SWINS

Sustainable Well-being through INvestment in Social Services

D4.4

The potential use of alternative data sources to estimate the returns of social services

Mapping alternative data sources, empirical applications, and research potential in the SWINS partner countries

September 2026

Funded by the European Union

Funded by European Union's Horizon Europe Programme under Grant Agreement No. 101177566

Authors

Annaelena Valentini – ARCO of PIN foundation

Federico Ciani - ARCO of PIN foundation

Irene Fattacciu - ARCO of PIN foundation

Arianna Vivoli - ARCO of PIN foundation

Peer Reviewers

Floore Bursens, University of Antwerp; Toon Van Hevere, University of Antwerp; Gerlinde Verbist, University of Antwerp

Cite as

Valentini, A., Ciani, F., Fattacciu, I., Vivoli, A., (2026), 'The potential use of alternative data sources to estimate the returns of social services - Mapping alternative data sources, empirical applications, and research potential in the SWINS partner countries", ARCO, research centre at PIN Foundation. Available at: www.swins.eu

Acknowledgements

The authors would like to thank all SWINS partners for the valuable feedback provided during the Budapest meeting. The authors are also grateful to Lisa Braito for her helpful advice and suggestions and to the peer reviewers for their constructive comments, which helped improve the paper.

Disclaimer

This deliverable is part of Task 4.5 "Surveying potential macro indicators". This task will cover two main areas. The first is using administrative data to study the economic and social returns of investment in social services. Administrative data have potential advantages, such as more complete information on individual service use, larger sample size, and longer timespan for following changes in individual status. The flipside of administrative data is the lack of cross-country comparability. The second area will deal with the potentialities of complex and unconventional data sources (including the so-called Big Data that will be at the centre of T6.3). The focus will be on how the information sources can be repurposed for the development of innovative approaches to the measurement of the return on investment in social services..

Funded by the European Union under Grant Agreement No.101177566. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Agency. Neither the European Union nor the granting authority can be held responsible for them.

Table of Contents

1.Introduction 5
2.Why look for alternative data sources? Limitations of existing sources for social investment evaluation .. 8
2.1 Existing data sources 8
2.2 Main limitations 8
3.A taxonomy of alternative data sources for social investment evaluation 9
3.1 Process-mediated data 10
3.2 Human-sourced data 10
3.3 Machine-generated and sensor data 11
3.4 Comparative summary 11
3.5 What alternative data can add to social-investment evaluation 12
3.6 Search scope and method 12
4.Alternative data in practice: Care 13
4.1 Process-mediated data 13
4.2 Human-sourced data 13
4.3 Machine-generated data 14
5.Alternative data in practice: ALMP and life-long learning 15
5.1 Process-mediated data 15
5.2 Human-sourced data 16
5.3 Machine-generated and sensor data 17
6.Alternative data in practice: Health 18
6.1 Process-mediated data 18
6.2 Human-sourced digital traces 19
6.3 Machine-generated data 19
6.4 Looking ahead: the European Health Data Space 20
7.Cross-country mapping for the SWINS consortium 21
Belgium 24
Germany 24
Hungary 25
Ireland 25
Italy 26
Norway 26
Serbia 26

Spain 27
8.Research potential and limitations 27
8.1 What can realistically be measured? 27
8.2 Cross-country comparability 29
8.3 Access, privacy and representativeness 30
9.Conclusions 30
References 33
Evidence note 39

1.Introduction

Governments across Europe invest substantial public resources in social services, such as early childhood education and care (ECEC), long-term care, active labour market policies (ALMP) and healthcare. Yet measuring the returns to those investments is inherently difficult, since many other factors such as demographic change or the economic cycle, influence economic and social outcomes and the effects of social investment often materialise only in the long term (IWGSI, 2024). The challenge is therefore one of identification and one of measurement at once: the relevant outcomes often unfold across domains and over periods that conventional evaluation datasets cannot observe in a single individual or household.

