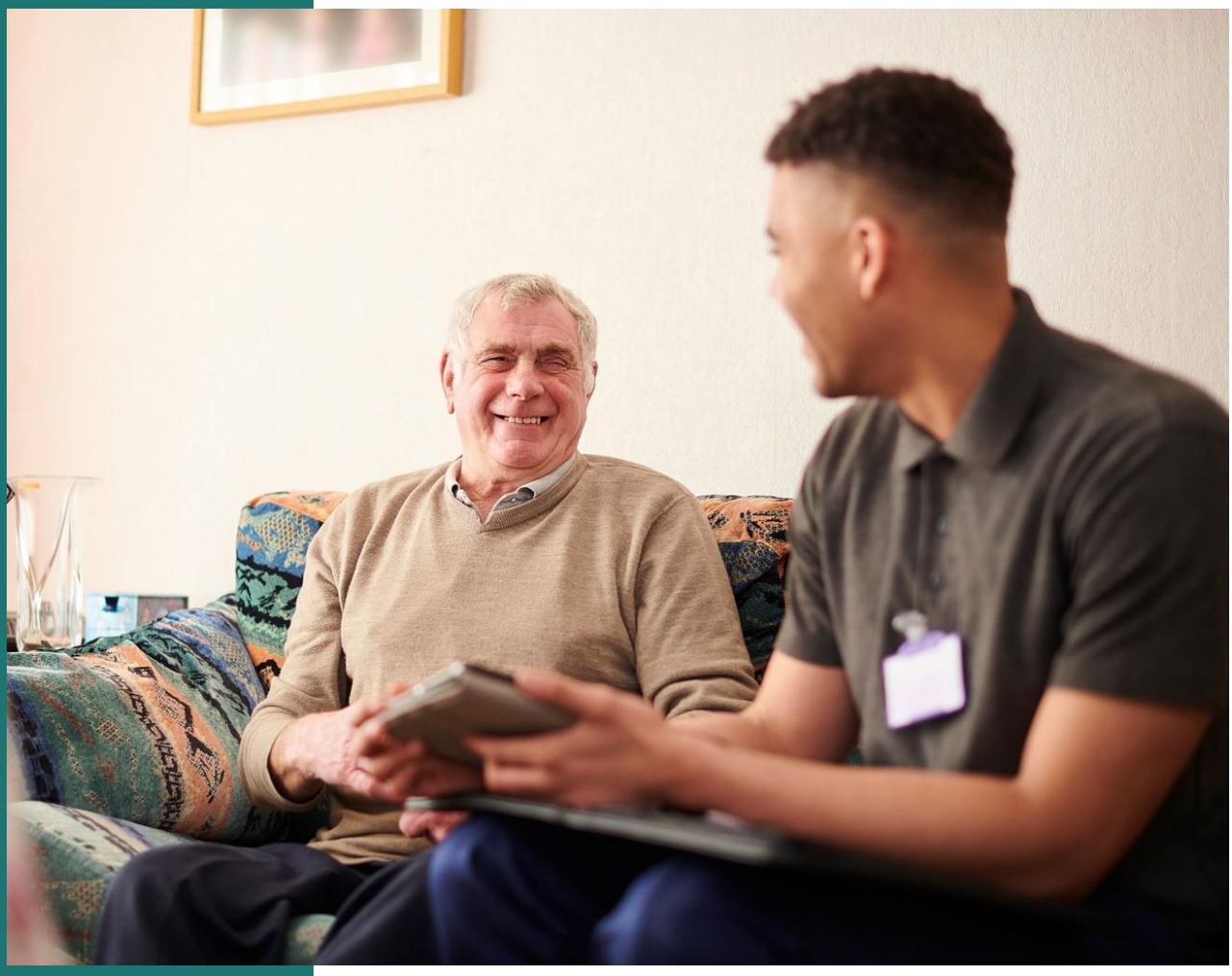
BENEFITS

Deliverable D1.1

Meta-Analysis of Existing Indicators



Funded by the European Union

Project Information

Project Acronym	BENEFITS
Project Name	Building Economic, Needs-Based and Environmental Evaluation Frameworks for Inclusive Transformation of Social Services in Europe
Call	CL2-2024-TRANSFORMATIONS-01
Grant Agreement No.	101179032
Type	HORIZON
Project Starting Date	January 2025
Project Duration	01/01/2025-31/12/2027
Coordinator	KMOP, Greece

Deliverable Information

Deliverable Name	D1.1 Meta-Analysis of Existing Indicators
Due Date	31/11/2025
Delivery Date	31/12/2025
Type	R – Document, Report
Dissemination Level	PU (Public)
Main Author(s)	Nikolaos Tzivanakis (UCL), George Melios (ETHOS), Henrietta Moore (UCL)
Contributors	Nikos Avgeris (ETHOS), Eleni Vossou (ETHOS), George Strofyllas (ETHOS), Nil Sari Aslam (UCL)
Document Reviewers	Pinar Cakiroglu (HEADWAY)
Security Reviewer	Panagiotis Katrakazas (KMOP), Eleni Plemmenou (KMOP)

Version History

Version	Issue Date	Author(s)	Content and Changes
0.1	30/6/2025	Nikolaos Tzivanakis (UCL)	ToC & Content Collection
0.2	28/11/2025	Nikolaos Tzivanakis (UCL)	First Complete Draft
0.3	9/12/2025	Pinar Cakiroglu (HEADWAY)	Internal Review
0.4	12/12/2025	Nikolaos Tzivanakis (UCL)	Second Draft – input in Sections 3.16.1, 4.5 and References, fixed format issues
0.5	16/12/2025	Panagiotis Katrakazas (KMOP)	Security Review – added missing deliverable information sections, fixed format issues
0.6	22/12/2025	Panagiotis Katrakazas (KMOP)	Final Draft

Funded by the European Union

Project number: 101179032

Acknowledgments

This report has been developed within the framework of the Horizon Europe project: Building Economic, Needs-Based and Environmental evaluation Frameworks for Inclusive Transformation of Social services in Europe (BENEFITS).

The project partners wish to thank all those who have contributed to the development of this report.

Deliverable Number and Title

D1.1 – Meta-Analysis of Existing Indicators

Authors and Contributors

Lead Author: Nikolaos Tzivanakis (UCL)
Authors: Georgios Melios (ETHOS), Henrietta Moore (UCL)
Contributors: Nikos Avgeris (ETHOS), Eleni Vossou (ETHOS), George Strofyllas (ETHOS), Nil Sari Aslam (UCL)
Reviewers: Pinar Cakiroglu (Headway)

This report should be cited as follows:
Tzivanakis, N., Melios, G., & Moore, H. (2025). Meta Analysis of Existing Indicators (Deliverable D1.1). London: BENEFITS Project (Grant Agreement No. 101179032) – Horizon Europe.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union
Project number: 101179032

Contents

Funded by the European Union
Project number: 101179032

Funded by the European Union
Project number: 101179032

List of acronyms, figures and tables

Acronyms

Acronym	Full Name
ANS	Adjusted Net Savings
ASCOT	Adult Social Care Outcomes Toolkit
BES	Benessere Equo e Sostenibile (Equitable & Sustainable Wellbeing, Italy)
BLI	Better Life Index
CBA	Cost-Benefit Analysis
CEA	Cost-Effectiveness Analysis
CIW	Canadian Index of Wellbeing
CPI	City Prosperity Index
CPIA	Country Policy & Institutional Assessment
CSV	Comma-Separated Values
CUA	Cost-Utility Analysis
DiD	Difference-in-Differences
EPI	Environmental Performance Index
ESS	European Social Survey
EU	European Union
EVS	European Values Study
FAO	Food and Agriculture Organization
GDI	Gender Development Index
GDP	Gross Domestic Product
GEP	Gross Ecosystem Product
GHG	Greenhouse Gas
GII	Gender Inequality Index
GNH	Gross National Happiness
GPI	Genuine Progress Indicator
HDI	Human Development Index
HPI	Human Poverty Index
IEA	International Energy Agency
IHDI	Inequality-Adjusted Human Development Index
IIAG	Ibrahim Index of African Governance
ILO	International Labour Organization
ISEW	Index of Sustainable Economic Welfare
IWI	Inclusive Wealth Index
JMP	Joint Monitoring Programme (WHO/UNICEF)
LSF	Living Standards Framework
MERGE	EU Horizon Project #101132524
MPI	Multidimensional Poverty Index
MSC	Most Significant Change
NGO	Non-Governmental Organisation
NICE	National Institute for Health and Care Excellence
NO_2	Nitrogen Dioxide
NUTS-2	Nomenclature of Territorial Units for Statistics (Level 2)

Funded by the European Union

Project number: 101179032

OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Student Assessment
PM2.5	Particulate Matter ≤2.5 Micrometers
PM10	Particulate Matter ≤10 Micrometers
PPI	Positive Peace Index
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY	Quality-Adjusted Life Year
R&D	Research and Development
RCT	Randomised Controlled Trial
RDD	Regression Discontinuity Design
SDG	Sustainable Development Goal
SDGI	Sustainable Development Goals Index
SO_2	Sulfur Dioxide
SPES	EU Horizon Project (Sustainable Human Development Framework)
SROI	Social Return on Investment
SWI	Sustainable Wellbeing Index
TIMSS	Trends in International Mathematics and Science Study
TOMs	National Themes, Outcomes and Measures
TPI	Thriving Places Index
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNFCCC	United Nations Framework Convention on Climate Change
UNICEF	United Nations Children's Fund
WEF	World Economic Forum
WGI	Worldwide Governance Indicators
WHO	World Health Organization
WISE	EU Horizon Project #101095219
WVS	World Values Survey

List of Figures

List of Tables

Funded by the European Union
Project number: 101179032

EXECUTIVE SUMMARY

This meta-analysis systematically examines 66 Beyond-GDP indicators developed between 1972 and 2023 to assess their coverage of social service domains. Despite five decades of innovation in wellbeing measurement, the analysis reveals systematic blind spots in domains central to social service provision, including care services, service integration, prevention, and user experience quality.

The review employed systematic methodology building on the comprehensive database of 213 wellbeing indicators compiled by Kubiszewski et al. (2025). Six sequential criteria addressing temporal relevance, implementation scale, documentation accessibility, multi-dimensional scope, social content, and service domain relevance, followed by recency verification excluding frameworks discontinued more than seven years prior, reduced this universe to 66 frameworks. The selected indicators span institutional diversity including international organisations (30%), national governments (29%), academic institutions (23%), non-governmental organisations (14%), and businesses (5%). Geographic coverage ranges from 35 global frameworks to 25 national and 6 regional initiatives. Methodological types include composite indices (52%), dashboards (33%), survey-based assessments (11%), and adjusted GDP measures (5%).

Analysis combined systematic domain mapping against a 22-domain taxonomy, methodological characteristic extraction, and co-occurrence analysis identifying which domains appear together across indicators, enabling both quantitative coverage assessment and qualitative evaluation of measurement approaches.

Domain coverage reveals pronounced concentration alongside systematic neglect. Health appears in 92% of indicators, knowledge and skills in 89%, and material wellbeing in 85%, reflecting robust measurement infrastructure and widespread recognition of their centrality to human development. Work and job quality features in 73%, environmental sustainability in 62%, economic security in 70%, housing in 58%, and governance in 47%.

Most significantly for social services, care services appear in only 20% of indicators despite their essential role supporting vulnerable populations. Service integration features in fewer than 11% of frameworks despite evidence that fragmentation harms users with complex needs. Prevention and early intervention prove virtually invisible at approximately 6% coverage. Subjective wellbeing appears in 33% of indicators, whilst culture and leisure receive minimal attention at 23% and 27% respectively.

Methodological analysis reveals additional constraints. Whilst 56% of indicators enable basic gender disaggregation, only 12% support intersectional analysis combining

Funded by the European Union

Project number: 101179032

multiple characteristics. Publication lags averaging 18 months limit policy responsiveness compared with GDP quarterly accounts. Composite indices obscure component performance critical for targeted improvement. Administrative data appearing in 89% provides comprehensive coverage but proves weak for quality dimensions requiring survey-based assessment.

Co-occurrence analysis reveals four robust empirical clusters validating a pillar structure grounded in observed practice. Economic foundations demonstrate 57 to 67% co-occurrence spanning material wellbeing, work quality, economic security, and housing. Social wellbeing shows 47 to 61% co-occurrence integrating health, knowledge, social connections, quality of life, inclusion, and equity. Environmental sustainability exhibits 31 to 36% co-occurrence across air quality, water, climate, ecosystems, and energy. Governance demonstrates 33 to 39% co-occurrence encompassing institutional quality, safety, transparency, and participation.

Critically, care services, subjective wellbeing, culture, and leisure show weak integration at 15 to 27% co-occurrence, indicating systematic neglect. The proposed framework addresses this through explicit acknowledgement of undervalued dimensions alongside empirically validated pillars, spanning access and adequacy, outcomes and experience, quality and governance, and sustainability dimensions.

This analysis provides the first systematic assessment of how Beyond-GDP frameworks address social services, identifies systematic measurement gaps through quantified analysis, and proposes empirically grounded architecture addressing identified deficiencies. The findings establish foundation for developing comprehensive service evaluation frameworks that capture dimensions critical to wellbeing but currently underrepresented in mainstream progress measurement.

Funded by the European Union

Project number: 101179032

1 INTRODUCTION

1.1 The Beyond-GDP Imperative

For more than half a century, Gross Domestic Product has dominated policy discourse as the primary metric of national progress. GDP's conceptual clarity, regular availability, and international comparability have secured its position as the headline indicator shaping public debate, political accountability, and policy priorities. Yet GDP's limitations as a comprehensive progress measure have been recognised since its inception. Simon Kuznets, architect of national income accounting, cautioned in 1934 that "the welfare of a nation can scarcely be inferred from a measure of national income" (Kuznets, 1934, p. 7). Early critiques of GDP's limitations also emerged from Nordhaus and Tobin (1972), who questioned whether economic growth remained beneficial beyond a certain threshold. GDP captures market transactions but remains blind to distributional equity, environmental sustainability, unpaid care work, health outcomes, educational attainment, social cohesion, and subjective wellbeing.

These conceptual limitations translate into practical policy distortions. Economic growth receives primacy in resource allocation decisions regardless of how gains are distributed or what costs are imposed on environmental systems or social relationships. Public services generating substantial wellbeing improvements, but modest market transactions appear peripheral to progress when GDP provides the dominant lens. Care services supporting children, persons with disabilities, and frail elderly individuals constitute major public investments directly targeting wellbeing enhancement, yet their contributions prove largely invisible in GDP-focused frameworks that measure inputs through public expenditure whilst ignoring outcomes in terms of lives improved, independence sustained, or family strain reduced.

Recognition of these limitations has generated sustained efforts to develop alternative progress measures capturing dimensions GDP neglects. The Beyond-GDP movement emerged in the 1970s through quality-of-life indices, including pioneering efforts to develop adjusted GDP measures such as the Genuine Progress Indicator (Cobb, Halstead, & Rowe, 1995) and the Index of Sustainable Economic Welfare (Stockhammer, Hochreiter, Obermayr, & Steiner, 1997), gained institutional legitimacy through the 1990 Human Development Index, and achieved political prominence following the 2008 financial crisis and the influential 2009 Stiglitz-Sen-Fitoussi Commission report (Stiglitz et al., 2009). The Sustainable Development Goals adopted in 2015 provide a global framework explicitly balancing economic, social, and environmental dimensions. National wellbeing dashboards have proliferated across OECD members since 2010, with 18 countries developing bespoke frameworks tailored to domestic policy contexts. The European Union's Beyond GDP initiative, launched in

Funded by the European Union
Project number: 101179032

2007 and reinforced through the European Pillar of Social Rights, positions comprehensive wellbeing measurement as essential infrastructure for evidence-informed social policy.

1.2 The Social Services Measurement Gap

Despite this proliferation of Beyond-GDP frameworks, a critical blind spot persists (Hoekstra, 2019; Terzi, 2021). Social services that constitute major welfare state investments and directly target wellbeing improvement receive inadequate attention in alternative progress measures. Healthcare and education feature prominently given robust international statistical infrastructure and widespread recognition of their importance. Material living standards prove comprehensively measured through income, consumption, and wealth indicators. Environmental sustainability has achieved growing integration particularly regarding climate change and air quality (Kubiszewski et al., 2025).

In contrast, care services supporting children, persons with disabilities, and elderly individuals remain systematically underrepresented. The International Labour Organization estimates that 2.1 billion people globally require care services, yet care work remains largely invisible in national progress metrics (ILO, 2018, 2022, 2024b). This invisibility perpetuates gender inequalities, as women perform over three-quarters of unpaid care work globally (UN Women, 2016a). Service integration and coordination addressing the needs of individuals requiring support across health, social care, housing, and employment domains prove virtually invisible. Prevention and early intervention that could avert crises through proactive strategies receive negligible measurement attention compared to reactive crisis responses. Subjective experience of service quality including dignity, respect, and autonomy in delivery contexts rarely features despite its importance to users. These gaps reflect structural biases including the devaluation of feminised reproductive labour, productivist frameworks privileging market production over care provisioning, and measurement conventions evolved for macroeconomic assessment rather than service evaluation purposes. These gaps reflect structural biases including the devaluation of feminised reproductive labour (Klasen, 2007; UN Women, 2016a), productivist frameworks privileging market production over care provisioning, and measurement conventions evolved for macroeconomic assessment rather than service evaluation purposes (Beck, 1999; Johnston, 1985).

This systematic neglect carries substantial consequences. Resource allocation decisions informed by incomplete wellbeing measures perpetuate underinvestment in unmeasured dimensions. Performance management systems optimising measured indicators whilst ignoring unmeasured dimensions create perverse incentives favouring

Funded by the European Union
Project number: 101179032

easily quantified outputs over complex qualitative outcomes. Policy evaluation assessing economic impacts whilst neglecting service experiences provides distorted evidence regarding intervention effectiveness. Democratic accountability suffers when citizens lack transparent information regarding service quality, accessibility, and outcomes enabling informed judgment of collective resource stewardship.

1.3 The BENEFITS Project Context

This meta-analysis forms part of the BENEFITS project, a European research initiative developing comprehensive frameworks for social service evaluation integrating wellbeing measurement, economic assessment, and equity analysis. The project addresses the measurement gap documented above through systematic evidence synthesis, methodological development, and practical application across European social service contexts spanning healthcare, education, social protection, housing, and care services.

The research questions guiding this meta-analysis are threefold. First, how comprehensively do existing Beyond-GDP indicators address social service domains, and what systematic gaps exist in coverage? Second, what methodological approaches characterise Beyond-GDP measurement, and what are their implications for social service applicability regarding data requirements, update frequency, equity integration, and user accessibility? Third, what conceptual frameworks and pillar structures emerge empirically from analysis of which domains co-occur across indicators, and how should these inform BENEFITS architecture?

1.4 Research Approach and Scope

This analysis employs systematic review methodology adapted from health sciences to policy and indicator literature. From an initial universe documented in academic literature, the review builds on the comprehensive database compiled by Kubiszewski et al. (2025), which systematically classified 213 wellbeing indicators through semantic analysis. Systematic application of six criteria addressing temporal relevance, implementation scale, documentation accessibility, multi-dimensional scope, social content prominence, and service domain relevance identified 66 frameworks meeting inclusion requirements. An additional criterion excluded frameworks discontinued more than seven years before 2024 to ensure analysis reflects contemporary measurement practice rather than historical approaches no longer considered viable.

The 66 selected indicators span five decades from 1972 to 2023, providing historical depth capturing both foundational frameworks establishing the Beyond-GDP movement and contemporary initiatives reflecting current measurement practice. Institutional diversity encompasses international organisations, national governments,

Funded by the European Union
Project number: 101179032

academic institutions, non-governmental organisations, and business entities. Geographic coverage ranges from global frameworks facilitating international comparison to national dashboards supporting domestic policy and regional initiatives addressing subnational contexts. Methodological heterogeneity includes composite indices aggregating components into headline scores, dashboards presenting disaggregated indicator sets, survey-based subjective assessments, and adjusted GDP measures modifying national accounts.

Analysis proceeded through mixed methods combining systematic categorisation with interpretive synthesis. Each indicator was mapped against a 22-domain taxonomy assessing coverage breadth and classified according to conceptual dimensions addressing wellbeing, sustainability, and inclusion. Methodological characteristics including data sources, update frequency, aggregation approaches, and equity integration were systematically extracted. Domain co-occurrence analysis identifying which dimensions appear together revealed empirical clustering patterns informing framework architecture. Synthesis integrated quantitative aggregation of coverage patterns with qualitative assessment of measurement approaches and their implications for social service evaluation.

1.5 Principal Findings and Framework Contributions

The analysis reveals pronounced coverage concentration alongside systematic neglect. Health (92%), knowledge and skills (89%), and material wellbeing (85%) dominate the indicator landscape, reflecting robust measurement infrastructure and widespread recognition of their centrality to human development. However, domains critical to social services remain severely underrepresented: care services appear in only 20% of indicators, service integration in 11%, and prevention in 6%. Whilst 56% of frameworks enable basic gender disaggregation, only 12% support intersectional analysis examining compounding disadvantages. Methodologically, composite indices (52%) facilitate communication but obscure component performance, whilst publication lags averaging 18 months limit responsiveness compared with GDP's six-week release cycle.

Domain co-occurrence analysis identifies four empirical clusters validating a pillar structure: economic foundations (material wellbeing, employment, economic security, housing), social wellbeing (health, knowledge, social connections, inclusion, equity), environmental sustainability (air, water, climate, ecosystems, energy), and governance and institutions (institutional quality, safety, transparency, participation). Care services, subjective wellbeing, culture, and leisure demonstrate weak integration with these clusters (15-27% co-occurrence), confirming systematic neglect requiring deliberate correction.