The methodologies available to estimate such returns split along two axes: ex-ante versus ex-post evaluation, and micro-level versus macro-level analysis (IWGSI, 2024). At the micro level, ex-ante approaches rely primarily on microsimulation models, of which EUROMOD is the standard EU instance: they apply tax-benefit rules to individual- or household-level records to simulate the effects of existing or hypothetical policies on income, poverty, and inequality. Ex-post micro approaches, including randomised controlled trials (RCT) and quasi-experimental designs such as propensity score matching, regression discontinuity, and difference-in-differences, can establish causality with relatively high robustness for a specific intervention, but their results are highly context-dependent and difficult to generalise beyond the setting in which they were estimated. This limitation is particularly relevant when the object of evaluation is not a discrete intervention but a system-level reform, for which the external validity of RCTs and other quasi-experimental approaches may be challenged by several factors. First, pilot interventions may be implemented under conditions that are not representative of those prevailing after scaling-up: services delivered within pilots often receive greater attention, dedicated resources and closer monitoring than their post-scale counterparts, potentially generating effects that are difficult to reproduce at system level. Second, the performance of both treated and untreated units may be shaped heterogeneously by institutional, socioeconomic and contextual factors, making it more difficult to construct a credible counterfactual when the conditions affecting the control group differ from those affecting the treatment group. Third, system-level reforms may generate second-round or indirect effects, including spillovers, behavioural responses and interactions with other policies and services, which may either amplify or attenuate the direct effects of the intervention and are difficult to capture within designs centred on a clearly defined treatment and control group.

At the macro level, ex-post approaches typically rely on fixed-effects panel regressions, often combined with System GMM estimators, relating country-year social expenditure to outcomes such as economic growth or inequality reduction. These approaches can account for observed and unobserved country-level factors but generally do not identify causality or the specific channels through which an effect operates. Part of the difficulty is that expenditure itself responds to outcomes, since governments tend to raise social spending when poverty or unemployment increase. Macro-level associations also average over micro units that may respond to the same policy in distinctly different ways, so that an aggregate association can conceal heterogeneous, and even opposite, effects across groups. Attempts to bridge the micro and macro levels directly - combining microsimulation with Social Accounting Matrices and computable general equilibrium models - remain the exception and face a structural trade-off between distributional detail and general-equilibrium effects (IWGSI, 2024).

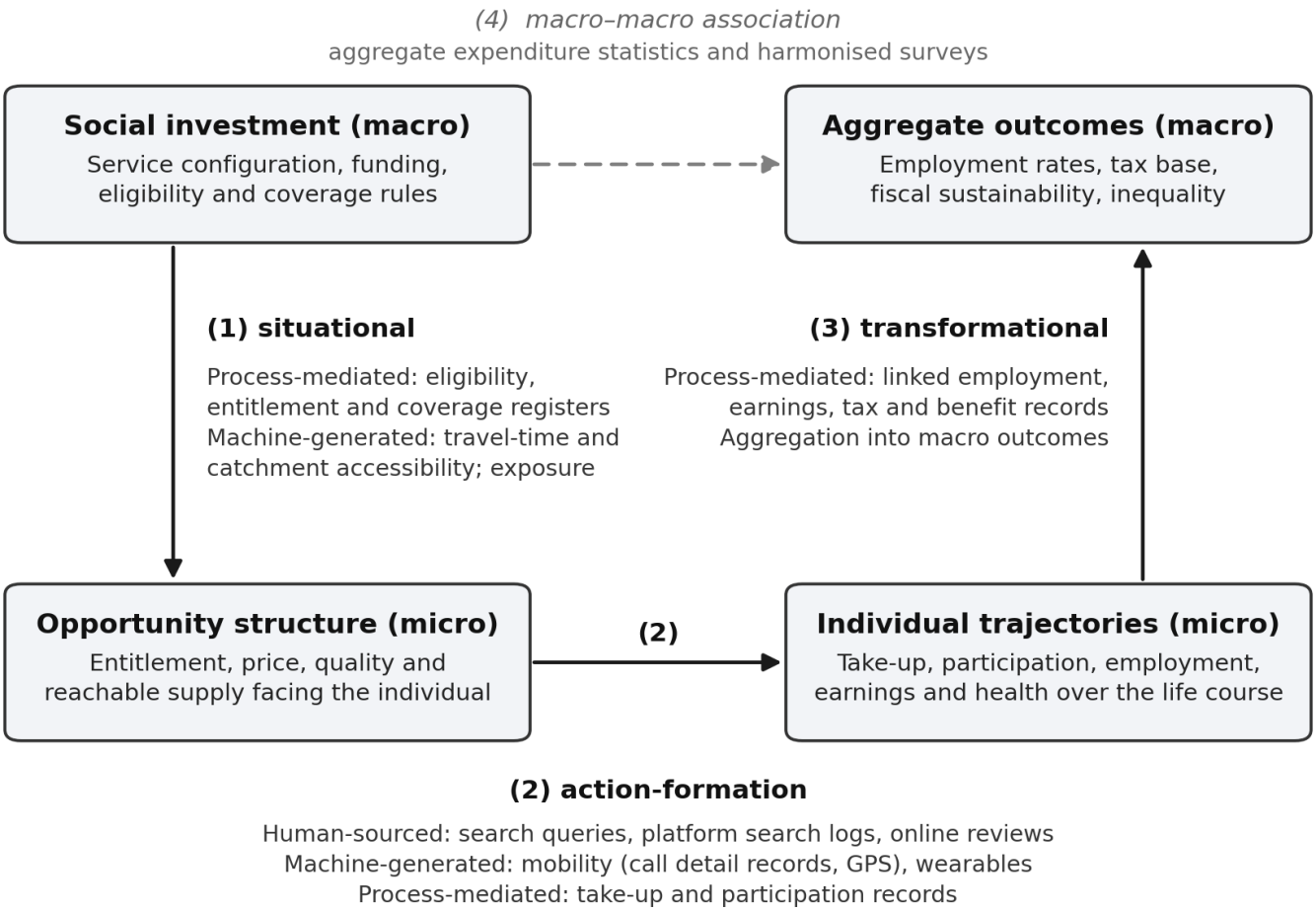
Standard evaluation frameworks in the EU social-policy toolkit therefore remain better suited to immediate distributional effects than to the medium- and long-run returns of in-kind social services. The SWINS project is designed to address this gap, with the objective of developing a framework for measuring the economic, social, and environmental returns generated by investments in social policies and services, with explicit attention to life-course dynamics and feedback mechanisms between micro-level behaviour and macro-level outcomes (Ciani and Fattacciu, 2025).