Funded by the European Union
Project number: 101179032

The BENEFITS framework addresses these gaps through empirically grounded architecture spanning access and adequacy, outcomes and experience, quality and governance, and sustainability. Social Value Analysis integrates within methodological pluralism combining economic evaluation, participatory assessment, and equity-focused analysis. The framework has implications for measurement infrastructure, service system design, and evaluation practice.

This analysis contributes to the Beyond-GDP movement by providing the first systematic assessment of how alternative progress measures address social services, identifying systematic measurement gaps through quantified analysis, and proposing empirically grounded framework architecture addressing identified deficiencies. For the BENEFITS project specifically, it provides essential foundation establishing measurement landscape context, justifying proposed indicator architecture through gap analysis, and informing operational development through synthesis of international experience. For European policymakers and practitioners, it offers evidence base for developing comprehensive service evaluation frameworks supporting evidence-informed improvement, transparent accountability, and equitable resource allocation.

Funded by the European Union

Project number: 101179032

2 METHODOLOGY

This meta-analysis employed a systematic review methodology to identify, assess, and synthesise Beyond-GDP indicators relevant to social service evaluation within the BENEFITS project framework. The research process followed PRISMA guidelines for systematic reviews, adapted for policy and indicator literature rather than clinical studies (see Annex I). The methodology comprised four sequential phases: comprehensive identification of candidate indicators, systematic screening against eligibility criteria, detailed assessment of relevance to social service contexts, and final inclusion based on transparent scoring thresholds.

The temporal scope of the review extended from the late 1970s to November 2024, capturing both foundational frameworks that established the Beyond-GDP movement and contemporary initiatives that reflect current measurement practice. This historical depth enabled analysis of how social services have been treated across different generations of indicator development, from early quality-of-life indices through the post-2015 SDG era.

2.1 Identification and Search Strategy

The identification phase involved comprehensive searches across multiple databases and institutional repositories between April and August 2025. Primary databases consulted included Google Scholar, Web of Science, OECD iLibrary, the UN Statistics Division, Eurostat, and institutional repositories maintained by the World Bank, International Monetary Fund, and regional development banks (Benczúr et al., 2025; Hoekstra, 2019; Jansen et al., 2024). Search terms combined Beyond-GDP terminology with wellbeing and sustainability concepts, including variations of "wellbeing indicator," "prosperity indicator," "happiness indicator," "quality of life indicator," "social progress indicator," "sustainable development indicator," and "beyond GDP indicator."

To ensure comprehensive coverage, targeted searches extended to recent outputs from EU Horizon research projects (WISE Horizon #101095219: Jansen et al., 2024; MERGE Horizon #101132524: Costanza et al., 2024; SPES: Biggeri et al., 2023), the European Commission's Beyond-GDP initiative (European Commission, 2009, 2023), national statistical office publications from OECD and EU member states, and grey literature from major think tanks and non-governmental organisations including the New Economics Foundation (NEF, 2006), Legatum Institute (2007), and Social Progress Imperative (2013).

Funded by the European Union
Project number: 101179032

The search strategy employed snowball sampling from reference lists of major systematic reviews in the field (Hoekstra, 2019; Costanza et al., 2014; Fleurbaey & Blanchet, 2013; Lawn, 2005). Most significantly, the review built upon the comprehensive semantic synthesis conducted by Kubiszewski et al. (2025), which provided a systematically classified database of 213 wellbeing indicators. This database served as both a primary source and a cross-reference to verify that independent searches had captured all major frameworks documented in academic and policy literature.

2.2 Selection Criteria and Application

To focus the analysis on indicators most relevant to social service evaluation within the BENEFITS framework, a multi-stage screening process employed six systematic criteria applied sequentially, followed by temporal recency verification and a final relevance threshold. This approach balanced comprehensive coverage of measurement diversity with focused attention on frameworks applicable to social service contexts.

Binary exclusion filters (Criteria 1-4). The first four criteria served as binary filters, establishing minimum requirements for temporal relevance, geographic scale, documentation accessibility, and multidimensional scope. These filters ensured that selected indicators reflected contemporary or historically influential measurement practice, demonstrated transferability beyond single local contexts, provided sufficient methodological transparency for assessment, and addressed multiple wellbeing dimensions rather than single-issue metrics.

Criterion 1: Temporal Relevance. This filter retained indicators that were either active between 2000 and 2024 or were foundational pre-2000 frameworks that significantly influenced subsequent measurement practice. The contemporary window (2000-2024) ensured relevance to current policy debates whilst the foundational exception preserved historically important frameworks including the Human Development Index (1990) and Gross National Happiness (1972) that shaped the Beyond-GDP movement. Application of this criterion to the initial database of 213 indicators excluded 26 measures, primarily historical quality-of-life indices from the 1970s and 1980s that had been discontinued without lasting methodological influence. This yielded 187 potentially relevant indicators.

Criterion 2: Scale of Implementation. This filter required sufficient geographic scope to demonstrate transferability beyond single local contexts and enable cross-jurisdictional learning. Three alternative pathways enabled inclusion: deployment in at least three countries, indicating methodology proven applicable across diverse national contexts; systematic application across EU or OECD member states, demonstrating regional relevance and coordination among multiple statistical systems;

Funded by the European Union
Project number: 101179032

or recognition as an international reference framework by major multilateral organisations including UN agencies, World Bank, OECD, or EU institutions, indicating international legitimacy and influence on measurement discourse even when implementation remained limited. This third pathway enabled inclusion of pioneering national frameworks such as Bhutan's Gross National Happiness Index or New Zealand's Living Standards Framework that, whilst developed for single countries, achieved international recognition and influenced framework development elsewhere.

Of the 187 indicators meeting temporal criteria, 63 were excluded under Criterion 2. Excluded indicators comprised single-city measures such as the Hamra Prosperity Index and Charlotte/Mecklenburg Quality of Life Explorer lacking evidence of replication or influence beyond original contexts, single-institution indices including campus wellbeing indicators and corporate sustainability metrics developed for internal use, pilot projects without documented wider adoption, and highly localised community indicators developed for specific neighbourhoods lacking transferable methodology or documentation enabling replication. This reduced the pool to 124 indicators demonstrating sufficient scale, multi-jurisdictional application, or international influence.

Criterion 3: Public Documentation. This filter required publicly accessible methodology documented in English, enabling independent assessment and potential replication. Whilst this language restriction excluded potentially innovative non-Anglophone frameworks, it proved necessary given resource constraints and the research team's capacity. Six of the 124 remaining indicators were excluded due to proprietary methodologies behind paywalls, references in literature without accompanying technical documentation, or frameworks announced but not yet operationalised with published results. This yielded 118 indicators with adequate documentation.

Criterion 4: Multi-Dimensional Scope. This filter mandated coverage of at least three distinct measurement domains to qualify as comprehensive progress frameworks rather than single-issue metrics. The Beyond-GDP movement explicitly seeks multidimensional assessment, making breadth a defining characteristic. Application of this criterion excluded 29 indicators focused on narrow thematic areas including pure environmental metrics such as isolated air quality or water poverty indices, single demographic focuses like the Children's Climate Risk Index or Arctic Social Indicators, and narrow sectoral measures addressing only urban health or agricultural wellbeing. This left 89 indicators meeting all four binary criteria.

Relevance scoring (Criteria 5-6). The remaining 89 indicators were then scored on two relevance dimensions using four-point scales (0-3) to assess social content prominence and service domain relevance. This approach enabled nuanced

Funded by the European Union
Project number: 101179032

differentiation among frameworks meeting basic inclusion criteria, identifying those with greatest applicability to social service evaluation.

Criterion 5: Social Content Prominence. This dimension assessed the extent to which social welfare, equity, or inclusion featured in the framework's conceptual architecture and component structure. Scoring guidelines defined: 3 = social welfare central to framework with multiple integrated social domains (e.g., SDGs, Human Development Index, Social Progress Index); 2 = moderate social content through several social components alongside economic or environmental dimensions; 1 = minimal social variables, typically one or two components within predominantly economic or environmental frameworks; 0 = purely economic or environmental indicators with no social dimension. This distribution yielded 31 indicators scoring 3, 38 scoring 2, 15 scoring 1, and 5 scoring 0.

Criterion 6: Service Domain Relevance. This dimension evaluated specific applicability to social service evaluation contexts. Scoring guidelines defined: 3 = explicit measurement of service delivery, access, quality, or user outcomes; 2 = partial relevance through inclusion of health, education, housing, or social protection domains; 1 = indirect connections through economic determinants or governance dimensions affecting but not directly measuring service provision; 0 = no discernible service connection. This assessment produced 18 indicators scoring 3, 42 scoring 2, 24 scoring 1, and 5 scoring 0.

Application of the combined relevance requirement proved straightforward given the scoring distributions. Final inclusion required a combined score of at least 3 across Criteria 5 and 6, ensuring that selected indicators demonstrated meaningful social content and relevance to service contexts. This threshold could be achieved through various combinations: frameworks with strong social content but moderate service relevance (3+0, 3+1, 2+1), frameworks with explicit service focus regardless of overall social breadth (0+3, 1+3), or balanced frameworks addressing both dimensions moderately (2+2). This flexible approach avoided excluding either socially focused frameworks with indirect service connections or service-specific frameworks operating within narrow social domains.

The threshold excluded 17 indicators scoring below 3, primarily those with strong environmental or economic focus but weak social service relevance, including several pure ecological footprint measures, economic competitiveness indices, and governance indicators lacking social outcome components. The final included set comprised 72 indicators spanning diverse institutional sources, geographic scopes, and methodological approaches whilst maintaining consistent relevance to comprehensive social service evaluation.

Funded by the European Union
Project number: 101179032

Following initial screening through the six criteria, frameworks were reviewed for active status to ensure analysis reflected contemporary measurement practice rather than historical approaches no longer considered viable. Indicators discontinued more than seven years before the review date (2025) were excluded unless they represented foundational methodological contributions directly informing active successors. This additional filter removed six frameworks discontinued between 2005 and 2016: Environmental Sustainability Index (replaced by Environmental Performance Index in 2006), Green GDP China (pilot discontinued 2007), National Accounts of Wellbeing (one-off 2009), Where-to-be-born Index (one-off 2013), Measures of Australia's Progress (discontinued 2014), and Sustainable Society Index (discontinued 2016).

The Human Poverty Index (discontinued 2010) was retained as direct methodological predecessor to the active Multidimensional Poverty Index, illustrating the evolution of multidimensional poverty measurement within UNDP's framework development. This decision reflected the principle that discontinued frameworks merit inclusion when they demonstrate clear methodological lineage to active successors, enabling analysis of measurement practice evolution. The recency verification reduced the pool from 72 to 66 indicators.

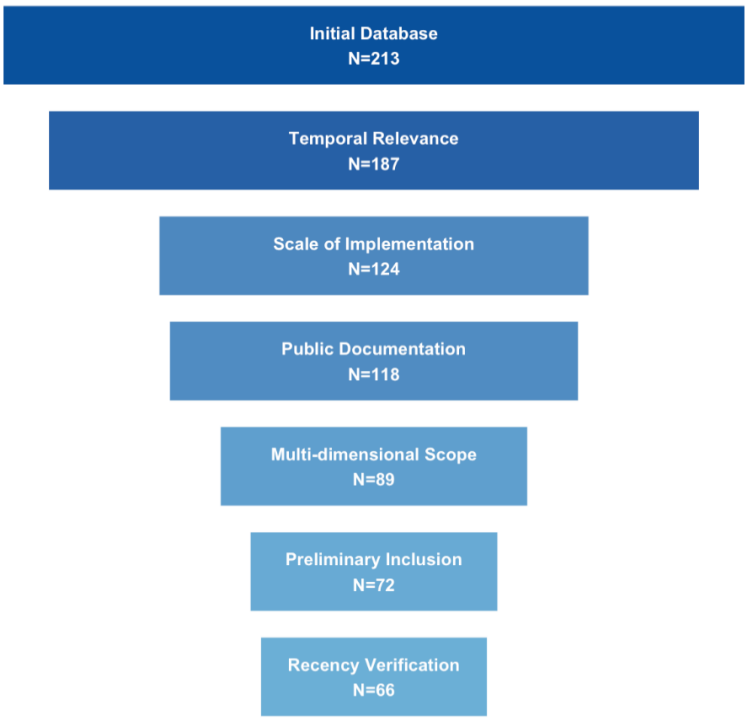


Figure 1: Selection Criteria and Application Results

Funded by the European Union
Project number: 101179032

2.3 Analytical Framework

The analytical framework employed mixed methods combining systematic categorisation with interpretive synthesis. Two complementary classification systems enabled structured comparison: a 22-domain taxonomy for assessing coverage breadth and a three-dimensional conceptual model for evaluating theoretical orientation.

The domain classification system adapted the MERGE project taxonomy (Rum et al., 2024), which identified approximately 20 distinct domains through unsupervised semantic clustering of indicator components across international frameworks. Three modifications were made to align this taxonomy with BENEFITS objectives. First, domains with minimal social service relevance were excluded, specifically Mineral Resources and pure Financial Capital. Second, service-specific domains absent from the original MERGE classification were added: Social Development, Social Inclusion, Social Equity, and Environmental Sustainability. Third, conceptually overlapping domains were consolidated: Knowledge Capital was integrated into Knowledge and Skills, while Trust and Social Connections were distinguished from general Social Capital to capture relationship dimensions specifically. This process yielded 22 analytical domains against which each indicator's coverage was systematically mapped.

Domain assignment proceeded through expert review of indicator documentation. For frameworks providing complete component lists, domains were coded based on explicit variable inclusion. Where full component inventories were unavailable, domain coverage was inferred from technical reports, academic publications describing the framework's construction, and institutional documentation detailing methodological approaches. Approximately 30% of indicators required this inferential approach due to incomplete public documentation.

The conceptual dimensions analysis followed the tripartite framework established by the WISE project (Jansen et al., 2023) and validated across multiple Beyond-GDP reviews. This model classifies indicators according to their primary theoretical orientation across three overarching dimensions. The Wellbeing dimension captures frameworks focused on quality of life, life satisfaction, and human flourishing, identified through presence of subjective wellbeing measures, health outcomes, educational attainment, and quality-of-life components. The Sustainability dimension encompasses frameworks emphasising environmental limits, natural capital preservation, and intergenerational equity, identified through environmental indicators, natural capital accounting, and future-oriented metrics. The Inclusion dimension addresses frameworks concerned with distribution, equality, and the principle of

Funded by the European Union
Project number: 101179032

leaving no one behind, captured through inequality measures, disaggregation capabilities, and explicit equity components. These dimensions were coded as binary present/absent classifications, with additional notation indicating primary focus where a framework clearly privileged one dimension over others.

2.4 Data Extraction Protocol

A structured extraction template was developed and pilot-tested on ten indicators before full application to the 66-indicator sample. This template systematically captured metadata, conceptual architecture, methodological characteristics, and social service relevance features for each framework. Basic metadata extraction recorded the indicator's official name and common abbreviations, developing organisation and institutional type (categorised as government, intergovernmental organisation, non-governmental organisation, academic institution, or business entity), year of introduction and subsequent update history, geographic coverage and scale of implementation, and status as active or discontinued. This information established the indicator landscape's institutional and temporal structure.

Conceptual architecture extraction documented each framework's stated purpose and policy objectives as articulated by developers, theoretical foundation (classified as capabilities-based, wellbeing-focused, sustainability-oriented, rights-based, or hybrid approaches), and target audience (policymakers, public, or specific sectors). This characterisation enabled assessment of how different theoretical traditions address social services.

Methodological characteristics extraction captured construction type (composite index, dashboard, adjusted GDP measure, or survey-based assessment), number of components or constituent indicators where available, aggregation methodology for composite indices, weighting scheme (categorised as equal weighting, expert-derived, statistically-derived, or participatory), data sources (administrative data, survey data, or mixed approaches), update frequency, and publication lag time from reference period to release. These features determine an indicator's practical utility for service evaluation and policy integration.

Social service relevance extraction focused on coverage of service-relevant domains (health, education, social protection, housing, care services), measurement approach (input, output, outcome, or impact focus), disaggregation capabilities across equity dimensions (gender, age, disability status, geographic location, income quintile), presence of explicit service delivery indicators, and documented links to service funding or provision mechanisms. This extraction enabled assessment of how comprehensively existing frameworks capture social service dimensions.

Funded by the European Union
Project number: 101179032

2.5 Synthesis and Analysis Approach

Synthesis combined quantitative aggregation with qualitative pattern identification to extract lessons for the framework design. Quantitative analysis generated frequency distributions across domains, dimensions, and methodological types to identify coverage patterns and systematic gaps. Cross-tabulations examined relationships between indicator characteristics, such as how composite indices versus dashboards differ in domain coverage or how different institutional developers emphasise distinct dimensions. Coverage heat maps visualised which domains receive attention across the indicator landscape, and which remain systematically neglected. Temporal analysis tracked how domain emphasis has evolved across indicator generations from the 1990s through the 2020s.

Qualitative synthesis employed thematic analysis to identify how social services appear conceptually across different indicator traditions, whether framed as human capital investments, rights entitlements, or consumption expenditures. Gap analysis compared indicator coverage systematically against the full range of social service functions, from prevention through crisis intervention to long-term support. Critical assessment evaluated the strengths and limitations of different measurement approaches for service evaluation contexts, considering issues of attribution, aggregation, actionability, and equity integration.

2.6 Limitations and Quality Assurance

Several limitations constrain this analysis and should be acknowledged. First, restriction to English-language documentation may exclude innovative approaches from non-Anglophone contexts, particularly from the Global South where alternative conceptualisations of progress may differ substantially from OECD frameworks. Second, the heterogeneity of indicators spanning different scales, purposes, and methodological traditions complicates direct comparison and may obscure important contextual nuances. A composite index designed for international benchmarking serves different functions than a national dashboard supporting domestic policy, yet both appear in this analysis.

Third, relevance scoring on criteria 5 and 6, whilst systematic and transparent, necessarily involves subjective judgment. Different researchers might reasonably assign different scores, particularly for boundary cases between score categories. To mitigate this limitation, scoring protocols were documented and inter-rater reliability was tested, but interpretive variation remains inevitable. Fourth, the rapid evolution of the Beyond-GDP field means very recent developments post-November 2024 may not

Funded by the European Union
Project number: 101179032

be captured, and some indicators may have been modified or discontinued since documentation was accessed.

Fifth, variation in documentation quality and detail limited the ability to conduct component-level analysis uniformly across all indicators. Approximately 30% of included indicators lacked publicly available complete component lists, requiring reliance on secondary descriptions that may not fully capture methodological complexity or accurately represent all dimensions measured. This asymmetry in available information may introduce bias toward better-documented frameworks, which tend to originate from well-resourced international organisations rather than national governments or civil society organisations operating with limited dissemination budgets.

Last, application of Criterion 2 regarding scale of implementation required subjective judgment, particularly for single-country frameworks. Whilst multi-country indicators (N=35 global frameworks) unambiguously met this criterion, inclusion of national dashboards and country-specific indices necessitated judgment regarding international influence or methodological transferability. The review adopted an inclusive approach, incorporating well-documented national frameworks from OECD and EU member states (N=25) on the rationale that these demonstrate diverse approaches to operationalising wellbeing measurement in advanced statistical systems, even when direct evidence of international replication or citation remained limited. This inclusive strategy may have admitted frameworks with primarily domestic rather than international relevance, potentially limiting generalisability of findings beyond high-income country contexts. More restrictive application excluding single-country frameworks without documented international influence would have yielded a smaller but more clearly transferable indicator set. The tension between comprehensive coverage of measurement diversity and strict focus on internationally validated approaches reflects broader methodological trade-offs in systematic reviews.

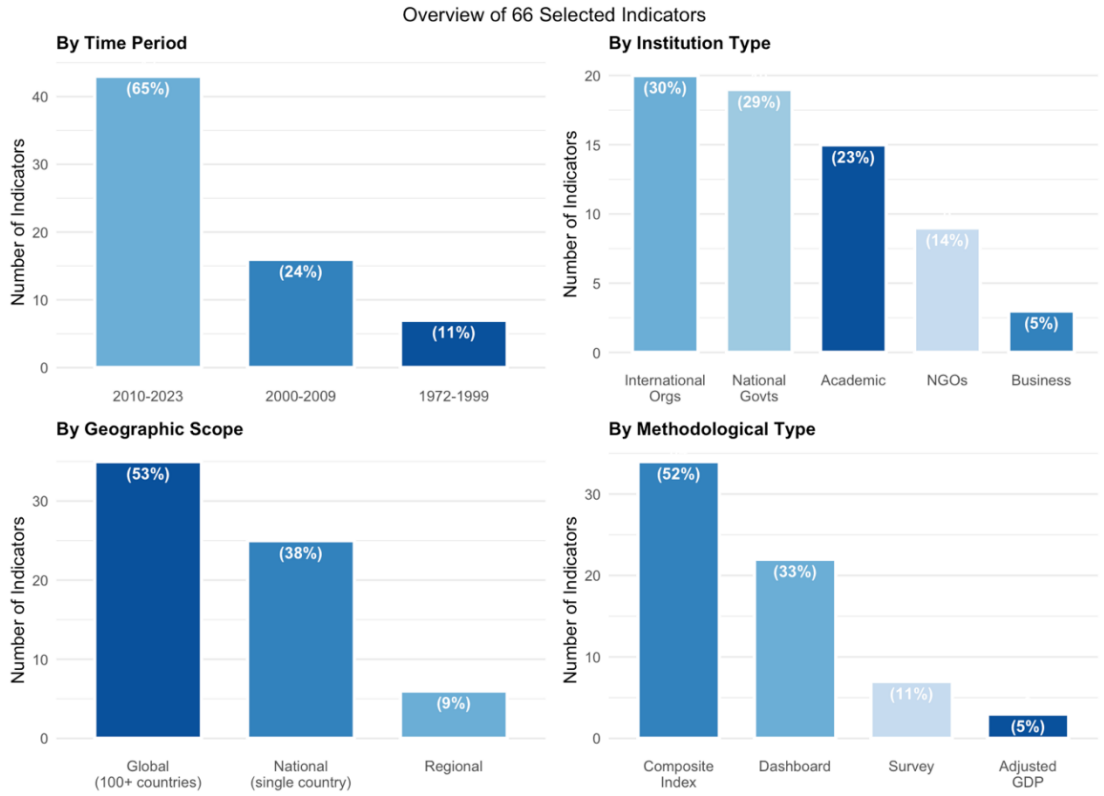
To ensure consistency and reliability, two reviewers independently assessed 20% of indicators against inclusion criteria, achieving 91% inter-rater agreement. Discrepancies were resolved through discussion and reference to original documentation. External validation cross-referenced the final included set against indicators identified in recent systematic reviews of Beyond-GDP measurement (Hoekstra, 2019; Costanza et al., 2014), confirming that no major frameworks documented in the literature had been inadvertently excluded.

Funded by the European Union

Project number: 101179032

3 FINDINGS

3.1 Overview of Selected Indicators



Note: Percentages may not sum to 100% due to rounding; some indicators classified in multiple categories

Figure 2: Overview of Selected Indicators

Systematic application of selection criteria identified 66 Beyond-GDP indicators meeting inclusion requirements for detailed analysis. These frameworks represent a diverse landscape of measurement spanning five decades of development from 1972 to 2023, providing robust foundation for understanding how alternative progress measures relate to social service evaluation contexts.

3.2 Temporal Distribution and Evolution

The temporal pattern of indicator development reveals three distinct waves of innovation corresponding to evolving policy contexts and intellectual currents shaping the Beyond-GDP movement. The first wave, extending from 1972 to 1999, produced seven foundational frameworks representing 11% of the sample. This period established core conceptual approaches that continue to influence contemporary measurement practice. Notable initiatives from this era include Bhutan's Gross

Funded by the European Union
Project number: 101179032

National Happiness framework introduced in 1972 and operationalised through comprehensive surveys from 2008, the Human Development Index launched by UNDP in 1990, the Gender Development Index established in 1995, and early adjusted GDP measures including the Index of Sustainable Economic Welfare first calculated in 1989. These pioneering efforts, whilst limited in number, demonstrated the feasibility of multidimensional progress assessment and established methodological precedents for component selection, normalisation, and aggregation. Their enduring influence justifies retention despite age, as subsequent frameworks build explicitly on conceptual foundations these initiatives established.

The second wave, spanning 2000 to 2009, marked significant expansion with 16 indicators representing 24% of the sample. This acceleration followed the Millennium Development Goals and growing recognition of GDP limitations in capturing social and environmental dimensions of development. Key frameworks from this period include the Environmental Performance Index introduced in 2006, the Legatum Prosperity Index launched in 2007, and multiple national wellbeing initiatives as countries began developing bespoke measurement systems. The temporal concentration reflects both the 2008 financial crisis, which prompted fundamental questioning of growth-focused metrics, and the influence of civil society organisations including New Economics Foundation advocating alternative measures of prosperity. Methodological innovation during this period included greater attention to subjective wellbeing and environmental sustainability, expanding Beyond-GDP scope beyond human development's traditional focus.

The third and largest wave, from 2010 to 2023, generated 43 indicators representing 65% of the sample. This proliferation reflects post-financial crisis rethinking of progress and the catalytic influence of the Stiglitz-Sen-Fitoussi Commission report in 2009, which provided intellectual legitimacy and policy momentum for Beyond-GDP measurement. The period witnessed emergence of national wellbeing frameworks, with numerous countries developing dashboards tailored to domestic policy contexts and integrated with budget processes. The 2015 adoption of the Sustainable Development Goals provided a unifying global framework, whilst simultaneously stimulating development of complementary measurement initiatives at national and subnational scales. Recent frameworks demonstrate growing sophistication in equity disaggregation, user experience assessment, and integration with policy cycles, suggesting maturation from advocacy tools toward operational performance management systems.

Funded by the European Union
Project number: 101179032

3.3 Institutional Distribution

Analysis of developing organisations reveals a balanced ecosystem of indicator production across five institutional categories, each bringing distinctive strengths and limitations to measurement practice. International organisations, including UN agencies, OECD, World Bank, and EU institutions, account for 30% (20 frameworks) and dominate global comparative measurement. These organisations leverage established statistical infrastructure, convening authority for methodological harmonisation, and political legitimacy deriving from intergovernmental mandates. UN agencies contribute the Human Development Index family, Multidimensional Poverty Index, and multiple gender-focused indices. OECD developed the Better Life Index and maintained extensive social protection statistics. World Bank produces Adjusted Net Savings and related wealth accounting frameworks. Eurostat coordinates EU regional quality of life measurement. International frameworks prioritise cross-country comparability and alignment with global policy architectures but may not adequately capture context-specific priorities or sub-national variation.

National governments, primarily from OECD member states, represent 29% (19 frameworks) and focus on wellbeing dashboards supporting domestic policy processes. This category includes pioneering national initiatives from UK, New Zealand, Netherlands, Italy, Canada, and Nordic countries that integrated Beyond-GDP measurement with budget cycles and parliamentary reporting. Government-led frameworks benefit from official statistical authority, policy relevance through alignment with decision-making structures, and sustainability through institutionalisation. However, political sensitivity may constrain measurement of controversial dimensions, and technical capacity limits implementation primarily to high-income contexts with advanced statistical systems.

Academic institutions and individual researchers contribute 23% (15 frameworks), often introducing methodological innovations subsequently adopted by official statistics. Academic contributions include ecological economics frameworks such as Genuine Progress Indicator and Inclusive Wealth Index, quality-of-life indices testing novel aggregation approaches, and conceptual frameworks influencing policy discourse even when not operationalised for regular publication. Academic freedom enables experimentation and critical perspectives, though limited resources constrain geographic coverage and update frequency compared to official statistical programs.

Non-governmental organisations and think tanks produce 14% (9 indicators), typically emphasising specific dimensions including inequality, environmental sustainability, or social progress. Organisations including New Economics Foundation, Social Progress Imperative, and various advocacy groups use measurement strategically to challenge

Funded by the European Union
Project number: 101179032

dominant policy narratives and mobilise constituencies. Civil society frameworks often prove innovative in incorporating user voice and addressing undervalued dimensions but may face sustainability challenges dependent on project funding and lack official statistical authority supporting data access and publication.

This institutional distribution suggests broad recognition across sectors of the need for alternative progress measures, though with varying motivations and methodological approaches. Successful Beyond-GDP implementation likely requires coordination across institutional types, combining international organisations' standardisation capacity, national governments' policy integration, academic institutions' methodological innovation, civil society's critical perspectives, and commercial actors' user-oriented design.

3.4 Geographic Coverage

The indicators display a pronounced hierarchy of geographic scope reflecting different measurement purposes and varying capacity for international comparison versus local policy relevance. Global frameworks designed for cross-national comparison represent 53% (35 indicators), typically covering 100 or more countries. These enable benchmarking performance across diverse contexts, facilitate identification of universal patterns and context-specific factors, and support diffusion of measurement practices through demonstration effects. Global frameworks prove particularly valuable for international development policy, where resource allocation and progress assessment require comparable metrics spanning low-income, middle-income, and high-income contexts. However, standardisation demands may obscure locally salient dimensions, and data quality variation across national statistical systems complicates interpretation when comparing countries with vastly different measurement capacity.

National frameworks tailored to country-specific circumstances comprise 38% (25 indicators), predominantly from OECD or EU member states. These frameworks integrate with domestic policy cycles, reflect national priorities and political settlements regarding valued outcomes, and exploit rich administrative and survey data available in advanced statistical systems. National dashboards prove particularly actionable for policy development, as indicator selection and disaggregation align with domestic governance structures and service delivery responsibilities. The geographic concentration in high-income contexts highlights measurement innovation clustering where statistical capacity, political interest in Beyond-GDP approaches, and fiscal resources for survey investment converge. This pattern raises questions about applicability to diverse social service contexts globally, as frameworks developed in settings with universal public services, strong welfare states, and comprehensive

Funded by the European Union
Project number: 101179032

administrative data may not adequately address challenges in contexts where service provision models, informal arrangements, and resource constraints differ substantially.

Regional frameworks focused on EU member states or subnational territories account for 9% (6 indicators). Regional initiatives including Eurostat's Quality of Life Index for NUTS-2 regions and various metropolitan area assessments enable within-country comparison whilst accounting for substantial sub-national variation in economic conditions, demographic composition, and service provision. Regional frameworks prove valuable for federal or decentralised governance systems where policy authority resides at provincial or municipal levels, enabling performance comparison across jurisdictions sharing national institutional frameworks whilst differing in local implementation.

The geographic distribution carries implications for our framework development. Whilst global frameworks provide templates for standardised measurement, European social service evaluation must draw primarily on national and regional initiatives reflecting governance contexts, service models, and data availability more comparable to EU member state circumstances. Learning from national pioneers including UK, New Zealand, and Netherlands proves particularly relevant, though adaptation to diverse European contexts requires attention to variation in welfare state architectures, statistical capacity, and policy integration mechanisms.

3.5 Methodological Types

Distribution across methodological approaches reveals enduring tension between communicative simplicity facilitating political uptake and analytical richness supporting evidence-informed policy development. Composite indices producing single aggregated scores remain dominant at 52% (34 indicators), employing various aggregation techniques including arithmetic means, geometric means, and sophisticated weighting schemes. These frameworks facilitate communication through headline numbers comparable to GDP, enable country ranking satisfying media interest and political accountability demands, and provide single summary metrics for tracking aggregate progress over time. Composite indices serve important communication and mobilisation purposes, translating complex multidimensional concepts into accessible narratives. However, aggregation obscures component performance, potentially masking deterioration in specific dimensions compensated by improvement elsewhere. Weighting decisions prove contentious, whether equal weights implying questionable value equivalence, statistical weights prioritising measurability over importance, or normative weights reflecting contestable value judgments. For social service evaluation, composite indices risk reducing complex service systems to single

Funded by the European Union
Project number: 101179032

numbers that cannot guide targeted improvement in specific domains requiring attention.

Dashboards presenting indicators without aggregation represent 33% (21 frameworks), showing growing preference particularly among national initiatives developed since 2015. Dashboard approaches preserve multidimensional complexity, enable users to examine performance across specific domains relevant to their interests, avoid contentious weighting decisions, and facilitate targeted policy response when specific indicators reveal problems. The shift toward dashboards in recent national and subnational frameworks suggests learning from earlier composite index limitations, particularly regarding transparency and actionability for diverse policy domains including social services. UK Measures of National Wellbeing, New Zealand's Living Standards Framework, and similar national initiatives demonstrate dashboard feasibility for policy integration. However, dashboards demand sophisticated users capable of synthesising information across multiple indicators without mechanical combination rules, potentially limiting accessibility compared to headline composite scores. Dashboard design choices regarding indicator selection, organisation, and visualisation substantially affect usability and interpretation.

Survey-based measures relying primarily on subjective assessments comprise 11% (7 indicators), offering direct capture of lived experience and user perspectives on wellbeing dimensions. Frameworks including the Australian Unity Wellbeing Index and European Quality of Life Survey prioritise subjective evaluations over objective conditions, revealing how circumstances translate into experienced quality of life and whether objective improvements generate corresponding subjective benefit. Survey approaches prove valuable for capturing dimensions resistant to objective quantification including dignity, autonomy, and sense of meaning. Limitations include cost constraining frequency and sample size, cultural variation in response styles complicating cross-national comparison, and limited actionability when satisfaction drivers remain unclear without accompanying objective indicators. For social services, survey measures prove essential for user voice and experience quality assessment but require complementary objective metrics for comprehensive evaluation.

Adjusted GDP measures modifying national accounts to incorporate social and environmental factors account for 5% (3 frameworks), maintaining connection to established economic measurement whilst attempting to broaden scope. Remaining examples include Genuine Progress Indicator and similar efforts to correct GDP for factors including environmental degradation, income distribution, and non-market household production. These approaches offer conceptual continuity with GDP, potentially facilitating uptake among economists and finance ministries, and demonstrate feasibility of monetary aggregation across heterogeneous impacts.

Funded by the European Union
Project number: 101179032

However, adjusted GDP approaches face criticism for forcing incommensurable dimensions onto monetary scales, limited international adoption suggesting lack of consensus regarding adjustment methodologies, and persistence of problematic growth-focused framing when GDP remains the starting point even with modifications.

This methodological distribution reflects competing priorities in indicator design. Composite indices serve communication and political purposes essential for Beyond-GDP legitimacy and uptake. Dashboards support analytical needs and policy integration, particularly for multisectoral challenges including social service improvement requiring attention to specific dimensions rather than aggregate scores. The meta-analysis finding that dashboards increasingly dominate recent national frameworks whilst composites persist in international comparison contexts suggests methodological pluralism may prove optimal, with different approaches suited to different purposes rather than single superior methodology.

3.6 Domain Coverage Across Indicators

Systematic mapping of the 66 indicators against the 22-domain taxonomy reveals pronounced patterns of coverage concentration and systematic neglect. Understanding which dimensions receive attention and which remain invisible provides essential context for framework development addressing social service evaluation requirements.

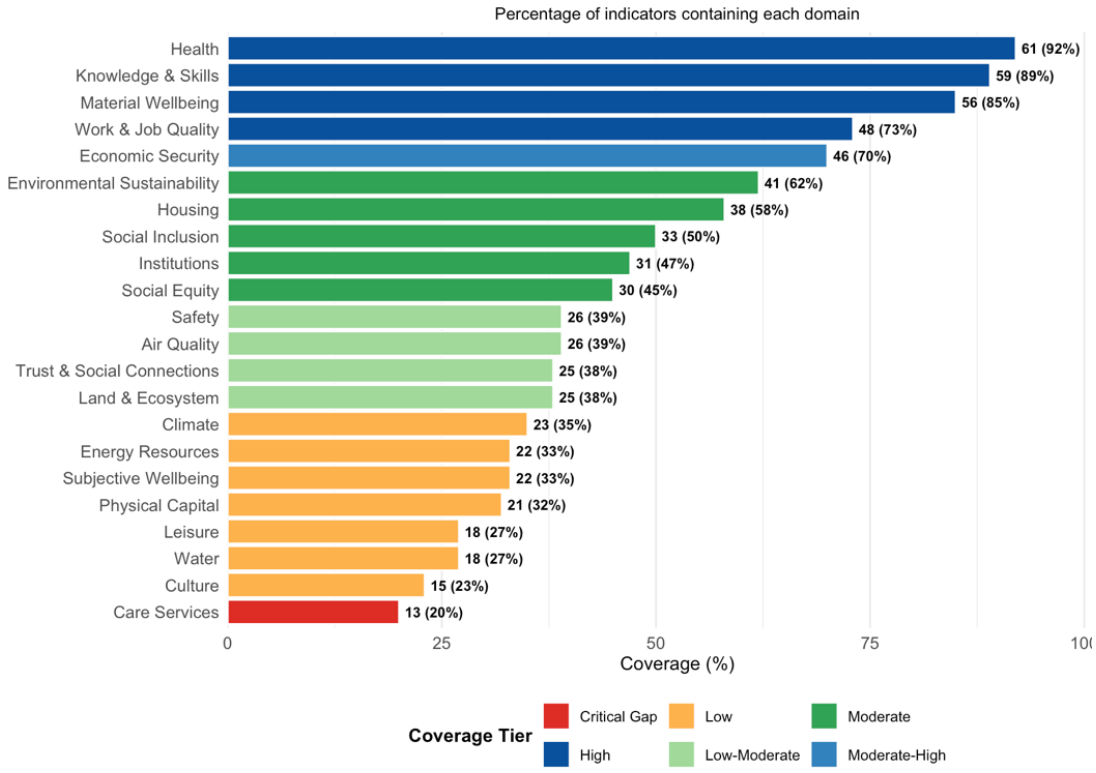


Figure 3: Domain Coverage Across 66 Indicators

Funded by the European Union
Project number: 101179032

3.7 High-Coverage Domains

Three domains achieve coverage exceeding 85% across the indicator sample, reflecting enduring influence of human development frameworks and centrality of these dimensions to wellbeing conceptualisations across diverse theoretical traditions.

Health appears in 61 of 66 indicators (92%), measured through diverse components including life expectancy at birth and at specific ages, mortality rates across age groups and causes, healthy life years and disability-adjusted metrics, morbidity prevalence for communicable and non-communicable diseases, healthcare access and utilisation patterns, quality of healthcare services, and health expenditure levels. This comprehensive coverage reflects robust international measurement infrastructure through WHO Global Health Observatory (World Health Organization, 2024) and national vital statistics systems, established comparability of mortality and morbidity statistics enabling cross-national analysis, recognised importance of health to quality of life across cultural contexts, and straightforward conceptual link between health outcomes and wellbeing making health a near-universal framework component. Health equity research demonstrates profound disparities in health outcomes across socioeconomic groups (Marmot, 2020), whilst catastrophic health expenditure metrics reveal financial protection gaps (Hsu, Flores, Evans, Mills, & Hanson, 2018). For social services, extensive health measurement provides strong foundation for outcome assessment whilst highlighting need for complementary measurement of service quality, user experience, and equity in access beyond aggregate population health statistics.

Knowledge and Skills features in 59 indicators (89%), captured through educational attainment levels and years of schooling, literacy and numeracy rates, school enrolment and completion across education levels, student performance on standardised international assessments, participation in lifelong learning and adult education, and research capacity through R&D personnel and publications. High coverage builds on UNESCO Institute for Statistics (UNESCO, 2024) infrastructure, comparability enabled by International Standard Classification of Education frameworks facilitating consistent attainment measurement, widespread acceptance of education as fundamental to human development and economic productivity, and established theoretical frameworks linking education to capability development, social mobility, and democratic participation. For social services, comprehensive education measurement supports outcome evaluation for education systems whilst indicating need for assessment of educational equity, quality beyond test scores, and non-cognitive skill development resistant to standardised testing.

Funded by the European Union
Project number: 101179032

Material Wellbeing appears in 56 indicators (85%), encompassing income levels and growth trajectories, poverty incidence and depth across multiple thresholds, consumption patterns and expenditure adequacy, wealth accumulation and distribution, and financial security measures. This domain's prominence reflects ease of monetary quantification enabling straightforward aggregation and comparison, persistent influence of economic thinking in progress measurement even within Beyond-GDP frameworks explicitly challenging GDP primacy, established national accounts infrastructure providing regular income and expenditure data, and material resources' instrumental role in enabling pursuit of diverse wellbeing objectives. For social services, material wellbeing measurement provides essential context for understanding service need, targeting resources toward disadvantaged populations, and assessing whether service provision adequately addresses economic vulnerability or merely manages symptoms whilst underlying material deprivation persists.

3.8 Moderate-Coverage Domains

A second tier of domains appears in 40 to 75% of indicators, representing recognised but incompletely integrated dimensions where measurement approaches continue evolving.

Work and Job Quality features in 48 indicators (73%), extending beyond simple employment rates to encompass unemployment and underemployment, job satisfaction and workplace conditions, work-life balance and time poverty, occupational safety and health impacts, skills utilisation and career development opportunities, and employment security and precariousness. Moderate-high coverage reflects relatively recent recognition that employment rates alone inadequately capture labour market wellbeing, growing policy attention to job quality following concerns about in-work poverty and precarious employment, and ILO Decent Work (International Labour Organization, 2012) framework providing conceptual architecture for quality assessment. Meta-analyses of active labour market programmes demonstrate variable effectiveness across intervention types and contexts (Card, Kluve, & Weber, 2018; Kluve et al., 2014). For social services, work measurement informs active labour market program evaluation, reveals employment barriers requiring support services, and contextualises economic security affecting service need.

Economic Security appears in 46 indicators (70%), addressing vulnerability to income shocks through social protection coverage across contingencies, household savings and financial buffers, debt levels and serviceability, insurance protection against major risks, and exposure to catastrophic expenditures. Coverage reflects growing recognition that income adequacy proves insufficient without attention to volatility and resilience, OECD and ILO initiatives developing social protection measurement

Funded by the European Union
Project number: 101179032

frameworks, and SDG emphasis on universal social protection. For social services, economic security measurement reveals populations lacking buffers who require crisis assistance, assesses social protection adequacy, and identifies catastrophic cost exposure when services require user contributions.

Environmental Sustainability features broadly in 41 indicators (62%), though fragmented across sub-domains including air quality (26 indicators, 39%), climate change (23 indicators, 35%), land and ecosystems (25 indicators, 38%), energy resources (22 indicators, 33%), and water (18 indicators, 27%). Aggregate environmental coverage reflects growing recognition of planetary boundaries and ecological foundations for human wellbeing, particularly following Paris Agreement and SDG adoption. However, no single environmental domain achieves the prominence of health, education, or income, suggesting persistent anthropocentric framing treating environment as context rather than constitutive of wellbeing. Fragmentation across sub-domains reflects measurement complexity requiring specialised scientific assessment and diverse policy communities addressing climate, pollution, biodiversity, and resource management through partially disconnected frameworks.