More precisely, SWINS approaches this problem through the analytical scaffold of Coleman's Boat (Coleman, 1990), which conceptualises social investment as operating across three connected mechanisms (Hedström and Ylikoski, 2010) rather than as a direct macro-to-macro relationship. First, situational mechanisms describe how macro-level service configurations reshape the incentives, constraints and opportunities individuals face (for example, how affordable ECEC shifts parental labour supply, or how eligibility rules shape take-up of a service). Second, action-formation mechanisms describe how individuals act within this altered environment, forming trajectories that accumulate across the life course through what is usually referred to as the life-course multiplier (Garritzmann et al., 2021): an intervention at one stage of life propagates through subsequent stages

and domains. Third, transformational mechanisms describe how the aggregation of these individual trajectories generates macro-level outcomes such as employment rates, tax bases and fiscal sustainability, which in turn reshape the institutional and economic conditions for the next cycle of investment.

This macro-micro-macro architecture has a direct methodological implication for the present paper. Each link in Coleman's Boat imposes different requirements on the data needed to observe it, requirements that in some cases traditional data sources cannot meet. The situational link, connecting policy configurations to individual opportunity structures, and the transformational link, connecting individual behaviour to aggregate outcomes, are what long-run administrative registers are, in principle, best equipped to document: population coverage, service participation, and downstream economic outcomes recorded as a matter of administrative routine. In practice this capacity is uneven: in several countries, participation in services such as ECEC or long-term care is recorded only at aggregate or provider level, or not at all where provision is private, so that coverage and participation cannot be observed at the individual level that the analysis requires. The action-formation link, by contrast, is where individuals actually convert opportunities into decisions and trajectories, a process that unfolds through everyday behaviour, mobility, service use and exposure that administrative records often do not capture at the relevant frequency or granularity. This is where big data sources have the greatest potential to complement registers, by making visible the micro-level mechanisms through which macro-level social investment is converted into individual, and ultimately collective, returns. Figure 1 summarises this architecture and indicates, for each link, the data families best placed to observe it.

Figure 1: Coleman's Boat and the data required to observe each link



Source: authors' elaboration on Coleman (1990) and Hedström and Ylikoski (2010).

This paper contributes to that framework by mapping the alternative data sources that can complement traditional data sources, such as surveys and aggregate statistics, in observing each of these links. The mapping is organised around a taxonomy proposed by UNECE (2013), which classifies sources according to how the data are generated. Within the UNECE taxonomy, three families are distinguished. (1) Process-mediated data, of which administrative and register-based sources are the principal instance, are generated as by-products of public administration, service delivery or transactions: social-security and tax records, benefit and childcare registers, employment-service records, hospital discharge data, prescriptions and health-insurance claims, as well as private-sector transaction records. These are the primary candidates for documenting the situational and transformational links, because they record participation and outcomes as a matter of administrative routine and at population scale. (2) Human-sourced digital traces (e.g., internet searches, online reviews, user-generated content and citizen reporting) and (3) machine-generated data, such as mobile-phone and GPS records, environmental and traffic sensors, wearables and satellite observations, are instead generated as by-products of individual behaviour or of automated observation. They are therefore the more promising candidates for the action-formation link, where opportunities are converted into decisions, movements and exposures. These three families are not, however, used in isolation. As the review will show, most of the strongest existing evidence on social-investment returns comes from designs that join sources across families: health records to employment histories, childcare registers to education and tax records, and what can be observed depends more often on the identifiers, legal bases and governance arrangements that make such joins possible than on any individual source.

Throughout this paper, “alternative” is defined relative to the data on which the literature on returns to social investment has conventionally relied, namely aggregate expenditure statistics and harmonised household surveys, and not relative to empirical social science in general. Two contributions that an alternative source can make should also be kept apart. A source can improve measurement, by observing a quantity that conventional sources miss or record with error, such as accessibility, service quality, environmental exposure or the timing of a shock; or it can support identification, by providing the variation, the sample size or the follow-up needed to estimate the causal effect of an investment.

This empirical review is a scoping exercise and focuses on the three SWINS policy domains: care (and in particular ECEC), ALMP and life-long learning and health. Within this set, greater attention is given to health, because it is the domain least well covered by traditional international and national surveys. By contrast, ECEC and ALMP are more often covered by conventional surveys, administrative data and established evaluation infrastructures. Particular attention is also given to the eight SWINS consortium partner countries: Italy, Spain, Belgium, Norway, Ireland, Hungary, Serbia and Germany. Beyond their membership of the consortium, these countries are analytically useful because they span both the main European welfare regimes and a wide range of data-infrastructure maturity: a Nordic register tradition (Norway), continental social-insurance systems with centralised social-security data warehouses (Belgium, Germany), Southern European systems with strong regional or sectoral infrastructures (Italy, Spain), a liberal welfare state (Ireland), a Central-Eastern European multi-domain administrative database (Hungary) and an EU candidate country (Serbia). However, where particularly relevant studies from other countries were identified, they were also included in the review. The mapping is deliberately conservative: a source is described as available or usable only where an official data infrastructure or a published study provides evidence for that claim. Because the search was purposive and not exhaustive, the statements of absence in this report should be read as gaps we were unable to fill and not as demonstrations that no such study exists.

The paper is structured as follows. Section 2 reviews the limitations of conventional data sources. Section 3 presents the taxonomy on which the mapping is built and sets out what the joining of sources can add to the observation of social-investment pathways. Sections 4, 5 and 6 review empirical applications in care, ALMP and health respectively, each organised by generative family. Section 7 maps the data infrastructures documented in the eight partner countries. Section 8 discusses research potential and limitations. Section 9 concludes.

2.Why look for alternative data sources? Limitations of existing sources for social investment evaluation

The case for exploring alternative data sources follows from a mismatch between the questions posed by the social investment approach and the data infrastructures on which much of the existing evidence relies. Social investment policies generate returns over long periods, across multiple policy domains, and through interconnected life-course trajectories. Most data sources used in evaluation were designed for other purposes and therefore observe only part of these processes.

2.1 Existing data sources

The empirical literature on returns to social investment relies heavily on aggregate cross-country statistics and household surveys. Aggregate indicators are drawn, for example, from OECD SOCX, Eurostat and the WHO Health for All Database and are typically used in macro-panel analyses relating social expenditure to GDP growth, employment, poverty or inequality. Household surveys include EU-SILC, the Labour Force Survey, Eurobarometer and SHARE. These sources remain indispensable because they provide harmonised information on income, labour-market status, subjective outcomes and other characteristics that are not systematically recorded in administrative systems.

EU-SILC has, for example, been widely used to study the redistributive effects of ECEC and publicly provided services (Verbist, Förster and Vaalavuo, 2012; Aaberge, Langørgen and Lindgren, 2013; Verbist and Matsaganis, 2014). SHARE provides substantially richer health information but focuses on individuals aged 50 and above, limiting its usefulness for interventions targeting earlier stages of the life-course.