Housing appears in 38 indicators (58%), measured through affordability relative to income, quality and adequacy to household needs, tenure security and homelessness risk, overcrowding and spatial adequacy, and access to basic services and amenities. Moderate coverage reflects housing's recognised importance whilst facing measurement challenges balancing objective conditions with affordability and adequacy perceptions. For social services, housing measurement proves essential given interconnections with health, employment, education attendance, and family stability. Gender dimensions of housing prove particularly neglected, with women disproportionately affected by housing insecurity and homelessness (Shelter Scotland & Engender, 2024).

Social Inclusion features in 33 indicators (50%), addressing participation in social, economic, cultural, and political life through employment integration, social participation and community engagement, cultural access, political participation and civic voice, and freedom from discrimination. Moderate coverage indicates recognition of inclusion as important wellbeing dimension whilst facing conceptual ambiguity regarding boundaries between inclusion as outcome versus process, measurement challenges distinguishing formal access from quality of participation, and debate regarding whether inclusion belongs within wellbeing frameworks or constitutes enabling condition.

Institutions appear in 31 indicators (47%), encompassing rule of law, control of corruption, government effectiveness, regulatory quality, and voice and accountability mechanisms. Moderate coverage suggests governance dimensions' growing integration

Funded by the European Union
Project number: 101179032

in comprehensive progress assessment, building on World Bank Worldwide Governance Indicators (Kaufmann et al., 2011) and similar frameworks. For social services, institutional quality determines whether legal entitlements translate into actual provision and resources reach intended beneficiaries.

Social Equity features in 30 indicators (45%), addressing distribution through income and wealth inequality, equality of opportunity, intergenerational mobility, and disparities across demographic groups. Moderate coverage reflects tension between aggregate wellbeing focus and distributional concerns, with some frameworks treating equity as distinct dimension whilst others rely on disaggregation of outcome indicators.

Safety appears in 26 indicators (39%), measured through crime rates and victimisation, violence against women and children, perceptions of personal security, and traffic accidents. Coverage reflects tension between objective crime statistics and subjective security perceptions, with considerable cross-national variation in reporting practices and cultural norms around acceptable violence levels complicating international comparisons.

3.9 Low-Coverage Domains

Several domains critical to social services appear in fewer than 40% of indicators, revealing systematic blind spots in Beyond-GDP measurement practice.

Trust and Social Connections features in only 25 indicators (38%), despite social capital theory emphasising relationship quality as fundamental to wellbeing. Components include generalised trust in others, trust in institutions, social network size and support, participation in community activities, loneliness and isolation, and reciprocity norms. Low coverage reflects measurement challenges including cultural variation in trust norms, difficulty distinguishing cause from effect when social cohesion both enables and results from other wellbeing dimensions, and lack of standardised instruments comparable to economic or health statistics. For social services, social connection measurement proves essential for mental health interventions, community development programs, and care services addressing isolation.

Subjective Wellbeing appears in 22 indicators (33%), concentrated in frameworks explicitly adopting wellbeing approaches post-Stiglitz-Sen-Fitoussi Commission. Components include life satisfaction, experienced happiness and positive affect, sense of meaning and purpose, and domain-specific satisfactions. Earlier generation indicators rarely included subjective measures, reflecting scepticism about validity, policy relevance, and comparability. Recent acceptance remains incomplete, with many national statistical offices treating subjective measures as supplementary rather

Funded by the European Union
Project number: 101179032

than core indicators. For social services, subjective wellbeing proves valuable for capturing user experience and ultimate service value beyond objectively measured capabilities.

Energy Resources features in 22 indicators (33%), addressing energy access, efficiency, renewable shares, and fuel poverty. Moderate-low coverage reflects energy's ambiguous status as either enabling condition or intrinsic wellbeing dimension. For social services, energy poverty directly affects health through cold housing whilst constraining activities including studying and economic participation.

Physical Capital appears in 21 indicators (32%), encompassing transport infrastructure, telecommunications access, public facilities, and housing stock. Coverage reflects ambiguity about whether infrastructure constitutes wellbeing directly or merely enables it, with outcome-focused frameworks tending to exclude infrastructure whilst capability approaches include access to quality built environments.

Water features in 18 indicators (27%), addressing access to safe drinking water, sanitation coverage, water quality, and water stress. Relatively low coverage despite SDG prominence partly reflects geographic concentration of indicators in high-income contexts where universal water access reduces salience, with frameworks developed in settings with piped water focusing on quality rather than basic access.

Leisure appears in 18 indicators (27%), typically through time poverty measures assessing work hours and commuting time rather than leisure quality or adequacy. Low coverage reflects productivist bias privileging paid work over other time uses, measurement focus on aggregate economic performance rather than individual time allocation, and difficulty quantifying leisure quality beyond time use. For social services, temporal dimensions prove relevant through work-life balance affecting care provision capacity, service opening hours determining access for time-constrained individuals, and person-centred scheduling respecting user routines.

Culture features in only 15 indicators (23%), addressing cultural participation, heritage preservation, and cultural vitality. Minimal coverage reflects difficulty quantifying cultural dimensions, diversity in cultural values across societies complicating standardised assessment, and instrumentalist frameworks treating culture as input to creativity and tourism rather than intrinsic flourishing dimension. For social services, cultural dimensions prove relevant through culturally appropriate provision, cultural participation for isolated populations, and community cultural assets strengthening informal support.

Funded by the European Union

Project number: 101179032

3.10 Critical Service Domains

The analysis confirms severe underrepresentation of domains directly relevant to social service evaluation, demonstrating systematic blind spots requiring deliberate correction.

Care Services appears explicitly in only 13 indicators (20%), despite essential role supporting children, persons with disabilities, frail elderly individuals, and those with chronic illness. Components addressing care include childcare availability and quality, eldercare provision, disability support services, informal care burden, unpaid care work time, and carer strain. This striking gap reflects systematic devaluation of care work in measurement frameworks, perpetuating invisibility of services predominantly serving vulnerable populations and relying on female labour. Care service neglect carries severe consequences including rendering substantial welfare state activity invisible in progress assessment, perpetuating gender inequalities by ignoring women's disproportionate care responsibilities, preventing evidence-informed policy regarding care financing, workforce development, and quality standards, and obscuring care-employment tensions affecting both service access and labour market participation.

Service Integration and Coordination appears in fewer than seven indicators (less than 11%). Most frameworks measure sectoral outcomes in health, education, and social protection separately without capturing cross-service collaboration or added value of integrated delivery for individuals with complex needs spanning multiple service domains. This measurement gap reflects siloed statistical traditions following administrative boundaries, conceptual difficulty defining and operationalising integration distinct from individual service effectiveness, and lack of established metrics for coordination quality. For social services addressing complex needs requiring health, social care, housing, and employment support coordination, integration measurement proves essential for identifying fragmentation costs and evaluating coordination initiatives.

Prevention and Early Intervention similarly lacks dedicated measurement, appearing in approximately four indicators (6%). Frameworks capturing crisis responses through hospitalisation rates, emergency service use, or benefit claims rarely assess preventive services successfully averting problems before they require intensive intervention. This gap creates perverse measurement incentives favouring reactive approaches where crises generate visible metrics whilst successful prevention produces non-events resistant to quantification. Prevention measurement requires counterfactual reasoning estimating what would occur without intervention, long-term follow-up matching prevention payback periods, and willingness to value problems averted rather than only crises managed.

Funded by the European Union
Project number: 101179032

The systematic underrepresentation of care services, integration, and prevention demonstrates that Beyond-GDP frameworks, despite breadth compared to GDP, inadequately address dimensions central to social service evaluation. Section 4 addresses these gaps through explicit framework architecture incorporating undervalued dimensions alongside empirically validated pillars.

3.11 How Social Services Appear Across Indicator Types

Classification of indicators by their measurement logic reveals four archetypal approaches to conceptualising and measuring social services, each with distinct implications for service evaluation.

3.11.1 System Performance Indicators

Approximately 26% of the sample treats social services as system outputs to be optimised according to efficiency and effectiveness criteria. Examples include the EU Social Scoreboard, OECD Social Protection indicators, and health system performance indices. These frameworks typically measure coverage rates indicating proportion of eligible or target populations receiving services, access metrics including waiting times and geographic proximity, efficiency ratios calculating cost per outcome achieved, system capacity through infrastructure and workforce levels, and processing speeds for applications and service delivery.

This approach offers clear accountability metrics suitable for performance management, enables identification of bottlenecks and inefficiencies in service delivery chains, facilitates benchmarking across providers or jurisdictions, and provides actionable signals when specific system dimensions underperform. Limitations include risk of reducing complex services to simple metrics that may not capture quality dimensions mattering to users, potential for gaming through selective focus on measured aspects whilst neglecting unmeasured dimensions including dignity and respect, difficulty attributing population outcomes to specific system performance given confounding factors, and tendency toward input/output measurement rather than ultimate wellbeing impacts.

3.11.2 Human Development Indicators

The largest group at approximately 35% frames social services as investments in human capabilities, following the theoretical tradition established by Sen and operationalised through UNDP frameworks. The Human Development Index and its variants including the Inequality-adjusted HDI demonstrate sustained methodological evolution (Nations, 2020; UN, 1997; Prados de la Escosura, 2021; Felice & Vasta, 2015). The HDI family, Social Progress Index, and most national wellbeing frameworks fall

Funded by the European Union
Project number: 101179032

within this category. They assess services through population capability outcomes including literacy rates, life expectancy, and educational attainment, opportunity measures capturing access to education and healthcare independent of utilisation, basic needs fulfilment across nutrition, shelter, and safety domains, and freedom expansion enabling choice over valued functioning.

Strengths include focus on ultimate goals rather than intermediate service outputs, international comparability through standardised capability metrics, resonance with rights-based frameworks emphasising entitlements to capability development, and attention to equity through disaggregated assessment revealing capability distribution. Limitations encompass weak attribution linking observed capabilities to specific service interventions given multiple influences on human development including family background and economic conditions, long lag times between service delivery and capability formation complicating accountability when outcomes manifest years after intervention, potential neglect of process values including dignity and autonomy in service delivery that matter independently of outcomes achieved, and insufficient attention to service quality versus coverage when capability metrics improve through expanding low-quality provision.

3.11.3 Subjective Wellbeing Indicators

Approximately 15% of indicators prioritise user experience and satisfaction, treating individual assessments as ultimate measure of service value. The World Happiness Report, Better Life Index user evaluations, and quality-of-life surveys exemplify this approach. Research on subjective wellbeing measurement demonstrates validity and policy relevance (Kahneman & Krueger, 2006; OECD, 2013; Cummins, Eckersley, Pallant, van Vugt, & Misajon, 2003; Veenhoven, 1996, 2014). These frameworks capture life satisfaction among service users, perceived service quality and adequacy, trust in institutions providing services, sense of security and support derived from service availability, and domain-specific satisfactions across health, education, and social protection.

This approach directly captures lived experience valued for its own sake rather than as means to other ends, reveals perception gaps where objective service availability does not translate to subjective security or satisfaction, enables identification of process dimensions affecting experience beyond measurable outcomes including respect and communication quality, and centres user voice in evaluation rather than external expert assessment. Limitations include susceptibility to adaptation effects where expectations adjust to service levels, making satisfaction unstable measure of absolute quality, cultural variation in response styles and satisfaction thresholds complicating cross-national comparison, limited actionability when satisfaction drivers remain unclear without accompanying objective indicators specifying which service

Funded by the European Union
Project number: 101179032

dimensions require improvement, and potential conservatism when disadvantaged populations express satisfaction with inadequate services due to low expectations or lack of awareness regarding alternatives.

3.11.4 Rights and Entitlements Indicators

Approximately 24% assess services against normative standards derived from legal obligations or ethical principles. The SDGs, human rights indicators, and equality indices exemplify this approach, evaluating fulfilment of legal entitlements established through domestic law or international conventions, non-discrimination in access across protected characteristics, accountability mechanisms enabling recourse when entitlements are denied, minimum standards achievement often specified as floors below which no one should fall, and progressive realisation tracking movement toward full entitlement fulfilment over time. Evidence on voice and agency in development contexts emphasizes rights-based approaches (Klugman, Hanmer, Twigg, Hasan, McCleary-Sills, & Santamaria, 2014).

Clear normative frameworks provide unambiguous assessment criteria independent of political preferences or resource constraints, legal backing strengthens claims for service adequacy and quality whilst enabling litigation when entitlements prove unfulfilled, rights approaches centre human dignity and universal claims rather than utilitarian aggregation, and standards focus attention on ensuring minimum acceptable provision for all rather than optimising average outcomes. Limitations include binary pass-fail logic that may not capture quality gradations or incremental improvement, difficulty establishing thresholds simultaneously ambitious and realistic across diverse contexts with varying resources and starting points, potential rigidity when fixed standards prove inappropriate for specific populations or circumstances, and tension between progressive realisation acknowledging resource constraints versus immediate obligation rhetoric.

None of these four approaches proves comprehensively adequate for social service evaluation. System performance indicators capture efficiency but miss ultimate value. Human development frameworks assess outcomes but struggle with attribution. Subjective wellbeing centres user voice but lacks actionable specificity. Rights frameworks provide normative clarity but resist empirical gradation. Comprehensive service evaluation requires integration across these traditions, combining efficiency metrics, outcome assessment, user experience, and normative standards within coherent frameworks. Section 4 develops such integrative architecture.

Funded by the European Union
Project number: 101179032

3.12 Methodological Patterns

Analysis of methodological approaches across the 66 indicators reveals patterns significantly affecting utility for social service evaluation. Data sources, update frequency, aggregation methods, and equity integration shape whether frameworks can support responsive service management and equitable resource allocation.

Funded by the European Union

Project number: 101179032

Table 1: Methodological Characteristics Summary

Characteristic	Distribution
Data Sources	
Administrative data	59 indicators (89%)
Survey data	41 indicators (62%)
Mixed sources	31 indicators (47%)
Update Frequency	
Annual	42 indicators (64%)
Multi-year/Irregular	18 indicators (27%)
Quarterly or more frequent	2 indicators (3%)
Discontinued/Unclear	4 indicators (6%)
Publication Lag (median)	18 months
Equity Disaggregation	
Gender/Sex	37 indicators (56%)
Age	31 indicators (47%)
Geographic	45 indicators (68%)
Income/Wealth	34 indicators (52%)
Disability	8 indicators (12%)
Migration status	6 indicators (9%)
Ethnicity/Race	12 indicators (18%)
Intersectional Analysis	8 indicators (12%)

3.13 Data Source Patterns

Administrative data appears in 89% (59 indicators), leveraging routinely collected statistics on service provision, utilisation, and population characteristics. Administrative sources include vital statistics registries, education enrolment and assessment systems, healthcare utilisation and clinical records, benefit and social assistance caseloads, employment and unemployment records, and housing and homelessness databases. Advantages include regular availability enabling frequent updates, comprehensive population coverage avoiding sampling limitations, relatively low marginal cost once collection systems are established, individual-level detail enabling disaggregation and longitudinal analysis, and integration across administrative systems through personal identifiers in countries with population registers.

Disadvantages encompass quality variation across jurisdictions and time periods reflecting differential investment in information systems, definitional inconsistencies impeding comparability when administrative categories follow program rules rather than conceptual frameworks, gaming risks when metrics directly influence performance assessment or funding allocation creating incentives for manipulation,

Funded by the European Union
Project number: 101179032

and systematic gaps for quality dimensions and user experience absent from administrative records focused on transactions and legal compliance. For social services, administrative data provides strong foundation for coverage and output measurement but proves weak for satisfaction, dignity, and process quality requiring direct user assessment.

Survey data features in 62% (41 indicators), typically through specialised wellbeing surveys, multipurpose household surveys, or service-specific user experience assessments. Surveys enable capture of subjective experience and user perspectives, measurement of constructs unavailable in administrative records including life satisfaction and relationship quality, user voice in assessment, and flexible adaptation to emerging measurement priorities. Limitations include cost constraining frequency and sample size particularly for granular geographic or demographic disaggregation, sampling limitations for rare populations including persons with severe disabilities or homeless individuals often excluded from household sampling frames, response bias when participation correlates with characteristics of interest, and difficulty maintaining consistent question wording across contexts, languages, and time periods whilst adapting to evolving concepts.

For service evaluation, surveys prove essential for satisfaction assessment, outcome attribution through baseline and follow-up measurement, and experience quality dimensions, but cannot support real-time monitoring or comprehensive coverage of service-intensive rare populations. Optimal survey design for service evaluation includes adequate sample sizes for disaggregation (minimum 2,000 respondents, with oversampling of service users), individual linkage to administrative service receipt enabling outcome attribution, validated instruments facilitating cross-national comparison, and integration with regular statistical programs ensuring sustainability.

Mixed-source approaches combining administrative and survey data appear in 47% (31 indicators), offering optimal potential for comprehensive assessment integrating objective conditions with subjective experience. Best practice examples include Nordic wellbeing dashboards integrating population registers with quality-of-life surveys, enabling individual-level linkage whilst protecting privacy through statistical disclosure control. Mixed approaches demand complex coordination across data collection systems operated by diverse organisations, legal frameworks enabling data sharing whilst protecting confidentiality, technical infrastructure for secure linkage, and analytical capacity for integrated analysis. Investment in mixed-source infrastructure proves worthwhile for comprehensive service evaluation requiring both population-level outcome tracking and user experience assessment.

Funded by the European Union

Project number: 101179032

3.14 Update Frequency and Timeliness

Update frequency critically affects policy relevance and integration with decision cycles. Annual updates characterise 63% (42 frameworks), enabling regular monitoring and year-on-year comparison whilst accommodating annual budget cycles and statistical production schedules. Most national dashboards and international comparative frameworks employ annual updates, balancing timeliness with data quality and resource demands. Irregular or multi-year updates apply to 27% (18 frameworks), often due to dependence on infrequent census data or intensive primary data collection including time-use surveys conducted quinquennially.

Quarterly or more frequent updates appear in only 3% (2 indicators), both drawing on monthly administrative series enabling near-real-time reporting. The ONS4 subjective wellbeing questions added to UK labour force surveys provide quarterly life satisfaction updates, whilst certain national dashboards including health system indicators report monthly. Five frameworks (8%) are discontinued or have unclear update status following institutional changes or funding cessation.

Median publication lag from reference period to data release spans 18 months across frameworks reporting this information. International frameworks coordinating across national statistical systems incur longest lags as data are compiled, validated, and standardised. National dashboards drawing primarily on domestic administrative sources achieve shorter lags, with best performers publishing within 6-8 months. Only frameworks exploiting monthly administrative series approach real-time reporting with lags under 3 months.

This 18-month median contrasts sharply with GDP quarterly national accounts typically published 6-8 weeks after period end. The timeliness gap partially explains GDP's persistence in policy discourse despite recognised limitations. Beyond-GDP frameworks struggling with timeliness face structural disadvantage in fast policy cycles where recent data command attention and enable responsive adjustment. For social services, where program management benefits from rapid feedback identifying emerging problems, publication lags undermine frameworks' utility for operational decision-making versus strategic assessment.

Emerging approaches attempting to improve timeliness include nowcasting techniques predicting recent outcomes combining timely partial data with historical patterns, high-frequency indicators exploiting monthly administrative series or real-time digital data, leading indicators predicting future outcomes based on early warning signals, and preliminary estimates subsequently revised as comprehensive data become available. Statistical agencies should prioritise temporal enhancement, potentially accepting

Funded by the European Union
Project number: 101179032

modest precision reduction for substantial timeliness gains when frameworks primarily serve monitoring rather than legal or financial obligations requiring maximum accuracy.

3.15 Aggregation and Weighting Approaches

Among the 34 composite indicators, weighting methodologies reveal fundamental philosophical choices with implications for how different dimensions, including social services, are valued relative to other domains.

Table 2: Composite Index Weighting Approaches

Weighting Method	N	% of Composites	Strengths	Limitations
Equal weighting	18	53%	Avoids value judgments; transparent	May not reflect importance; cascading effects
Statistical weighting (PCA, DEA)	6	18%	Data-driven; reduces subjectivity	Prioritises measurability; context instability
Expert weighting (Delphi, panels)	8	24%	Incorporates judgment	Elite bias risk
Geometric mean (implicit)	2	6%	Penalises imbalance	No explicit weights; less intuitive

Note: Weighting methodology classifications based on publicly documented approaches; some frameworks use hybrid methods

Equal weighting appears in 18 indicators (53% of composites[1]), embodying the implicit assumption that all dimensions contribute equally to overall progress. Equal weighting offers transparency, avoids contentious value judgments, proves easy to explain and replicate, and maintains political neutrality not privileging particular constituencies. However, weights cascade through hierarchical structures such that equal weighting at higher levels does not guarantee equal influence of underlying components. Service domains aggregated with broader dimensions through equal weighting may receive systematically lower effective weight when broader dimensions contain more sub-components. Additionally, equal weighting across incommensurable dimensions with different units and scales proves arbitrary, as normalisation choices substantially affect relative influence.

[1] Extensive methodological literature addresses composite index construction including weighting, aggregation, and robustness issues (OECD, 2008; Greco, Ishizaka, Tasiou, & Torrisi, 2019; Mazziotta & Pareto, 2013; Foster, McGillivray, & Seth, 2013).

Funded by the European Union
Project number: 101179032

Statistical weighting through methods including principal components analysis, factor analysis, and data envelopment analysis appears in 6 indicators (18% of composites). These data-driven approaches reduce subjectivity, enable weights to emerge from observed variation patterns, and maximise variance explained by composite indices. However, statistical weighting may prioritise measurability over importance when easily measured dimensions receive higher weights due to greater variation, prove unstable when applied across contexts with different covariance structures, and lack normative foundation as empirical correlation does not imply relative value.