2.2 Main limitations

The limitations of these sources can be grouped into three types: limitations of sample, of time, and of content. Traditional sources, such as harmonised surveys, are designed to produce reliable estimates for national populations, not for the specific groups that social investment policies target. Sample sizes can therefore become too small for reliable analysis of vulnerable subgroups (Gábos, Ciani, Tomka and Biggeri, 2021), which is particularly significant because these are the groups for whom returns are most heterogeneous and the policy case strongest. The same design logic constrains spatial resolution: representativeness is typically ensured at national or major-region level, leaving within-country variation in the availability, density and accessibility of services largely invisible. This matters for the SWINS framework in particular, since service configurations translate into territorial functionings at the meso-level, where they enable or constrain individual and collective capabilities (Biggeri and Ferrannini, 2014).

As per the time-related limitation, rotating panel designs limit the observation of life-course returns. EU-SILC typically follows individuals for four consecutive years, while for instance returns to ECEC may take one or more decades to emerge. A second temporal constraint concerns not duration but latency: income information in EU-SILC refers to the year preceding the interview and is released with a further delay, so the resulting series are poorly suited to observing the timing of a shock or the speed of a policy response. This is a significant constraint for assessing the buffer function of social protection, whose value lies precisely in how quickly a

disturbance is absorbed. Retrospective questions offer a partial remedy for duration, but introduce recall and reporting error, particularly for events distant in time or sensitive in nature.

The last constraint has to do with the content. Surveys often capture for example self-reported health but provide limited information on diagnoses, treatment, healthcare utilisation or the precise timing of health events, that are the variables required to trace a health shock into its economic consequences. A related constraint is associated with the value of in-kind services. Household income surveys measure cash income, but the benefit a household derives from an in-kind service, such as a childcare place, does not enter disposable income as conventionally measured. As reported in Van Havere and Verbist (2026), the literature on the redistributive effect of publicly provided services addresses this by imputing a value to those services, typically at average production cost (Marical, d'Ercole, Vaalavuo and Verbist, 2006; Matsaganis and Verbist, 2009; Verbist, Förster and Vaalavuo, 2012). This is a defensible convention for distributional analysis, but it substitutes a spending-based proxy for the quantity of interest: what the service actually delivered to that household. The same substitution occurs at the aggregate level, where expenditure data record how much a government spends on a service, but not the quality, accessibility or intensity of the service received.

3.A taxonomy of alternative data sources for social investment evaluation

If the previous section identified what conventional sources cannot observe, we now turn to what is available. Generally, when we move from traditional to alternative data sources, a general distinction used is the one between administrative and big data. Administrative data are often population-scale, longitudinal and linkable across domains. Big data are often characterised by high frequency, high volume, fine spatial resolution or the ability to observe behaviour without relying on self-report. However, the two should not be treated as competing categories: a health record, for example, can simultaneously be an administrative record and a high-volume digital data stream.

Administrative data are particularly valuable for following individuals over long periods and linking service participation to subsequent outcomes. Examples include childcare participation linked to later employment, ALMP participation linked to subsequent earnings, and health events linked to employment and social-security outcomes. Their main limitations are fragmented institutional ownership, access restrictions, variable linkage quality and poor cross-country comparability because national registers are products of national administrative systems rather than harmonised research instruments. A register also records only what an administration has reason to record: entitlement, take-up and receipt are visible, whereas eligibility that was never exercised, or need that was never presented to the system, generally is not.

Big data add a different comparative advantage. Mobile-phone records can reveal mobility and accessibility at high temporal and spatial resolution; online searches can provide high-frequency indicators of health concerns; wearables can provide continuous measures of physical activity or sleep; online job vacancies can reveal labour demand and skill requirements; satellite and environmental sensors can measure exposures such as temperature, pollution, land cover or accessibility. Beyond this granularity, big data sources also expand the range of outcomes that can be analysed. This is particularly relevant if we assume that returns of social investment are economic, social and civic, contributing to the reduction of inequalities and the reinforcement of social cohesion (Hemerijck, 2017). Big data can therefore help capture outcomes and returns that go beyond the purely economic, shedding light on dimensions that traditional sources struggle to observe. Big data can also contribute to identifying and analysing emerging forms of deprivation that are difficult to capture through conventional socioeconomic indicators. Cooling poverty (Mazzone et al. 2023), for instance, is likely to become increasingly relevant under climate change and illustrates how deprivation can emerge from the interaction between individual characteristics, such as income and health status, and contextual factors, including the quality and energy performance of buildings, infrastructure networks and the urban environment. By combining individual-level information with geospatial, environmental and infrastructural data, big data can help identify

where these overlapping vulnerabilities translate into unequal exposure and unequal capacity to cope with changing environmental conditions. However, these sources may be proprietary, recent, difficult to validate, and unrepresentative of digitally excluded groups. These limitations are particularly important for social investment because disadvantaged populations may be those least likely to generate digital traces.