Expert weighting through Delphi methods or expert panels features in 8 indicators (24% of composites). This approach incorporates professional judgment and substantive knowledge, enables sophisticated assessment of relative importance, and proves flexible adapting to evolving understanding. Risks include elite bias when experts' priorities diverge from those of service users or general populations, limited transparency when weighting rationales remain implicit, and instability as expert composition changes over time.

Two composite indicators (6%) avoid explicit weighting by employing geometric means or minimum operators penalising imbalance across dimensions. These approaches implicitly treat deficiency in any domain as requiring compensation through superior performance elsewhere, preventing high aggregate scores masking severe deprivation in specific dimensions. Geometric aggregation proves particularly appropriate for frameworks treating dimensions as substitutable only at diminishing rates, but interpretation proves less intuitive than arithmetic means.

Dashboard approaches presenting disaggregated indicators without aggregation characterise 21 frameworks (33% of total sample), circumventing weighting dilemmas entirely whilst requiring sophisticated users capable of synthesising multidimensional information without mechanical combination rules. Dashboards reflect judgment that transparency and actionability outweigh communicative simplicity, enabling diverse stakeholders to focus on dimensions relevant to their interests and responsibilities.

For social service evaluation, weighting choices prove consequential. Equal weighting across broad domains including economic, social, environmental, and governance dimensions may systematically undervalue specialised service concerns when aggregated with expansive categories. Participatory weighting engaging service users proves normatively attractive but faces practical challenges. Dashboard approaches avoiding aggregation enable focused attention on service-specific indicators whilst complicating communication to broad audiences. Optimal frameworks likely employ dashboards for detailed service evaluation whilst developing selective composites for high-level communication when necessary.

Funded by the European Union
Project number: 101179032

3.16 Equity and Disaggregation

Despite rhetorical commitment to inclusive measurement and leaving no one behind, systematic equity analysis remains limited across the indicator landscape, with pronounced gaps in dimensions most relevant to social service evaluation.

Gender disaggregation[2] appears in 37 indicators (56%), age disaggregation in 31 indicators (47%), and geographic disaggregation in 45 indicators (68%). These relatively high rates reflect established statistical practice routinely collecting and reporting age, sex, and location. Gender disaggregation enables assessment of whether women and men experience comparable outcomes and whether gender gaps narrow over time, though often limited to binary categories insufficient for analysing gender diversity. Age disaggregation typically employs broad categories (children, working-age, elderly) rather than granular age groups, limiting precision for life-course analysis. Geographic disaggregation ranges from crude urban-rural dichotomies to detailed regional or municipal breakdowns, with granularity depending on sample sizes and privacy protection requirements.

Income or wealth disaggregation features in 34 indicators (52%), often through quintile comparisons or Gini coefficients. Socioeconomic disaggregation proves essential for assessing distributional progressivity and identifying whether progress reaches disadvantaged populations. However, many frameworks present only aggregate inequality metrics rather than disaggregated outcome indicators across income groups, limiting insight into specific mechanisms generating or reducing disparities.

Disability disaggregation appears in only 8 indicators (12%), migration status disaggregation in merely 6 indicators (9%), and ethnicity/race disaggregation in approximately 12 indicators (18%). These low rates represent serious blind spots for social service evaluation, as persons with disabilities, migrants, and ethnic minorities constitute populations with intensive service needs and particular vulnerabilities to access barriers, discrimination, and inadequate service quality. Disability disaggregation proves methodologically challenging given definitional variation across jurisdictions, spectrum from mild to severe impairments, and measurement dependent on self-report versus clinical assessment versus functional limitation criteria. Migration status disaggregation faces similar challenges distinguishing foreign-born from foreign nationals, recent from established migrants, refugees from economic migrants, and documented from undocumented status.

[2] Gender-sensitive statistical frameworks provide guidance for equity-focused disaggregation (Beck, 1999; Gardner, 2016; OXFAM, 2014), yet implementation remains inconsistent across Beyond-GDP indicators.

Funded by the European Union
Project number: 101179032

Intersectional analysis combining multiple characteristics simultaneously appears in only 8 indicators (12%). This severe limitation prevents examination of how disadvantages compound, such as elderly women with disabilities facing distinctive challenges not captured by separate analysis of age, gender, and disability effects. Intersectional analysis requires either very large samples enabling joint categories with adequate statistical power, or sophisticated modelling techniques including multilevel analysis. The scarcity suggests that technical capacity, sample size constraints, and limited awareness of intersectionality's importance prevent widespread implementation despite growing recognition of its analytical value.

The gap between disaggregation capability and actual reporting practice further limits equity analysis. Many frameworks technically enabling disaggregation through underlying survey data do not routinely publish disaggregated results, leaving analytical potential unrealised in accessible outputs. Publication selectivity reflects privacy concerns when cell sizes prove small, resource constraints limiting production of multiple report variants, and insufficient user demand when policy attention focuses on aggregate performance rather than distributional patterns.

For social service evaluation, disaggregation proves essential rather than optional. Services explicitly target disadvantaged groups, distributional equity constitutes primary policy objective, and aggregate metrics mask wide variation in access, quality, and outcomes across populations. Comprehensive frameworks must incorporate routine disaggregation by gender, age, disability, geography, socioeconomic status, and migration background at minimum, with intersectional analysis for populations facing compounding disadvantages. Investment in sample sizes, administrative data linkage, small area estimation techniques, and privacy-preserving publication methods enables equity-focused measurement whilst protecting confidentiality.

3.16.1 Gender Dimensions in Beyond-GDP Measurement

Building on the disaggregation patterns documented above, gender dimensions warrant dedicated synthesis given their intersection with the care service measurement gaps identified in Section 3.10. Whilst 56% of indicators enable gender disaggregation, this capacity remains underutilised when frameworks fail to connect gender analysis with domains where gendered patterns prove most consequential for social service evaluation.

The systematic underrepresentation of care services (20% coverage) carries profound gender implications. Care work remains disproportionately performed by women, with the International Labour Organization documenting that women perform over three-quarters of unpaid care work globally. The invisibility of care domains in Beyond-GDP measurement perpetuates this gendered devaluation, rendering substantial welfare

Funded by the European Union
Project number: 101179032

state activity invisible whilst obscuring women's disproportionate care responsibilities. Care service neglect prevents evidence-informed policy regarding care financing, workforce development, and quality standards that would address gendered care burdens.

The care-employment tension particularly affects women's labour market participation. Women face distinctive barriers to employment arising from unpaid care responsibilities, including reduced working hours, career interruptions, and occupational segregation into care-related sectors offering lower wages and fewer advancement opportunities. Beyond-GDP frameworks inadequately capture these dynamics when measuring work quality and economic security without attention to how care responsibilities constrain labour market access differentially by gender.

Dedicated gender indices including the Gender Development Index, Gender Inequality Index, and Global Gender Gap Index demonstrate methodological feasibility of comprehensive gender assessment. These frameworks integrate gender dimensions across education, health, economic participation, and political empowerment domains. The feminist economics literature referenced throughout this analysis provides theoretical grounding for gender-sensitive measurement, emphasising recognition of reproductive labour, analysis of time poverty, and assessment of how social institutions allocate care responsibilities. Yet these specialised gender frameworks remain peripheral to mainstream Beyond-GDP measurement, treated as supplementary rather than integral to comprehensive progress assessment.

The intersectional analysis limitations documented above (12% coverage) prove particularly concerning for gender-sensitive evaluation. Elderly women, women with disabilities, migrant women, and women from ethnic minorities face compounding disadvantages not captured through simple gender disaggregation. Intersectional analysis examining how gender intersects with age, disability, migration status, and socioeconomic position proves essential for identifying populations facing distinctive service barriers and designing appropriately targeted interventions.

These findings inform BENEFITS framework development. Gender-sensitive evaluation should integrate care work visibility through time use assessment and service quality measurement, incorporate gender disaggregation as standard analytical dimension, enable intersectional analysis combining gender with other characteristics, and assess gendered impacts of social services on labour market participation and economic security. The gap between existing gender measurement capacity and its integration into mainstream wellbeing frameworks represents an opportunity for BENEFITS to address gender dimensions systematically rather than peripherally.

Funded by the European Union
Project number: 101179032

3.17 Current Methods for Social Services Impact Measurement

Understanding how social services are currently evaluated provides essential context for comprehensive framework development. Existing evaluation methodologies reflect diverse disciplinary traditions and measurement purposes, from economic efficiency assessment to lived experience documentation. This section synthesises approaches documented in evaluation literature and policy practice, organised by their primary evaluative logic.

3.17.1 Economic Evaluation Methods

Economic evaluation methods address value-for-money questions by comparing costs of service provision against quantified benefits or outcomes. These approaches prove particularly relevant for resource allocation decisions and investment justification in constrained fiscal environments.

Cost-benefit analysis represents the most comprehensive economic approach, monetising both intervention costs and resulting social benefits to calculate benefit-cost ratios or net present values. Applications to social services require valuing non-market outcomes through techniques including willingness-to-pay surveys eliciting maximum amounts individuals would pay for services or outcomes, revealed preference inferring values from observed behaviour and trade-offs, and quality-adjusted life year valuations enabling health and longevity monetisation. Whilst CBA provides clear investment criteria comparable across sectors and enables comparison of incommensurable outcomes through common monetary metric, monetisation of dimensions including dignity, social inclusion, or family stability proves contentious and technically challenging. Critics argue that reducing human wellbeing to financial equivalents proves ethically problematic and that monetary valuation imposes utilitarian logic inappropriate for rights-based services. Applications to social services including the Genuine Progress Indicator demonstrate alternative approaches to valuing non-market dimensions (Lawn, 2005; Cobb, Halstead, & Rowe, 1995).

Cost-effectiveness analysis evaluates relative costs of achieving specified outcomes without requiring full monetisation of benefits. Applications calculate cost per job placement, housing unit secured, quality-adjusted life year gained, or child lifted above poverty threshold. CEA proves more feasible than CBA for many social service contexts whilst maintaining economic rigor, avoids contentious monetisation of intrinsic values, and provides clear efficiency comparisons across interventions pursuing similar objectives. Limitations include inability to compare interventions with incommensurable outcome measures, failure to determine whether any intervention offers sufficient value to justify costs in absolute terms, and potential neglect of secondary outcomes and spillover effects not captured in primary effectiveness metric.

Funded by the European Union
Project number: 101179032

Cost-utility analysis specialises CEA for health and quality-of-life outcomes, typically expressing results as cost per quality-adjusted life year gained. CUA employs standardised preference weights for health states, enabling comparison across diverse health interventions. Whilst widely applied in healthcare priority-setting through NICE (National Institute for Health and Care Excellence, 2013) and similar bodies, CUA translation to broader social services requires adaptation of quality-of-life instruments and disability weights to non-health contexts including education, housing, and social care. Studies of publicly provided services demonstrate significant distributional impacts often invisible in market-based metrics (Verbist, Förster, & Vaalavuo, 2012; Aaberge & Langorgen, 2006).

Social return on investment adapts financial ROI concepts to social contexts, calculating ratios of monetised social value created to investment costs. SROI incorporates stakeholder engagement in outcome identification, applies financial proxies to non-market benefits, accounts for attribution and deadweight, and disaggregates value creation by stakeholder group. Whilst this broadens value assessment beyond conventional economic evaluation and centres stakeholder perspectives, SROI faces criticism for subjectivity in proxy selection, limited comparability across applications employing different proxies and assumptions, insufficient accounting for displacement and substitution effects, and overconfidence in precise monetary valuations of inherently uncertain impacts. Systematic reviews of SROI applications reveal both promise and methodological challenges (Corvo, Pastore, Mastrodascio, & Cepiku, 2022; Marques, Rodrigues, Zerth, & Orrego, 2025; Shields, Roy, Khan, Read, & Ottaway, 2025). Section 4 elaborates SROI conceptual foundations and appropriate application contexts.

3.17.2 Experimental and Quasi-Experimental Methods

Causal identification methods address the attribution challenge, distinguishing service effects from confounding influences through counterfactual logic. These approaches prove essential when demonstrating that observed outcomes result from service provision rather than external factors or selection effects.

Randomised controlled trials assign eligible individuals to treatment and control conditions through random allocation, eliminating systematic selection bias and enabling unbiased effect estimation under minimal assumptions. RCTs provide gold-standard causal evidence and have proliferated in social policy evaluation. However, practical and ethical constraints limit applicability, including costs of implementation for adequately powered trials, sample size requirements for adequate statistical power when effects prove modest, ethical concerns about withholding potentially beneficial services from control groups, limited external validity when experimental conditions differ from routine practice, and political resistance when randomisation proves

Funded by the European Union
Project number: 101179032

incompatible with universal service entitlements or professional discretion in allocation. Evidence from labour market interventions (Card, Kluve, & Weber, 2018; Imbert & Papp, 2015) and social assistance programmes (Shahidi, Ramraj, Sod-Erdene, Hildebrand, & Siddiqi, 2019) demonstrates both the potential and limitations of experimental approaches for service evaluation.

Difference-in-differences designs compare outcome trajectories between groups exposed and unexposed to interventions, controlling for time-invariant differences. DiD requires parallel pre-intervention trends but accommodates observational data, making it feasible for retrospective service evaluation exploiting policy reforms affecting some jurisdictions but not others. DiD applications increasingly incorporate event-study frameworks examining dynamic treatment effects and testing parallel trends assumptions.

Regression discontinuity design exploits threshold-based assignment rules to compare individuals just above and below eligibility cutoffs, who differ only marginally in characteristics determining assignment but differ in treatment receipt. RDD provides credible causal estimates under less restrictive assumptions than RCTs whilst using observational data, though estimates apply only to populations near assignment thresholds rather than full eligible populations.

3.17.3 Participatory and Qualitative Methods

Participatory approaches privilege lived experience and stakeholder perspectives, capturing dimensions of value not reducible to pre-specified metrics. These methods prove particularly important for understanding service quality, dignity in delivery, and outcomes as perceived by users rather than external observers.

Most Significant Change methodology engages stakeholders in collecting and systematically selecting stories describing significant service-related changes. Participants including service users, staff, and community members document changes they consider most significant. Iterative panel review identifies recurring themes and valued outcomes through structured selection processes, revealing what matters to participants beyond researcher-defined measures. MSC proves valuable for surfacing unexpected benefits, understanding mechanisms through which services generate value, and incorporating diverse stakeholder perspectives in complex programs where outcomes cannot be fully anticipated at design stage.

Outcome Harvesting collects evidence of outcomes achieved and works backward to determine whether and how services contributed, rather than measuring predefined indicators against targets. This emergent approach accommodates complex environments where outcomes cannot be fully anticipated, multiple actors contribute

Funded by the European Union
Project number: 101179032

to change making attribution ambiguous, and rigid indicator frameworks risk missing significant but unexpected developments. It emphasises contribution analysis over attribution, acknowledging multiple influences whilst identifying service roles in outcome achievement.

Participatory scorecards combine participatory and quantitative approaches, engaging communities to identify quality dimensions, score current performance using simple rating scales, and monitor changes over time. Communities define locally salient indicators, assess services against these criteria, share findings with providers through interface meetings, and collaboratively develop improvement plans. Scorecards function as continuous improvement tools enabling user voice in accountability processes whilst maintaining quantitative tracking.

3.17.4 Framework-Based Indicator Systems

Standardised outcome frameworks provide consistent measurement scaffolding across service providers and contexts whilst enabling benchmarking and aggregation.

The Adult Social Care Outcomes Toolkit (ASCOT) (Netten et al., 2012) developed for UK social care measurement provides validated instruments assessing quality of life across eight domains including control over daily life, personal cleanliness and comfort, food and nutrition, personal safety, social participation, occupation, accommodation, and dignity. ASCOT enables baseline and follow-up measurement revealing service impacts on user wellbeing, supports cross-provider comparison when consistently applied, and facilitates economic evaluation through preference weights enabling QALY calculation. Limitations include implementation costs requiring trained assessors or survey administration, potential ceiling effects when high-functioning users score near maximum at baseline, and cultural adaptation requirements for international application.

National Themes, Outcomes and Measures (TOMs) (Social Value UK, 2020) frameworks provide standardised outcome categories and proxy monetisation values for social value reporting in UK public procurement. TOMs facilitate cross-organisational comparison, aggregation into social value estimates, and integration with commissioning processes requiring quantified social value. However, proxy values prove contentious and context-dependent, and standardisation may not accommodate local priorities or innovative outcome types.

Beyond-GDP dashboards at national and subnational levels increasingly incorporate service-relevant indicators, particularly in health and education domains. Integration with fiscal reporting cycles enhances policy uptake, whilst routine publication enables longitudinal tracking. However, dashboards risk reinforcing existing measurement

Funded by the European Union
Project number: 101179032

conventions rather than addressing gaps, and policy integration depends on sustained political commitment potentially vulnerable to electoral cycles.

This synthesis reveals that current evaluation methods provide valuable building blocks for comprehensive social service assessment but prove incomplete when applied individually. Economic methods excel at efficiency assessment but struggle with equity integration and non-market value capture. Experimental methods provide causal rigor but face feasibility constraints and ethical concerns. Participatory approaches centre lived experience but resist aggregation across contexts. Framework-based systems enable standardisation but may not capture local priorities. Comprehensive frameworks must integrate insights across these traditions, combining strengths whilst mitigating limitations. Section 4 develops such integrative architecture.

Funded by the European Union

Project number: 101179032

4 FRAMEWORK ARCHITECTURE

4.1 Derivation from Empirical Patterns

Our framework architecture derives from systematic analysis of domain co-occurrence patterns across the 66 indicators rather than from a priori theoretical commitments. This empirical approach ensures alignment with established measurement practice whilst identifying gaps requiring deliberate correction. Analysis of which domains appear together reveals four robust clusters corresponding to distinct but interrelated dimensions of progress: economic foundations, social wellbeing, environmental sustainability, and governance and institutions.

Co-occurrence analysis calculates the frequency with which domain pairs appear within the same indicator. High co-occurrence rates indicate domains that existing frameworks treat as conceptually related or mutually constitutive. Material wellbeing, work and job quality, and economic security demonstrate strong clustering, with 57 to 67% of indicators including at least two of these three domains and 44% including all three. This pattern reflects the economic dimension's theoretical coherence and practical interdependence, where income security, employment quality, and material living standards form an integrated whole.

Health, knowledge and skills, housing, and social protection similarly cluster, appearing together in 47 to 61% of indicators. This social dimension encompasses both outcomes directly constitutive of wellbeing and services enabling capability development. The high co-occurrence rate validates treating these domains as a unified pillar whilst acknowledging internal diversity between outcome and enabler components.

Environmental domains including air quality, water, climate, land and ecosystem, and energy resources demonstrate moderate clustering at 31 to 36% co-occurrence. Lower rates than economic or social domains partially reflect the environmental dimension's relative recency in Beyond-GDP measurement and ongoing debate about whether environment constitutes wellbeing directly or merely conditions enabling it. Nonetheless, consistent joint appearance justifies pillar status.

Institutions, safety, social equity, and social inclusion co-occur in 33 to 39% of indicators, forming a governance dimension addressing power relations, procedural fairness, and enabling conditions for agency. This pillar's moderate co-occurrence reflects diverse theoretical traditions framing governance variously as intrinsic to wellbeing, instrumental to other outcomes, or contextual background rather than direct component.

Funded by the European Union
Project number: 101179032

Critically, several domains relevant to social services demonstrate weak integration with these four clusters. Care services appear alongside other social domains in only 15% of indicators, subjective wellbeing in 27%, culture in 23%, and leisure in 27%. These low rates indicate systematic neglect requiring explicit correction rather than domains naturally fitting within established pillars. Section 4 addresses this challenge through dedicated attention to undervalued dimensions.

The four-pillar structure balances empirical grounding in observed measurement practice with normative commitment to comprehensive assessment. It provides familiar architecture facilitating uptake by policymakers accustomed to economic, social, environmental, and governance categorisations whilst creating space to highlight systematic blind spots requiring focused attention.

4.1.1 Pillar One: Economic Foundations

The economic foundations pillar encompasses material wellbeing, employment quality, economic security, and housing. These domains provide essential context for social service evaluation because economic circumstances determine both the need for services and people's capacity to access provision that requires user contributions.

Material poverty reveals the scale of populations requiring income support and housing assistance. Employment precarity generates financial instability that creates demand for emergency assistance, whilst poor working conditions increase physical and mental health pressures on healthcare systems. Economic insecurity identifies populations lacking financial buffers who require immediate support during crises, whilst also revealing whether unemployment and sickness benefits prove adequate. Housing instability frequently triggers service crises, whereas stable affordable housing enables effective service delivery across multiple domains.

For service evaluation frameworks, the economic pillar establishes baselines for assessing whether service provision genuinely reduces economic vulnerability or merely manages symptoms whilst underlying material insecurity persists. Comprehensive frameworks must therefore integrate economic circumstance assessment alongside service-specific metrics.

4.1.2 Pillar Two: Social Wellbeing

The social wellbeing pillar encompasses health, knowledge and skills, social inclusion, equity, subjective wellbeing, and social connections. These are dimensions that directly constitute human flourishing. Unlike the instrumental economic foundations that enable wellbeing, social domains represent ultimate goals that services explicitly target.

Funded by the European Union
Project number: 101179032

Health and education measurement must distinguish between population outcomes and service provision characteristics. Population health reflects multiple influences including genetics, behaviour, environment, and services, not just healthcare provision alone. Similarly, educational attainment reflects family background, peer effects, and schooling quality. Service evaluation therefore requires methods that isolate the contribution of services from these confounding factors.

Subjective wellbeing captures user experience and dignity in service delivery, revealing how services affect experienced quality of life beyond objectively measured capabilities. Social connections prove both service outcomes (particularly for care services and mental health interventions) and determinants of service access, as socially connected individuals navigate complex service systems more effectively.

For service evaluation, social outcomes provide the ultimate criteria for assessing whether interventions translate into better lives. Measurement frameworks must therefore capture both objective circumstances and subjective experiences to reflect wellbeing comprehensively.

4.1.3 Pillar Three: Environmental Sustainability

The environmental sustainability pillar addresses air quality, water, climate, ecosystems, and energy. Whilst environmental considerations may appear peripheral to social service evaluation, the linkages prove substantial. The concept of safe operating space for humanity (Rockström et al., 2009) and doughnut economics frameworks (Raworth, 2012) provide conceptual foundation for integrating environmental limits with social foundations. Wealth accounting approaches including Inclusive Wealth (Arrow, Dasgupta, Goulder, Mumford, & Oleson, 2012; UN, 2023) and environmental performance indices (Färe, Grosskopf, & Hernandez-Sancho, 2004; WEF, 2001) demonstrate diverse methodological approaches.

Air pollution exposure generates increased healthcare utilisation among vulnerable populations, particularly children, elderly persons, and those with chronic respiratory conditions. Climate change creates demands for emergency response and displacement assistance. Energy poverty directly affects health through cold housing whilst constraining activities such as studying and income generation.

Service infrastructure energy consumption affects both operational costs and carbon footprints, creating sustainability imperatives for facility efficiency and renewable energy adoption. Environmental sustainability proves the least developed pillar for social service application, facing attribution difficulties (environmental changes result from multiple diffuse causes) and scale mismatches (environmental processes operate regionally or globally whilst service provision remains local).

Funded by the European Union
Project number: 101179032

Nonetheless, environmental integration proves essential for comprehensive service evaluation, addressing health determinants, climate adaptation requirements, and long-term service system sustainability in the face of resource constraints.

4.1.4 Pillar Four: Governance and Institutions

The governance and institutions pillar addresses institutional quality, safety, transparency, and civic participation. These are the processes and structures that enable or constrain effective service delivery. Quality governance proves essential for translating resources into effective services and ensuring that services respect rights and dignity.

Institutional quality determines whether legal entitlements translate into actual provision, whether resources reach intended beneficiaries rather than dissipating through corruption or inefficiency, whether service standards receive enforcement through effective oversight, and whether user complaints receive fair hearing and redress. Safety proves fundamental both directly and by affecting service access (crime or violence can deter use of public facilities or community-based programmes). Transparency enables external oversight that identifies waste whilst facilitating learning from performance data.

User involvement in service design and evaluation ensures responsiveness to actual needs rather than provider assumptions. Co-production approaches that engage service users as partners rather than passive recipients improve service relevance whilst respecting agency.

For service evaluation, governance dimensions operating at delivery organisation level (workforce stability, person-centred planning processes, safeguarding effectiveness) complement national governance indicators by capturing the process quality that directly affects user experience. This proves particularly important because governance failures can dissipate resources through corruption, erode trust that reduces service uptake, and perpetuate inequalities when powerful groups capture benefits intended for disadvantaged populations.

4.2 Addressing Systematic Gaps: The Undervalued Dimensions

Empirical analysis reveals that several dimensions critical to wellbeing and central to social service provision appear infrequently across Beyond-GDP indicators and fail to cluster consistently with the four major pillars. This systematic neglect requires explicit acknowledgement and deliberate correction.

Funded by the European Union
Project number: 101179032

Table 3: Undervalued Dimensions - Coverage Gap Analysis

Dimension	Coverage in 66 Indicators	Why Undervalued	Framework Response
Care Services	13 (20%)	Devaluation of feminised reproductive labour; measurement difficulty; productivist bias	Dedicated attention across framework dimensions
Subjective Wellbeing & Experience	22 (33%)	Scepticism about subjective validity; limited policy relevance perception	Life satisfaction; user experience measures
Culture & Meaning	15 (23%)	Difficulty quantifying; instrumentalist frameworks	Quality consideration; respect for diversity
Leisure & Time Use	18 (27%)	Productivist bias; focus on paid work	Work-life balance; person-centred scheduling
Service Integration	<7 (11%)	Siloed measurement; sector-specific frameworks	Continuity of care; cross-service coordination
Prevention	~4 (6%)	Success = non-events; long time horizons	Repeat crisis tracking; preventive logic

Care services appear in only 20% of indicators despite their essential role for children, persons with disabilities, frail elderly persons, and individuals with chronic illness. This invisibility reflects historical devaluation of feminised reproductive labour, measurement difficulty when care operates informally through family arrangements, and productivist frameworks prioritising market production over care provisioning. The consequences prove severe: inadequate measurement renders substantial welfare state activity invisible, perpetuates gender inequalities by ignoring women's disproportionate care responsibilities, and prevents evidence-informed policy regarding care financing and quality standards. Measurement approaches exist (time use surveys, service infrastructure indicators, carer wellbeing surveys, user outcome measures such as the Adult Social Care Outcomes Toolkit) but remain peripheral to mainstream frameworks.

Subjective wellbeing and experience quality appear in only 33% of indicators, reflecting persistent scepticism about subjective measurement validity despite growing evidence of predictive validity and stability. This neglect proves particularly problematic for social services given their inherently relational nature. Services addressing complex needs operate through ongoing professional relationships where trust, communication quality, and procedural fairness fundamentally shape effectiveness. Process values including respect, dignity, autonomy, and participation matter independently of

Funded by the European Union
Project number: 101179032

measured outcomes, particularly when services involve involuntary intervention such as child protection or mental health treatment. Comprehensive frameworks should incorporate life satisfaction as ultimate outcome measure, dignity and respect in service delivery through patient-reported experience measures, and user voice in service design through person-centred planning.

Culture and meaning appear in only 23% of indicators, reflecting difficulty quantifying cultural vitality and instrumentalist frameworks treating culture as input to creativity rather than intrinsic dimension of flourishing. For social services, cultural dimensions prove relevant through culturally appropriate provision (respecting linguistic diversity, religious practices, dietary requirements), cultural participation sustaining identity for isolated persons, and community cultural assets strengthening informal care arrangements. Whilst comprehensive frameworks need not position culture as distinct pillar, cultural considerations should inform quality assessment.

Leisure and time use appear in 27% of indicators, typically through time poverty measures assessing work hours rather than leisure quality. Time use proves fundamental to care services both as outcome and constraint. Service users value control over daily schedules. Service effectiveness depends partly on adequate time for care workers to perform tasks without excessive pressure. Family carers face time poverty balancing care responsibilities with employment, creating demands for respite services. Temporal accessibility (service opening hours, scheduling flexibility) determines whether time-constrained individuals can access provision.

The systematic undervaluation of these domains creates distorted policy incentives prioritising measurable dimensions over unmeasured values. Service systems optimise measured health outcomes whilst neglecting care quality affecting daily dignity. Economic policy emphasises employment maximisation without adequate attention to time poverty constraining care provision. Comprehensive frameworks address these distortions through deliberate inclusion of undervalued dimensions, explicit acknowledgement of measurement gaps, and commitment to continued indicator development.

4.3 Methodological Pluralism for Service Evaluation

The framework integrates multiple evaluation traditions, recognising that no single method adequately captures service complexity. Economic evaluation methods (cost-benefit analysis, cost-effectiveness analysis, social return on investment) enable efficiency assessment and resource allocation decisions but struggle with non-market value capture and equity integration. Experimental and quasi-experimental methods (randomised controlled trials, difference-in-differences, regression discontinuity) provide causal rigour for attribution but face feasibility constraints and ethical

Funded by the European Union
Project number: 101179032

concerns with withholding services. Participatory approaches (Most Significant Change, Outcome Harvesting, participatory scorecards) centre lived experience and user voice but resist aggregation across contexts. Standardised frameworks (Adult Social Care Outcomes Toolkit, National TOMs) enable benchmarking but may not capture local priorities.

Comprehensive service evaluation combines these approaches: economic methods for value-for-money assessment, causal identification for attribution, participatory processes for user voice, and standardised indicators for comparability. This methodological pluralism, detailed further in Section 3, achieves comprehensiveness unattainable through isolated methods.

4.4 Framework Design Principles for Social Service Evaluation

Based on the empirical analysis of 66 Beyond-GDP indicators and the identified systematic gaps, comprehensive social service evaluation frameworks should adhere to several design principles emerging from international measurement experience.

Multi-dimensional scope proves essential. Single-domain frameworks whether focused exclusively on health outcomes, economic efficiency, or user satisfaction provide incomplete assessment. Effective frameworks integrate at minimum four evaluation dimensions. Access and adequacy dimensions capture coverage rates, waiting times, geographic reach, affordability barriers, and unmet need. Outcome and experience dimensions assess changes in user circumstances across quality of life, housing stability, employment participation, social connection, subjective wellbeing, and family carer impacts. Quality and governance dimensions evaluate workforce stability, person-centred approaches, service continuity, safety, complaints responsiveness, transparency, safeguarding effectiveness, and user voice. Sustainability dimensions examine financial viability, capital adequacy, workforce supply, environmental footprint, and resilience under shock.

Equity integration must move beyond aggregate assessment to systematic disaggregation. The finding that only 12% of Beyond-GDP indicators enable intersectional analysis demonstrates that current practice inadequately reveals how disadvantages compound for multiply marginalised groups. Methodologies for gender-sensitive indicator development exist (Beck, 1999; Gardner, 2016; OXFAM, 2014; Lin et al., 2007) but remain poorly integrated into mainstream wellbeing frameworks. The Gender Equality Index (EIGE, 2013) and Global Gender Gap Index (WEF, 2006) demonstrate feasibility of comprehensive gender assessment, yet most Beyond-GDP indicators treat gender as optional disaggregation rather than central analytical dimension. Service evaluation frameworks should routinely disaggregate by gender, age, disability status, migration background, geographic location, and socioeconomic

Funded by the European Union
Project number: 101179032

position, with intersectional analysis combining characteristics to identify groups facing distinctive barriers or outcomes. Small area estimation, targeted oversampling, and privacy-preserving publication methods enable granular equity assessment whilst protecting confidentiality.

Temporal responsiveness requires innovation beyond annual reporting cycles. The 18-month median publication lag characterising Beyond-GDP indicators undermines policy responsiveness and accountability. Service frameworks should exploit administrative data for quarterly or monthly monitoring where feasible, employ nowcasting techniques predicting recent outcomes from partial data, develop leading indicators anticipating future performance, and implement real-time user feedback mechanisms. Whilst comprehensive wellbeing assessment appropriately employs annual cycles, service management benefits from higher-frequency signals enabling rapid course correction.

User voice and experience quality warrant dedicated measurement beyond objective outcomes. The systematic underrepresentation of subjective wellbeing (33% coverage) and virtual absence of patient-reported experience measures indicates that current frameworks inadequately capture dimensions mattering most to service users including dignity, respect, autonomy, and procedural fairness. Service evaluation should incorporate life satisfaction as ultimate outcome measure, experience quality assessment through validated instruments, user involvement in service design captured through person-centred planning metrics, and complaints mechanisms as accountability signals when experience proves unsatisfactory.

Prevention and early intervention require measurement approaches valuing problems averted. The finding that fewer than 6% of indicators assess prevention reflects measurement bias favouring reactive interventions over proactive strategies. Frameworks should track repeat crises indicating prevention failure, employ counterfactual methods demonstrating problems averted through early intervention, conduct long-term follow-up matching prevention payback periods, and utilise simulation modelling projecting intervention impacts when empirical data require extended timeframes. Economic research on skill formation and early intervention demonstrates substantial long-term returns to preventive investment in human development (Heckman, 2006; Cunha & Heckman, 2007), though life-course perspectives reveal measurement challenges spanning decades (Elder, 1998).

Care services demand visibility commensurate with their welfare state importance. The 20% coverage of care domains despite their centrality to wellbeing for vulnerable populations demonstrates systematic neglect requiring deliberate correction. Comprehensive frameworks should assess childcare availability and quality, eldercare coverage and adequacy, disability support provision, informal care

Funded by the European Union
Project number: 101179032

burden through time use and strain measures, and care work economic valuation recognising reproductive labour contributions.

Service integration and coordination warrant dedicated attention. The finding that fewer than 11% of indicators assess integration reflects siloed measurement traditions. Frameworks should evaluate continuity of care for individuals receiving support from multiple providers, care coordination effectiveness, information sharing across organisational boundaries, and multi-disciplinary collaboration supporting holistic responses to complex needs.

Methodological pluralism proves superior to single-method approaches. No single measurement tradition whether economic evaluation, experimental causal identification, participatory assessment, or standardised frameworks adequately captures service complexity. Effective evaluation combines economic methods for efficiency assessment, causal identification strategies for attribution, participatory approaches centring lived experience, and standardised indicators enabling comparability, with integration achieving comprehensiveness unattainable through isolated methods.

These design principles derived from international Beyond-GDP experience provide guidance for developing operational measurement systems adapted to specific service contexts, governance structures, and policy objectives. Application of these principles to European social services requires context-specific framework development, with detailed indicator specifications, data collection protocols, and implementation guidance constituting work streams beyond this meta-analysis scope.

4.5 Implications for BENEFITS Work Packages

The findings from this meta-analysis provide essential foundation for subsequent BENEFITS work packages, informing methodological development, pilot design, and knowledge dissemination across the project lifecycle.

The four-pillar framework architecture derived from co-occurrence analysis provides conceptual grounding for the WP2 European Green Book on social services evaluation (D2.1). The identified systematic gaps in care services (20% coverage), service integration (11%), and prevention (6%) establish priority domains requiring enhanced measurement attention. The documented 18-month median publication lag and limited intersectional analysis capacity (12%) inform recommendations for timeliness improvements and equity-focused disaggregation in European social service assessment.

Funded by the European Union

Project number: 101179032

The 22-domain taxonomy presented in Annex III provides operational definitions for the WP2 microsimulation model specification (D2.2), ensuring alignment between indicator categories and simulation parameters. The analysis of methodological patterns across Beyond-GDP indicators, including weighting approaches, aggregation methods, and data source combinations, informs microsimulation model architecture. The finding that 89% of indicators rely on administrative data whilst 62% incorporate survey data guides decisions regarding input data requirements and validation approaches for simulating social service value added.

The synthesis of evaluation methodologies in Section 3.17, encompassing economic evaluation methods, experimental and quasi-experimental approaches, and participatory methods, provides methodological foundation for WP3 pilot design across partner countries (D3.1, D3.2, D3.3). The identified strengths and limitations of each approach inform selection of appropriate methods for different pilot contexts. The framework design principles derived from international experience guide pilot protocol development, ensuring pilots test measurement innovations addressing the systematic gaps identified in this analysis whilst maintaining methodological rigour suitable for subsequent scaling.

The complete inventory of 66 indicators presented in Annex II and the domain co-occurrence matrix in Annex IV provide content architecture for the WP4 Social Services Policy Observatory Platform (D4.1, D4.2), enabling systematic organisation of indicator documentation, comparative assessment tools, and policy guidance materials. The analysis of institutional distribution informing stakeholder engagement strategies and content curation priorities. The dashboard versus composite index distinction guides platform design decisions regarding presentation formats appropriate for different user needs and policy contexts.

This meta-analysis thus establishes the empirical and conceptual foundation upon which subsequent BENEFITS work packages build, ensuring coherent framework development grounded in systematic assessment of international measurement experience.

Funded by the European Union

Project number: 101179032

5 CONCLUSION

5.1 Summary

This meta-analysis of 66 Beyond-GDP indicators spanning five decades reveals both substantial progress in comprehensive wellbeing measurement and persistent blind spots constraining application to social service evaluation. Health (92% coverage), knowledge and skills (89%), and material wellbeing (85%) demonstrate sophisticated measurement with robust international infrastructure. Environmental sustainability achieves moderate integration (62%), whilst governance and institutional quality appear in 47% of frameworks, reflecting growing recognition of their importance.

However, domains central to social service provision remain severely underrepresented. Care services appear in only 20% of indicators despite their essential role for vulnerable populations. Service integration features in fewer than 11% of frameworks, and prevention in approximately 6%. Whilst 56% of indicators enable basic gender disaggregation, only 12% support intersectional analysis examining compounding disadvantages. Publication lags averaging 18 months limit policy responsiveness. Composite indices (52%) facilitate communication but obscure component performance critical for targeted improvement.

These gaps reflect structural biases: devaluation of feminised reproductive labour, productivist frameworks privileging market production over care provisioning, and measurement conventions evolved for macroeconomic assessment rather than service evaluation.

The proposed framework addresses identified gaps through empirically grounded architecture. The four-pillar structure (economic foundations, social wellbeing, environmental sustainability, governance and institutions) derives from co-occurrence analysis revealing robust domain clustering (33-67% co-occurrence rates) rather than imposed theoretical commitments. This data-driven approach ensures alignment with established practice whilst deliberately incorporating undervalued dimensions (care services, subjective wellbeing, culture, leisure) through explicit acknowledgement rather than forced integration.

The conceptual framework spanning access and adequacy, outcomes and experience, quality and governance, and sustainability provides comprehensive architecture for service assessment. This multi-dimensional structure addresses the finding that single-domain frameworks provide incomplete evaluation, enabling assessment of coverage gaps, user circumstance changes, process quality, and long-term system viability.

Funded by the European Union
Project number: 101179032

Domain co-occurrence patterns validate pillar coherence whilst revealing systematic neglect requiring correction. Care services' weak integration with established clusters (15-27% co-occurrence) confirms that addressing this gap demands deliberate framework design rather than relying on organic emergence through existing measurement traditions.

5.2 Implications for Social Service Evaluation

The meta-analysis demonstrates that existing Beyond-GDP frameworks, despite breadth compared with GDP, inadequately address dimensions central to social service evaluation. This creates several challenges. First, resource allocation decisions informed by incomplete measures perpetuate underinvestment in unmeasured dimensions. Second, performance management systems optimising measured indicators whilst ignoring unmeasured dimensions create perverse incentives. Third, policy evaluation assessing economic impacts whilst neglecting service experiences provides distorted evidence regarding intervention effectiveness.

The identified gaps suggest that comprehensive service evaluation requires frameworks explicitly incorporating care provision, service coordination, prevention, user experience quality, and intersectional equity analysis. Methodological patterns indicate that no single approach (composite indices, dashboards, surveys, adjusted GDP measures) proves comprehensively adequate. Effective evaluation demands methodological pluralism combining economic methods for efficiency assessment, causal identification for attribution, participatory approaches for user voice, and standardised indicators for comparability.

Temporal patterns reveal acceleration in framework development post-2010 (65% of sample), suggesting momentum in Beyond-GDP measurement. However, geographic concentration in OECD contexts (89% of national frameworks) raises questions about applicability to diverse social service contexts globally, as frameworks developed in high-income settings may not adequately address challenges in contexts where service provision models, informal arrangements, and resource constraints differ substantially.

5.3 Limitations and Future Research

Several limitations constrain this analysis. First, restriction to English-language documentation may exclude innovative approaches from non-Anglophone contexts, particularly from the Global South where alternative progress conceptualisations may differ substantially from OECD frameworks. Future work should systematically review non-English literature and engage Southern perspectives ensuring universal applicability.

Funded by the European Union
Project number: 101179032

Second, variation in documentation quality limited component-level analysis uniformity, with approximately 30% of indicators lacking complete public inventories. This asymmetry may introduce bias towards better-documented frameworks from well-resourced international organisations. Third, the rapid evolution of the Beyond-GDP field means very recent developments post-November 2024 may not be captured. Periodic updates prove necessary for currency, ideally on three to five year cycles.

Fourth, criterion 4 required coverage of at least three distinct domains to qualify as comprehensive progress frameworks. Relaxing this threshold to two domains would increase the sample by approximately 20-25 indicators, yielding N≈90. However, sensitivity analysis suggests this would marginally increase care services coverage from 20% to 22% whilst substantially increasing environmental sub-domains (air quality 39%→58%, water 27%→43%). The stricter three-domain threshold better operationalises the research question regarding comprehensive frameworks whilst acknowledging that specialised two-domain frameworks serve complementary purposes.

Fifth, whilst the framework proposes comprehensive conceptual architecture, operational specifications including precise definitions, calculation algorithms, data sources, and quality criteria require continued development through pilot implementation. The transition from conceptual framework to operational measurement system demands substantial technical work beyond this meta-analysis scope.

Future research should pursue several directions. First, cost-effectiveness evidence regarding different measurement approaches remains limited. Research comparing costs and benefits of alternative measurement strategies would inform resource allocation decisions. Second, the framework does not fully address attribution challenges distinguishing service contributions from confounding influences on outcomes. Continued methodological development integrating causal identification strategies within routine monitoring systems proves essential.

Third, real-time measurement exploiting digital service channels could dramatically reduce publication lags enabling responsive management. Fourth, participatory measurement engaging service users in outcome definition, data collection, and interpretation could strengthen relevance whilst respecting agency. Fifth, longitudinal measurement tracking individuals across services and time could enable causal analysis and life-course assessment. Sixth, spatial measurement at granular geographic scales could reveal service deserts. Seventh, intersectional measurement examining compounding disadvantages across multiple characteristics simultaneously could advance equity. Eighth, care work valuation through time use, economic equivalents, and wellbeing impacts could address systematic devaluation.

Funded by the European Union
Project number: 101179032

5.4 Significance

This analysis provides a systematic assessment of how Beyond-GDP frameworks address social services, quantifying coverage gaps and proposing empirically grounded architecture for correction. The contribution proves timely as societies confront demographic ageing, migration, climate change, and technological transformation increasing social service centrality to wellbeing and social cohesion.