This contrast between administrative data and big data is a useful first approximation, and it captures something real about comparative advantage. It does not, however, amount to a classification: it is binary where the underlying landscape is not, and one of its two terms is purely residual, since "big data" groups together sources that have little in common beyond not being administrative. Online patient reviews, anonymised call detail records and satellite imagery differ in who produces them, who owns and governs them, what processes generate them, which legal regime governs their reuse, and what form of validation they require before they can support an inference. Treating them as a single category obscures the distinctions that determine whether, and how, a given source can actually be used. The two groups are not competing options, and they are most useful when combined within a single design.

Before classifying sources, however, it is necessary to be explicit about the basis on which they are classified, because two different distinctions are in play in literature. The first is functional: it asks what a source makes observable, which segment of the social-investment pathway, at what frequency, and for which population. The second is generative: it asks how the data came into being, whether as a by-product of an administrative process, of human expression, or of machine observation. These two axes cut across one another. A hospital discharge record is administrative in origin yet constitutes a high-volume digital data stream; mobile-phone records are machine-generated yet held and governed as commercial assets. A review that intends to map sources systematically needs a finer classification, built on a criterion that does not shift with the use to which a source is put.

The taxonomy that we refer to is built on the generative axis, following UNECE (2013) and drawing also on Blazquez and Domenech (2018), where three data types are identified: process-mediated data, human-sourced data and machine-generated data. The reason for this choice is that how a dataset came into being is a stable property of the dataset itself, whereas what it makes observable depends on the question being asked: the same mobility data serve as an accessibility measure in one design and as an exposure measure in another. The functional axis however is not thereby discarded: it is applied on top of the generative categories throughout the empirical review.

3.1 Process-mediated data

Process-mediated data are generated as by-products of public administration, service delivery or transactions. UNECE (2013) distinguishes within this family between data produced by public agencies and data produced by businesses. The first branch is the more important for social-investment evaluation and includes social-security and tax records, benefit registers, education and childcare registers, employment-service records, hospital discharge data, prescriptions and health-insurance claims. The second comprises commercial transaction records (e.g., payments, bookings, e-commerce) which are less central here, but relevant wherever private provision or private expenditure forms part of the service landscape. Access to commercial records of this kind is typically governed by bilateral agreements with the data holder and is therefore neither guaranteed nor stable over time. Process-mediated data are especially valuable when they can be linked through person-, household-, firm- or place-level identifiers.

3.2 Human-sourced data

Human-sourced digital traces include social-network content, internet searches, mobile-app activity, user-generated content, online reviews and citizen reporting. Their main potential is to measure behaviour, perceptions, concerns or service quality at a higher frequency than conventional surveys. Their interpretation requires particular attention to selection, platform-specific populations and changes in platform use over time.

3.3 Machine-generated and sensor data

Machine-generated data arise from automated observation rather than from human reporting or administrative recording and include fixed environmental or traffic sensors, smart meters, GPS-enabled devices, wearables and satellite observations. Because these data are produced by instruments rather than by respondents, they are free of recall and social-desirability bias and are generated continuously, which makes them well suited to outcomes that are spatially structured or that vary at high frequency. Examples relevant to SWINS include air pollution, heat, green space, mobility, physical activity and access to services.

3.4 Comparative summary

Table 1 below provides an overview of the main families of data sources considered in this deliverable, organised according to the UNECE taxonomy. Following the classification into process-mediated, human-sourced and machine-generated data, the table maps each family to representative examples, its main contribution to the analysis, the typical unit of observation, and the principal limitations affecting its use.