The Beyond-GDP movement has established that comprehensive wellbeing assessment proves both conceptually sound and practically feasible. The 66 indicators reviewed demonstrate sophisticated measurement spanning health, education, material living standards, environmental sustainability, and governance quality. However, systematic blind spots persist regarding social services constituting major public investments directly targeting wellbeing improvement. OECD countries spend an average of 1.7% of GDP on long-term care services, yet measurement of care service quality and outcomes remains underdeveloped (Colombo, Llena-Nozal, Mercier, & Tjadens, 2011). Workforce challenges including high turnover and low staffing levels significantly affect care quality (Bostick et al., 2006; Spilsbury et al., 2011; Hussein & Manthorpe, 2012).

Success in developing comprehensive frameworks would demonstrate feasibility of comprehensive measurement for concrete policy applications, provide template for other service domains, and strengthen the broader Beyond-GDP project by moving from abstract advocacy towards operational integration. European statistical capacity, research excellence, and political commitment to social rights create favourable conditions for pioneering social service evaluation frameworks.

Ultimately, measurement reform proves necessary but insufficient for transformative change. Indicators illuminate problems, reveal inequalities, and track progress, but do not automatically generate political will addressing deficiencies. The value of comprehensive measurement lies in strengthening democratic deliberation through transparent evidence regarding collective wellbeing and its distribution. When measurement systems reflect what societies genuinely value rather than only what proves easily quantified, they enable more informed choices regarding resource allocation and institutional design affecting human flourishing.

This analysis contributes by synthesising international evidence from 66 Beyond-GDP indicators, identifying systematic gaps in social service coverage through quantified analysis, proposing coherent architecture addressing deficiencies through empirically derived pillar structure, and establishing foundation for operational framework development in subsequent BENEFITS project phases.

Funded by the European Union
Project number: 101179032

6 BIBLIOGRAPHY

Aaberge, R., & Langørgen, A. (2006). Measuring the benefits from public services: The effects of local government spending on the distribution of income in Norway. *Review of Income and Wealth, 52*(1), 61–83. https://doi.org/10.1111/j.1475-4991.2006.00176.x

Arrow, K. J., Dasgupta, P., Goulder, L. H., Mumford, K. J., & Oleson, K. (2012). Sustainability and the measurement of wealth. *Environment and Development Economics, 17*(3), 317–353. https://doi.org/10.1017/S1355770X12000137

Beck, T. (1999). *Using gender-sensitive indicators: A reference manual for governments and other stakeholders* (Gender Management System Series). Commonwealth Secretariat.

Benczúr, P., Boskovic, A., Cariboni, J., Chevallier, R., Le Blanc, J., Sandor, A., & Zec, S. (2024). *Sustainable and inclusive wellbeing: The road forward*. Publications Office of the European Union. https://doi.org/10.2760/83967

Benczúr, P., Boskovic, A., Giovannini, E., Pagano, A., & Sandor, A.-M. (2025). *Measuring sustainable and inclusive wellbeing: A multidimensional dashboard approach*. Publications Office of the European Union. https://data.europa.eu/doi/10.2760/4186342

Biggeri, M., Ferrannini, A., Lodi, L., Cammeo, J., & Francescutto, A. (2023). The "winds of change": The SPES framework on sustainable human development. *SPES Working Paper*.

Bostick, J. E., Rantz, M. J., Flesner, M. K., & Riggs, C. J. (2006). Systematic review of studies of staffing and quality in nursing homes. *Journal of the American Medical Directors Association, 7*(6), 366–376. https://doi.org/10.1016/j.jamda.2006.01.024

Bragg, R., & Atkins, G. (2016). *A review of nature-based interventions for mental health care* (Natural England Commissioned Report No. 204). Natural England. https://publications.naturalengland.org.uk/publication/4513819616346112

Card, D., Kluve, J., & Weber, A. (2018). What works? A meta-analysis of recent active labor market program evaluations. *Journal of the European Economic Association, 16*(3), 894–931. https://doi.org/10.1093/jeea/jvx028

Cobb, C., Halstead, T., & Rowe, J. (1995). *The genuine progress indicator: Summary of data and methodology*. Redefining Progress.

Colombo, F., Llena-Nozal, A., Mercier, J., & Tjadens, F. (2011). *Help wanted? Providing and paying for long-term care* (OECD Health Policy Studies). OECD Publishing. https://doi.org/10.1787/9789264097759-en

Corvo, L., Pastore, L., Mastrodascio, M., & Cepiku, D. (2022). The social return on investment model: A systematic literature review. *Meditari Accountancy Research, 30*(1), 49–86. https://doi.org/10.1108/MEDAR-05-2021-1307

Costanza, R., Fioramonti, L., & Kubiszewski, I. (2016). The UN Sustainable Development Goals and the dynamics of well-being. *Frontiers in Ecology and the Environment, 14*(2), 59. https://doi.org/10.1890/1540-9295-14.2.59

Funded by the European Union

Project number: 101179032

Costanza, R., Hoekstra, R., Eastoe, J., Rum, I., Kubiszewski, I., O'Neill, D., Biggeri, M., Pollio, C., & Zinutti, M. (2024). *A synthesis of models, metrics, and policies for a sustainable and inclusive wellbeing future (Deliverable D1.3)*. MERGE Project. https://mergeproject.eu/synthesis/

Costanza, R., Kubiszewski, I., Giovannini, E., Lovins, H., McGlade, J., Pickett, K. E., Ragnarsdóttir, K. V., Roberts, D., De Vogli, R., & Wilkinson, R. (2014). Time to leave GDP behind. *Nature, 505*(7483), 283–285. https://doi.org/10.1038/505283a

Cummins, R. A., Eckersley, R., Pallant, J., van Vugt, J., & Misajon, R. (2003). Developing a national index of subjective wellbeing: The Australian Unity Wellbeing Index. *Social Indicators Research, 64*(2), 159–190. https://doi.org/10.1023/A:1024704320683

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review, 97*(2), 31–47. https://doi.org/10.1257/aer.97.2.31

Elder, G. H. (1998). The life course as developmental theory. *Child Development, 69*(1), 1–12. https://doi.org/10.2307/1132065

European Commission. (2009). *GDP and beyond: Measuring progress in a changing world* (COM(2009) 433 final). European Commission.

European Commission. (2023). *Beyond GDP: Delivering sustainable and inclusive wellbeing*. Joint Research Centre. https://joint-research-centre.ec.europa.eu/projects-and-activities/beyond-gdp-delivering-sustainable-and-inclusive-wellbeing_en

European Institute for Gender Equality. (2013). *Gender Equality Index*. https://eige.europa.eu/gender-equality-index

Färe, R., Grosskopf, S., & Hernandez-Sancho, F. (2004). Environmental performance: An index number approach. *Resource and Energy Economics, 26*(4), 343–352. https://doi.org/10.1016/j.reseneeco.2003.10.003

Felice, E., & Vasta, M. (2015). Passive modernization? The new human development index and its components in Italy's regions (1871–2007). *European Review of Economic History, 19*(1), 44–66. https://doi.org/10.1093/ereh/heu018

Fioramonti, L. (2013). *Gross domestic problem: The politics behind the world's most powerful number*. Zed Books.

Fleurbaey, M., & Blanchet, D. (2013). *Beyond GDP: Measuring welfare and assessing sustainability*. Oxford University Press.

Foster, J. E., McGillivray, M., & Seth, S. (2013). Composite indices: Rank robustness, statistical association, and redundancy. *Econometric Reviews, 32*(1), 35–56. https://doi.org/10.1080/07474938.2012.690641

Foster, J., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica, 52*(3), 761–766. https://doi.org/10.2307/1913475

Gardner, J. (2016). *Using gender statistics*. UNECE.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research, 141*(1), 61–94. https://doi.org/10.1007/s11205-017-1832-9

Funded by the European Union

Project number: 101179032

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(5782), 1900–1902. https://doi.org/10.1126/science.1128898

Hoekstra, R. (2019). *Replacing GDP by 2030: Towards a common language for the well-being and sustainability community*. Cambridge University Press.

Hsu, J., Flores, G., Evans, D., Mills, A., & Hanson, K. (2018). Measuring financial protection against catastrophic health expenditures: Methodological challenges for global monitoring. *International Journal for Equity in Health, 17*, 69. https://doi.org/10.1186/s12939-018-0749-5

Hussein, S., & Manthorpe, J. (2012). The dementia social care workforce in England: Secondary analysis of a national workforce dataset. *Aging & Mental Health, 16*(1), 110–118. https://doi.org/10.1080/13607863.2011.596808

Imbert, C., & Papp, J. (2015). Labor market effects of social programs: Evidence from India's employment guarantee. *American Economic Journal: Applied Economics, 7*(2), 233–263. https://doi.org/10.1257/app.20130401

International Labour Organization. (2012). *Decent work indicators: Concepts and definitions*. ILO.

International Labour Organization. (2018). *Care work and care jobs for the future of decent work*. ILO.

International Labour Organization. (2022). *Care at work: Investing in care leave and services for a more gender equal world of work*. ILO. https://doi.org/10.54394/LRRN6830

International Labour Organization. (2024a). *World social protection report 2024–26: Universal social protection for climate action and a just transition*. ILO. https://doi.org/10.54394/ZMDK5543

International Labour Organization. (2024b). *The impact of care responsibilities on women's labour force participation*. ILO. https://doi.org/10.54394/LPTT5569

Jansen, A., Hoekstra, R., Kaufmann, R., & Gerer, A. (2023). *A synthesis of Beyond-GDP metrics for wellbeing, inclusion, and sustainability including a deep dive into EU metrics and their role in governance (WISE Horizon 2020 Project #101095219).* European Commission Joint Research Centre.

Jansen, A., Wang, R., Behrens, P., & Hoekstra, R. (2024). Beyond GDP: A review and conceptual framework for measuring sustainable and inclusive wellbeing. *The Lancet Planetary Health, 8*(9), e695–e705. https://doi.org/10.1016/S2542-5196(24)00178-4

Johnston, D. F. (1985). The development of social statistics and indicators on the status of women. *Social Indicators Research, 16*(3), 233–261. https://doi.org/10.1007/BF00415125

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives, 20*(1), 3–24. https://doi.org/10.1257/089533006776526030

Funded by the European Union

Project number: 101179032

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The Worldwide Governance Indicators: Methodology and analytical issues. *Hague Journal on the Rule of Law, 3*(2), 220–246. https://doi.org/10.1017/S1876404511200046

Klasen, S. (2007). Gender-related indicators of well-being. In M. McGillivray (Ed.), *Human well-being: Concept and measurement* (pp. 167–192). Palgrave Macmillan. https://doi.org/10.1057/9780230625600_7

Klugman, J., Hanmer, L., Twigg, S., Hasan, T., McCleary-Sills, J., & Santamaria, J. (2014). *Voice and agency: Empowering women and girls for shared prosperity*. World Bank Publications. https://doi.org/10.1596/978-1-4648-0359-8

Kluve, J., Puerto, S., Stoeterau, J., Weidenkaff, F., Witte, M., & Robalino, D. (2014). Interventions to improve labour market outcomes of youth: A systematic review of active labour market programmes. *Campbell Systematic Reviews, 10*(1), 1–109. https://doi.org/10.1002/cl2.123

Kubiszewski, I., Costanza, R., Eastoe, J., Lu, T., Mulder, K., Patteson Hernandez, G., Benczúr, P., & Dixson-Declève, S. (2025). Building consensus on societal wellbeing: A semantic synthesis of indicators to move beyond GDP. *Ecological Indicators, 178*, 114076. https://doi.org/10.1016/j.ecolind.2025.114076

Kuznets, S. (1934). *National income, 1929–1932: A report to the U.S. Senate, 73rd Congress, 2nd Session*. U.S. Government Printing Office.

Lawn, P. A. (2005). An assessment of the valuation methods used to calculate the index of sustainable economic welfare (ISEW), genuine progress indicator (GPI), and sustainable net benefit index (SNBI). *Environment, Development and Sustainability, 7*(2), 185–208. https://doi.org/10.1007/s10668-005-7312-4

Legatum Institute. (2007). The Legatum Prosperity Index. Legatum Institute. https://www.prosperity.com/

Lin, V., Gruszin, S., Ellickson, C., Glover, J., Silburn, K., Wilson, G., & Poljski, C. (2007). Comparative evaluation of indicators for gender equity and health. *International Journal of Public Health, 52*(Suppl. 1), S19–S26. https://doi.org/10.1007/s00038-006-6050-1

Marmot, M. (2020). Health equity in England: The Marmot review 10 years on. *BMJ, 368*, m693. https://doi.org/10.1136/bmj.m693

Marques, S. R., Rodrigues, R., Zerth, J., & Orrego, C. (2025). The use of social return on investment approaches to evaluate integrated long-term care in high-income countries: A scoping review. *Health Policy, 161*, 105414. https://doi.org/10.1016/j.healthpol.2025.105414

Mazziotta, M., & Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one? *Rivista Italiana di Economia, Demografia e Statistica, 67*(2), 67–80.

Michalos, A., Smale, B., Labonté, R., Muharjarine, N., Scott, K., Moore, K., Swystun, L., Holden, B., Bernardin, H., Dunning, B., Graham, P., Guhn, M., Gadermann, A. M., Zumbo, B. D., Morgan, A., Brooker, A.-S., & Hyman, I. (2010). *The Canadian Index of Wellbeing: Technical paper — CIW 1.0*. Canadian Index of Wellbeing, University of Waterloo.

Funded by the European Union

Project number: 101179032

National Institute for Health and Care Excellence. (2013). *Guide to the methods of technology appraisal 2013*. NICE.
Netten, A., Burge, P., Malley, J., Potoglou, D., Towers, A.-M., Brazier, J., Flynn, T., Forder, J., & Wall, B. (2012). Outcomes of social care for adults: Developing a preference-weighted measure. *Health Technology Assessment, 16*(16), 1–166. https://doi.org/10.3310/hta16160
New Economics Foundation. (2006). *Happy Planet Index*. New Economics Foundation.
New Economics Foundation. (2006). *The Happy Planet Index: An index of human well-being and environmental impact*. New Economics Foundation.
Nordhaus, W., & Tobin, J. (1972). Is growth obsolete? In *Economic research: Retrospect and prospect* (Vol. 5, pp. 1–80). NBER.
OECD. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD. https://doi.org/10.1787/9789264043466-en
OECD. (2013). *OECD guidelines on measuring subjective well-being*. OECD. https://doi.org/10.1787/9789264191655-en
OECD. (2018). *The future of education and skills: Education 2030*. OECD Publishing.
Ouyang, Z., Song, C., Zheng, H., Polasky, S., Xiao, Y. [Yi], Bateman, I. J., Liu, J., Ruckelshaus, M., Shi, F., Xiao, Y. [Yang], Xu, W., Zou, Z., & Daily, G. C. (2020). Using gross ecosystem product (GEP) to value nature in decision making. *Proceedings of the National Academy of Sciences, 117*(25), 14593–14601. https://doi.org/10.1073/pnas.1911439117
Oxfam International. (2014). *Quick guide to gender-sensitive indicators*. Oxfam International.
Prados de la Escosura, L. (2021). Augmented human development in the age of globalization. *The Economic History Review, 74*(4), 946–975. https://doi.org/10.1111/ehr.13064
Raworth, K. (2012). *A safe and just space for humanity* (Discussion Paper). Oxfam.
Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., … Foley, J. A. (2009). A safe operating space for humanity. *Nature, 461*(7263), 472–475. https://doi.org/10.1038/461472a
Rum, I., Hoekstra, R., Behrens, P., & Jansen, A. (2024). A semantic taxonomy of wellbeing and sustainability indicators (MERGE Project Working Paper). European Commission Joint Research Centre.
Shahidi, F. V., Ramraj, C., Sod-Erdene, O., Hildebrand, V., & Siddiqi, A. (2019). The impact of social assistance programs on population health: A systematic review of research in high-income countries. *BMC Public Health, 19*, 2. https://doi.org/10.1186/s12889-018-6337-1
Shelter Scotland, & Engender. (2024). *Gender and the housing emergency*. Shelter Scotland.

Funded by the European Union
Project number: 101179032

Shields, P., Roy, R., Khan, S., Read, A., & Ottaway, S. (2025). Social value—are we measuring up? *Public Money & Management*, 1–6. https://doi.org/10.1080/09540962.2025.2539355

Social Progress Imperative. (2013). *Social Progress Index*. Social Progress Imperative. https://www.socialprogress.org/

Social Value UK. (2020). *National themes, outcomes and measures (TOMs): Social value measurement*. Social Value UK.

Spilsbury, K., Hewitt, C., Stirk, L., & Bowman, C. (2011). The relationship between nurse staffing and quality of care in nursing homes: A systematic review. *International Journal of Nursing Studies, 48*(6), 732–750. https://doi.org/10.1016/j.ijnurstu.2011.02.014

Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. OECD.

Stockhammer, E., Hochreiter, H., Obermayr, B., & Steiner, K. (1997). The index of sustainable economic welfare (ISEW) as an alternative to GDP in measuring economic welfare: The results of the Austrian (revised) ISEW calculation 1955–1992. *Ecological Economics, 21*(1), 19–34. https://doi.org/10.1016/S0921-8009(96)00088-2

Terzi, A. (2021). *Economic policy-making beyond GDP: An introduction* (European Economy Discussion Papers No. 142). European Commission.

UN Women. (2016a). *Redistributing unpaid care and sustaining quality care services: A prerequisite for gender equality* (UN Women Policy Briefs No. 5). https://doi.org/10.18356/40becb1b-en

UNESCO. (2024). *UNESCO Institute for Statistics database*. Retrieved 6 October, 2025, from http://uis.unesco.org

United Nations. (1997). *Human development report 1997*. UNDP.

United Nations. (2013). *City Prosperity Index*. UN-Habitat.

United Nations. (2020). *Human development report 2020*. UNDP.

United Nations. (2023). *Inclusive wealth report 2023: Measuring sustainability and equity*. UN Environment Programme. https://doi.org/10.59117/20.500.11822/43131

Ura, K., Alkire, S., & Zangmo, T. (2012). *GNH and GNH Index: A short guide to Gross National Happiness Index*. The Centre for Bhutan Studies.

Veenhoven, R. (1996). Happy life-expectancy. *Social Indicators Research, 39*(1), 1–58. https://doi.org/10.1007/BF00300831

Veenhoven, R. (2014). Happiness adjusted life years. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 2641–2643). Springer. https://doi.org/10.1007/978-94-007-0753-5_1225

Verbist, G., Förster, M., & Vaalavuo, M. (2012). The impact of publicly provided services on the distribution of resources: Review of new results and methods (OECD Social, Employment and Migration Working Papers No. 130). OECD. https://doi.org/10.1787/5k9h363c5szq-en

World Bank. (2019). *World development report 2019: The changing nature of work*. World Bank. https://doi.org/10.1596/978-1-4648-1328-3

Funded by the European Union

Project number: 101179032

World Commission on Environment and Development. (1987). *Our common future*. Oxford University Press.
World Economic Forum. (2001). *Environmental Sustainability Index*. World Economic Forum.
World Economic Forum. (2006). *The Global Gender Gap Report 2006*. World Economic Forum.
World Health Organization. (2024). *Global Health Observatory data repository*. Retrieved 6 October, 2025, from https://www.who.int/data/gho

Funded by the European Union
Project number: 101179032

7 ANNEX I: PRISMA Diagram

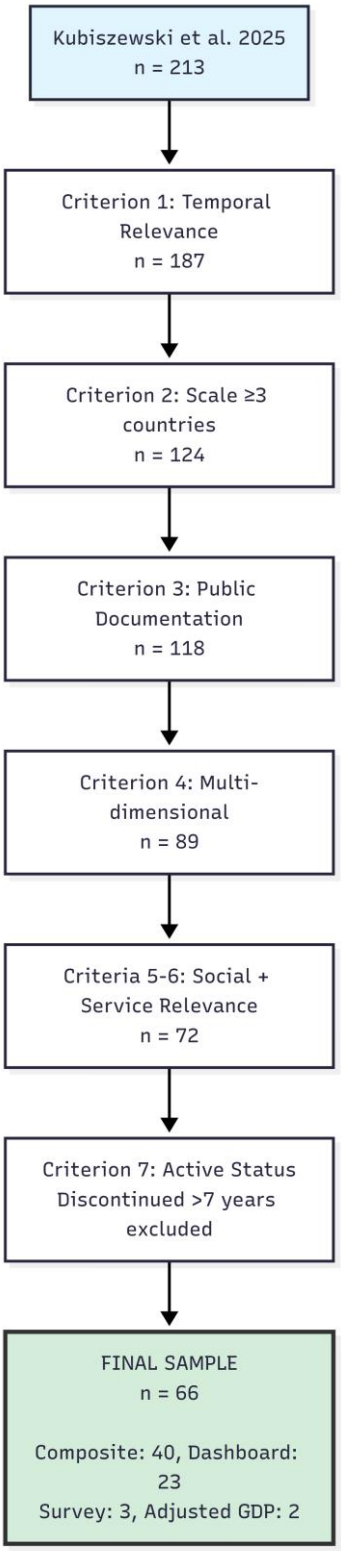


Figure 4: PRISMA Diagram

Funded by the European Union
Project number: 101179032

8 ANNEX II: COMPLETE INVENTORY OF 66 INCLUDED INDICATORS

Table 4: Inventory of Included Indicators

#	Indicator Name	Developer Institution	Type	Year	Geographic Scope	Compon. (N)	Update Freq
1	Genuine Progress Indicator (GPI)	Academic (Redefining Progress)	Adjusted GDP	1995	National (USA, adapted globally)	27	Irregular
2	Human Development Index (HDI)	UN Development Programme	Composite	1990	Global (191 countries)	4	Annual
3	Inequality-adjusted HDI (IHDI)	UN Development Programme	Composite	2010	Global (151 countries)	5	Annual
4	OECD Better Life Index (BLI)	OECD	Dashboard	2011	OECD members (38 countries)	24	Biennial
5	Inclusive Wealth Index (IWI)	UN Environment Programme	Composite	2012	Global (140+ countries)	17	Irregular (4-year)
6	Happy Planet Index	New Economics Foundation (NGO)	Composite	2006	Global (140+ countries)	3	Irregular
7	Index of Sustainable Economic Welfare (ISEW)	Academic (Daly & Cobb)	Adjusted GDP	1989	National (multiple adaptations)	8	Irregular
8	Legatum Prosperity Index	Legatum Institute (NGO)	Composite	2007	Global (167 countries)	67	Annual
9	Canadian Index of Wellbeing (CIW)	University of Waterloo	Dashboard	2011	National (Canada)	62	Biennial
10	City Prosperity Index (CPI)	UN-Habitat	Composite	2012	Global cities (400+)	62	Irregular
11	Sustainable Wellbeing Index (SWI)	Academic (multiple)	Composite	2010	National (various)	3	Irregular
12	Genuine Savings/Adjusted Net Savings (ANS)	World Bank	Adjusted GDP	1999	Global (140+ countries)	7	Annual
13	Gallup-Sharecare Wellbeing Index	Gallup (Business)	Survey	2008	National (USA)	42	Continuous

Funded by the European Union
Project number: 101179032

#	Indicator Name	Developer Institution	Type	Year	Geographic Scope	Compon. (N)	Update Freq
14	AARP Liveability Index	AARP (NGO)	Composite	2015	National (USA communities)	59	Annual
15	National Wellbeing: Life Satisfaction	UK Office for National Statistics	Dashboard	2010	National (UK)	6	Quarterly
16	Regional Quality of Life Index	Eurostat	Composite	2015	EU NUTS-2 regions (263)	17	Biennial
17	Australian Unity Wellbeing Index	Australian Centre on Quality of Life	Survey	2001	National (Australia)	16	Biannual
18	Social Progress Index	Social Progress Imperative (NGO)	Composite	2013	Global (169 countries)	60	Annual
19	Multidimensional Poverty Index (MPI)	UNDP & Oxford Poverty & Human Development Initiative	Composite	2010	Global (111 countries)	10	Annual
20	Global Competitiveness Index	World Economic Forum	Composite	2004	Global (141 countries)	12	Annual
21	Gender Development Index (GDI)	UN Development Programme	Composite	1995	Global (167 countries)	5	Annual
22	Positive Peace Index (PPI)	Institute for Economics & Peace	Composite	2009	Global (163 countries)	24	Biennial
23	European Quality of Life Survey	Eurofound	Survey	2003	EU28 + candidate countries	12	4-year cycle
24	Quality of Life Index	Numbeo (Business)	Composite	2009	Global cities (continuous)	26	Continuous
25	Bloomberg Healthiest Countries Index	Bloomberg (Business)	Composite	2019	Global (169 countries)	10	Irregular
26	Living Standards Framework (LSF)	New Zealand Treasury	Dashboard	2011	National (New Zealand)	21	Annual
27	Thriving Places Index (TPI)	Local Trust (UK NGO)	Composite	2017	National (England localities)	22	Annual

Funded by the European Union
Project number: 101179032

#	Indicator Name	Developer Institution	Type	Year	Geographic Scope	Compon. (N)	Update Freq
28	The Good Country Index	Good Country (NGO)	Composite	2014	Global (169 countries)	14	Irregular
29	European Social Survey (ESS)	European Research Infrastructure	Survey	2002	European (30+ countries)	19	Biennial
30	Mercer Quality of Living	Mercer (Business)	Composite	1999	Global cities (230+)	39	Annual
31	Global Liveability Index	Economist Intelligence Unit	Composite	2002	Global cities (173)	33	Annual
32	Human Sustainable Development Index	Academic (Togtokh & Gaffney)	Composite	2010	Global (151 countries)	9	Irregular
33	Gender Inequality Index (GII)	UN Development Programme	Composite	2010	Global (162 countries)	10	Annual
34	World Values Survey (WVS)	World Values Survey Association	Survey	1981	Global (120+ countries)	295	5-year waves
35	European Values Study (EVS)	EVS Foundation	Survey	1981	European (47+ countries)	70	9-year waves
36	Human Poverty Index (HPI)	UN Development Programme	Composite	1997	Global (replaced by MPI 2010)	5	Annual
37	Gross National Happiness (GNH) Index	Government of Bhutan	Composite	1972	National (Bhutan)	33	5-year
38	Inclusive Development Index	World Economic Forum	Composite	2017	Global (103 countries)	9	Annual
39	Sustainable Development Goals Index (SDGI)	Sustainable Development Solutions Network	Composite	2016	Global (193 countries)	97	Annual
40	UK Measures of National Wellbeing	UK Office for National Statistics	Dashboard	2010	National (UK)	46	Annual
41	Measuring What Matters Framework	Statistics Canada	Dashboard	2021	National (Canada)	81	Annual
42	Measuring Ireland's Progress	Central Statistics Office Ireland	Dashboard	2003	National (Ireland)	43	Annual

Funded by the European Union

Project number: 101179032

#	Indicator Name	Developer Institution	Type	Year	Geographic Scope	Compon. (N)	Update Freq
43	Finland's Findicators	Statistics Finland	Dashboard	2009	National (Finland)	83	Annual
44	Israel Wellbeing Indicators	Israel Central Bureau of Statistics	Dashboard	2015	National (Israel)	19	Annual
45	Quality of Life Indicators in Slovenia	Statistical Office Slovenia	Dashboard	2015	National (Slovenia)	16	Annual
46	Quality of Life Indicators in Korea	Statistics Korea	Dashboard	2011	National (South Korea)	8	Annual
47	Italy BES (Equitable & Sustainable Wellbeing)	Istat (Italian National Statistical Institute)	Dashboard	2010	National (Italy)	12	Annual
48	Belgium Sustainable Development Indicators	Federal Planning Bureau Belgium	Dashboard	2000	National (Belgium)	7	Annual
49	Chile Social Wellbeing Survey	Ministry of Social Development Chile	Survey	2015	National (Chile)	48	Biennial
50	Canada's Quality of Life Framework	Statistics Canada	Dashboard	2021	National (Canada)	8	Annual
51	European Social Progress Index	European Commission	Composite	2016	EU regions (272 NUTS-2)	11	Biennial
52	Quality of Life in Norway	Statistics Norway	Dashboard	2006	National (Norway)	8	Annual
53	Quality of Life Indicators in Spain	Spanish National Statistics Institute	Dashboard	2013	National (Spain)	37	Annual
54	Iceland Wellbeing Indicators	Statistics Iceland	Dashboard	2019	National (Iceland)	17	Annual
55	Indicators Aotearoa New Zealand	Stats NZ & Treasury	Dashboard	2018	National (New Zealand)	43	Annual
56	Switzerland MONET 2030	Swiss Federal Statistical Office	Dashboard	2003	National (Switzerland)	98	Annual
57	Netherlands Monitor of Wellbeing	Statistics Netherlands	Dashboard	2013	National (Netherlands)	21	Annual

Funded by the European Union
Project number: 101179032

#	Indicator Name	Developer Institution	Type	Year	Geographic Scope	Compon. (N)	Update Freq
58	Wellbeing in Germany	Federal Statistical Office Germany	Dashboard	2016	National (Germany)	14	Biennial
59	Doughnut Economics	Academic (Kate Raworth)	Dashboard	2017	City-level applications	Variable	Variable
60	Comprehensive Wealth (World Bank)	World Bank	Composite	2021	Global (146 countries)	17	Irregular
61	Ecological Footprint	Global Footprint Network	Composite	1992	Global (190+ countries)	6	Annual
62	Environmental Performance Index (EPI)	Yale & Columbia Universities	Composite	2006	Global (180 countries)	40	Biennial
63	Country Policy & Institutional Assessment (CPIA)	World Bank	Composite	1977	Low-income countries (78)	50	Annual
64	Ibrahim Index of African Governance (IIAG)	Mo Ibrahim Foundation	Composite	2007	African countries (54)	42	Annual
65	Global Gender Gap Index	World Economic Forum	Composite	2006	Global (146 countries)	16	Annual
66	World Happiness Report	UN Sustainable Development Solutions Network	Composite	2012	Global (143 countries)	10	Annual

Notes:

- Six frameworks discontinued more than seven years before 2024 were excluded: Environmental Sustainability Index (2005), Green GDP China (2007), National Accounts of Wellbeing (2009), Where-to-be-born Index (2013), Measures of Australia's Progress (2014), and Sustainable Society Index (2016)
- Human Poverty Index retained as direct methodological predecessor to Multidimensional Poverty Index, illustrating framework evolution
- Component counts reflect distinct variables or sub-indicators included in each framework
- Gross National Happiness Index was operationalised much later in 2008

Funded by the European Union
Project number: 101179032

9 ANNEX III: 22-DOMAIN TAXONOMY WITH DEFINITIONS AND EXAMPLES

Table 5: Domain Taxonomy and Definitions

Domain	Definition	Example Components	Typical Data Sources
1. Health	Physical and mental health outcomes, healthcare access, and health system performance	Life expectancy at birth; mortality rates (infant, child, maternal); healthy life years; morbidity prevalence (communicable and non-communicable diseases); mental health status; healthcare coverage; access to healthcare services; health expenditure	Vital statistics; health surveys; administrative health records; WHO databases
2. Knowledge & Skills	Educational attainment, literacy, cognitive skills, and lifelong learning capacity	Years of schooling (mean and expected); literacy and numeracy rates; school enrolment and completion; student performance (PISA, TIMSS); tertiary attainment; adult education participation; research capacity (R&D personnel)	Education administrative data; international assessments; household surveys; UNESCO statistics
3. Material Wellbeing	Income, consumption, wealth, and material living standards	GDP per capita; household income (mean, median, disposable); consumption expenditure; wealth accumulation; poverty rates (absolute and relative); Gini coefficient; financial assets	National accounts; household budget surveys; income and wealth surveys; tax records
4. Work & Job Quality	Employment conditions, job security, work satisfaction, and work-life balance	Employment rate; unemployment rate; underemployment; job satisfaction; working conditions; work-life balance; workplace safety; skills	Labor force surveys; enterprise surveys; working conditions surveys; ILO statistics

Funded by the European Union
Project number: 101179032

Domain	Definition	Example Components	Typical Data Sources
		utilisation; labour productivity; precarious employment	
5. Economic Security	Financial resilience, social protection coverage, and vulnerability to economic shocks	Social protection coverage; unemployment benefits adequacy; pension coverage; savings rates; debt levels; insurance coverage; exposure to catastrophic expenditure; economic vulnerability index	Social protection administrative data; household financial surveys; insurance records; ILO social security databases
6. Housing	Housing quality, affordability, tenure security, and adequacy	Housing affordability (price-to-income ratio, rent burden); housing quality (overcrowding, basic amenities); tenure security; homelessness rates; energy efficiency; housing deprivation	Census data; housing surveys; administrative records (social housing); homelessness counts
7. Social Inclusion	Participation in social, economic, cultural, and political life; freedom from discrimination	Employment integration; social participation; community engagement; cultural access; political participation; discrimination experiences; social isolation; civic voice	Social surveys; participation surveys; discrimination reports; civic engagement data
8. Social Equity	Distribution of resources, opportunities, and outcomes across population groups	Income inequality (Gini, Palma, income share ratios); wealth inequality; equality of opportunity; intergenerational mobility; gender pay gap; regional disparities; distributional progressivity of services	Income/wealth distribution data; mobility studies; administrative data with demographic detail
9. Environmental Sustainability	Natural resource use, pollution, biodiversity, and ecological integrity	GHG emissions; renewable energy share; resource productivity; waste generation; biodiversity loss; ecosystem health; natural capital depletion; environmental footprint	Environmental accounts; emissions inventories; biodiversity monitoring; resource extraction data

Funded by the European Union
Project number: 101179032

Domain	Definition	Example Components	Typical Data Sources
10. Institutions	Governance quality, rule of law, corruption control, and institutional effectiveness	Rule of law; control of corruption; government effectiveness; regulatory quality; voice and accountability; institutional trust; civil service capacity	Governance indicators (WGI, CPIA); perception surveys; institutional assessments; transparency indices
11. Safety	Personal security, crime, violence, and freedom from harm	Crime rates (violent, property); victimisation; violence against women and children; homicide rate; perceptions of safety; traffic accidents; workplace injuries	Police statistics; victimisation surveys; health records (injury data); traffic accident data
12. Trust & Social Connections	Social capital, trust in others and institutions, and relationship quality	Generalised trust; trust in institutions (government, police, courts); social network size; social support; loneliness; community belonging; reciprocity norms	Social capital surveys; European Social Survey; World Values Survey; national wellbeing surveys
13. Subjective Wellbeing	Self-reported life satisfaction, happiness, and sense of meaning	Life satisfaction (Cantril ladder); experienced happiness and positive affect; absence of negative affect; sense of meaning and purpose; domain satisfactions (health, work, relationships)	Wellbeing surveys; European Quality of Life Survey; Gallup World Poll; national life satisfaction modules
14. Care Services	Provision and accessibility of childcare, eldercare, disability support, and informal care	Childcare enrolment and availability; eldercare coverage; home care provision; disability support services; informal care burden; time spent on unpaid care; carer strain	Time use surveys; service administrative data; care sector statistics; household surveys on care arrangements
15. Air Quality	Ambient air pollution levels and exposure	PM2.5 and PM10 concentrations; NO2 and SO2 levels; ozone concentrations;	Air quality monitoring networks; satellite

Funded by the European Union
Project number: 101179032

Domain	Definition	Example Components	Typical Data Sources
		population exposure to air pollution; air quality index	data; environmental health assessments
16. Water	Water access, quality, and resource sustainability	Access to safe drinking water; sanitation coverage; water quality (pollutant levels); water stress; freshwater availability; wastewater treatment coverage	Water service providers; water quality monitoring; FAO AQUASTAT; WHO/UNICEF JMP
17. Climate	Climate change impacts, mitigation, and adaptation	GHG emissions (total, per capita, intensity); renewable energy share; climate vulnerability; extreme weather events; temperature changes; climate adaptation measures	National GHG inventories; UNFCCC reporting; climate vulnerability indices; meteorological data
18. Land & Ecosystem	Land use, biodiversity, ecosystem services, and natural habitats	Forest cover; protected area coverage; biodiversity indicators (species threat status); land degradation; ecosystem service flows; habitat fragmentation	Land use surveys; biodiversity monitoring; protected area databases; ecosystem assessments
19. Energy Resources	Energy access, efficiency, and transition to renewables	Access to electricity and clean cooking fuels; energy intensity; renewable energy share; energy security; energy poverty; energy efficiency	Energy balances; IEA statistics; household energy surveys; utility data
20. Physical Capital	Infrastructure quality and access to built environment	Transport infrastructure; telecommunications access; public facilities; housing stock; utilities coverage; infrastructure quality	Infrastructure inventories; national accounts (fixed capital formation); access surveys
21. Culture	Cultural participation,	Cultural participation rates; arts access; heritage	Cultural participation

Funded by the European Union
Project number: 101179032

Domain	Definition	Example Components	Typical Data Sources
	heritage, and cultural vitality	preservation; cultural diversity; creative sector employment; cultural identity	surveys; administrative data (museums, theatres); cultural sector statistics
22. Leisure	Time use, leisure quality, and work-life balance	Time spent on leisure; work hours; commuting time; time poverty; leisure satisfaction; vacation time; temporal autonomy	Time use surveys; labour force surveys; employer records (working time)

- **Note:** Domain boundaries are analytical constructs; many real-world phenomena span multiple domains (e.g., mental health intersects Health, Subjective Wellbeing, and Social Connections).

Funded by the European Union

Project number: 101179032

10 ANNEX IV: DOMAIN CO-OCCURRENCE MATRIX

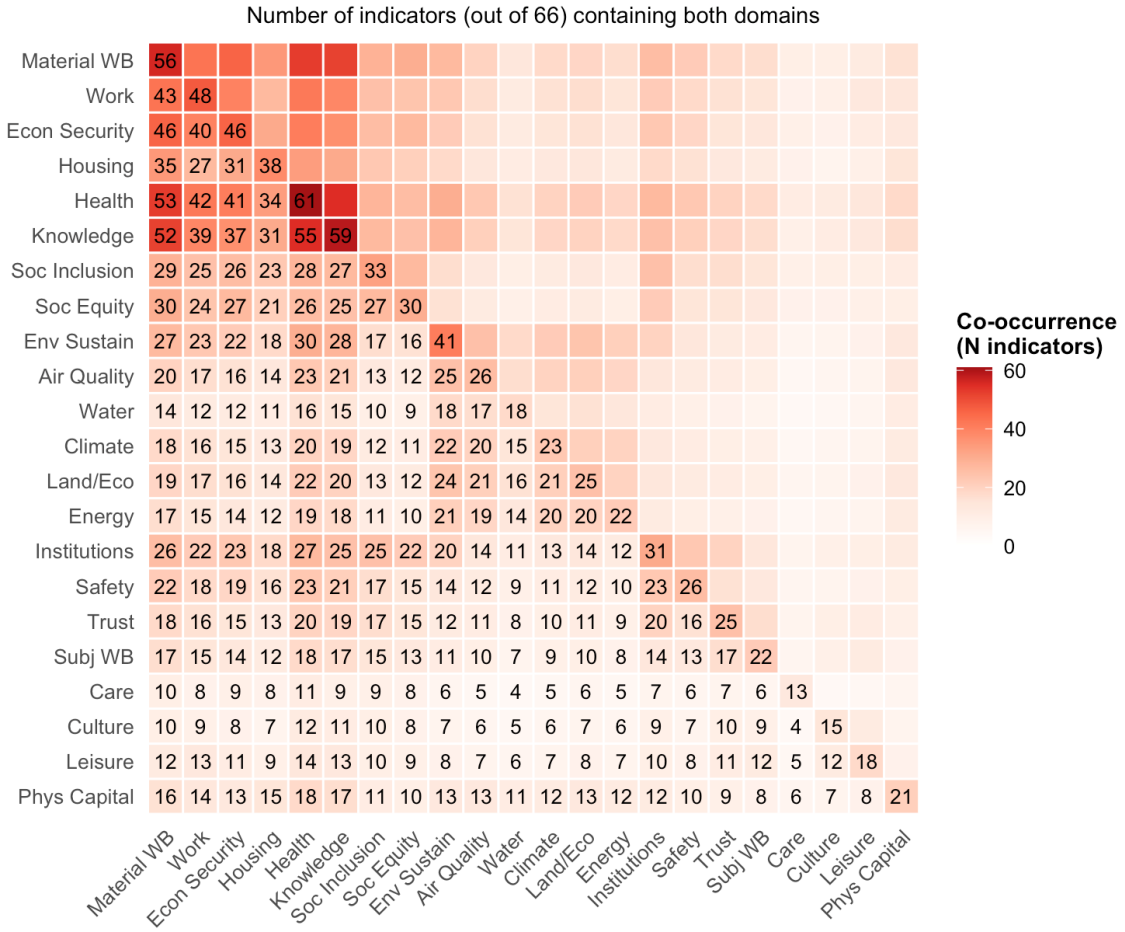


Figure 5: Domain Co-occurrence Matrix

Note: Due to size constraints, presenting upper triangle of symmetric matrix with co-occurrence counts (number of indicators containing both domains)

Key Insights:

- **Highest co-occurrence**: Health + Knowledge (55), Material WB + Econ Security (46), Health + Material WB (53)
- **Economic cluster coherence**: Material WB, Work, Econ Security show 40-46 pairwise co-occurrence (61-70%)
- **Environmental fragmentation**: Environmental domains show 17-25 co-occurrence with each other
- **Care services isolation**: Care shows only 4-11 co-occurrence with other domains, confirming systematic neglect
- **Governance moderate**: Institutions shows 20-27 co-occurrence with social domains

Funded by the European Union
Project number: 101179032

11 ANNEX V: Domain Co-occurrence Table

Table 6: Domain Co-occurrence Table

Domain Pair	Co-occurrence (N)	Co-occurrence (%)	Pillar Assignment
ECONOMIC FOUNDATIONS			
Material Wellbeing + Work Quality	43	65%	Economic
Material Wellbeing + Economic Security	46	70%	Economic
Economic Security + Work Quality	40	61%	Economic
Material Wellbeing + Housing	35	53%	Economic
SOCIAL WELLBEING			
Health + Material Wellbeing	53	80%	Social/Economic overlap
Health + Knowledge & Skills	55	83%	Social
Material Wellbeing + Knowledge & Skills	52	79%	Social/Economic overlap
Health + Housing	34	52%	Social
ENVIRONMENTAL SUSTAINABILITY			
Environmental Sustainability + Climate	22	33%	Environmental
Environmental Sustainability + Air Quality	25	38%	Environmental
Environmental Sustainability + Energy	21	32%	Environmental
Air Quality + Climate	20	30%	Environmental
GOVERNANCE & INSTITUTIONS			
Social Inclusion + Social Equity	27	41%	Governance
Social Inclusion + Institutions	25	38%	Governance
Institutions + Safety	23	35%	Governance
Social Equity + Institutions	22	33%	Governance
UNDERVALUED DIMENSIONS			
Care + Health	11	17%	Orphaned
Care + Social Inclusion	9	14%	Orphaned
Subjective WB + Health	18	27%	Isolated
Culture + Leisure	12	18%	Peripheral

Funded by the European Union
Project number: 101179032

BENEFITS

Building Economic, Needs-Based and Environmental evaluation Frameworks for Inclusive Transformation of Social services in Europe

Funded by the European Union