Table 1: Overview of data source families by mode of generation

GENERATIVE FAMILY	SOURCE FAMILY	EXAMPLES	MAIN CONTRIBUTION	TYPICAL UNIT	MAIN LIMITATION
Process-mediated	Administrative registers	Social security, tax, education, childcare, PES	Longitudinal coverage; observed participation and outcomes	Person/household/firm	Access; identifiers; cross-country comparability
	Health records and claims	EHR, hospital discharge, prescriptions, insurance claims	Clinical events and service use	Person/event	System-specific coverage and coding
	Transaction and platform records	Payments, bookings, e-commerce	Demand, use and market behaviour	Transaction/person/firm	Commercial access; partial coverage
Human-sourced	Digital traces	Internet searches, social media, online reviews, citizen reporting	High-frequency expressed concerns and perceptions	Person/area/platform	Selection; platform-specific populations
	Online job advertisements	Web-scraped vacancy postings	Labour demand and skill requirements	Vacancy/firm/area	Duplication across boards; non-advertised hiring unobserved
Machine-generated	Mobile and mobility data	CDR, GPS, app location	Movement and effective accessibility	Person/area	Privacy; representativeness

	Environmental sensors and satellite	Air quality, temperature, land cover, night lights	Fine spatial and temporal exposure	Grid/pixel/area	Indirect measurement; requires geocoded personal records

Source: Authors' elaboration

3.5 What alternative data can add to social-investment evaluation

The three mechanisms at the basis of the Coleman's Boat theory introduced in Section 1 imply a specific empirical object: the return to a social investment is a pathway traversed over time rather than an effect measured at a single point in it. In health, that pathway can be represented as: public investment → availability, accessibility and quality of services → service utilisation → health outcome → employment and earnings → fiscal and social return. Analogous sequences apply to ECEC (investment → enrolment and service quality → child development and parental labour supply → later educational attainment and earnings) and to ALMP (investment → programme participation → job search and matching → re-employment and earnings → fiscal effects). These sequences combine short-run returns, such as labour-market participation and income gains that reduce poverty, with long-run returns in human development and attainment; the two unfold over different horizons and require different observation windows (see Bursen et al., (2026)). No single source normally observes the entire sequence for a given individual or household. Each generative family is nonetheless better suited to some segments of the pathway than to others: process-mediated sources to participation and to downstream earnings and fiscal effects, machine-generated sources to availability and accessibility, and human-sourced sources to perceived quality and to the behaviour that intervenes between entitlement and outcome. The empirical question is therefore not which single source is best, but how these segments can be joined for the same individual, household or place.

The stock–flow–buffer typology that SWINS adopts as an operational principle (Plavgo and Hemerijck, 2021) offers a useful way of allocating source families to segments of this pathway. Stock effects and flow effects are by construction the components most likely to generate an administrative record and are therefore best documented by process-mediated sources. Buffer functions, and more generally the intermediate mechanisms connecting an investment to its eventual outcome, are harder to observe in this way: whether a service was reachable, whether it was actually used, how its quality was perceived, how quickly a household adjusted after a shock.

The value of alternative data depends less on the richness of any individual source than on the possibility of joining sources that observe different segments of the same pathway. A childcare register may identify exposure to a service, an education register may capture intermediate outcomes, and employment or tax records may reveal longer-run economic consequences; health records may identify diagnoses and utilisation, while social-security and tax records capture the subsequent employment and income response. As Sections 4 and 5 will show, the studies that succeed in tracing an investment through to a measurable return are, with few exceptions, studies of linked data.

The conditions that make such designs possible are however institutional before they are statistical. Linkage requires common identifiers, an appropriate legal basis, governance arrangements that satisfy data-protection requirements, and, where cross-country comparison is intended, conceptual harmonisation of the linked variables. Such conditions are unevenly distributed, and one consequence is visible throughout the review that follows: the geography of the available evidence tracks the geography of data infrastructure more than the geography of policy interest, and the countries appearing most frequently in this literature are those whose registers were designed, decades ago, in ways that made them joinable.

3.6 Search scope and method

The evidence base underlying the domain sections and the mapping in Section 7 was assembled through purposive, targeted search. Coverage was bounded along two criteria. Firstly, the search was restricted to

sources addressing early childhood education and care and long-term care, active labour market policies and life-long learning, or health, in line with the three SWINS domains. Secondly, priority was given to the SWINS partner countries, but the search was not restricted to them: relevant studies from other EU and non-EU countries were included wherever they were found to speak usefully to the domain and data-source questions at hand. Search combined targeted queries on Google Scholar with citation snowballing: once a relevant study or infrastructure was identified, its reference list and, where feasible, the literature citing it were screened for further sources meeting the same domain criterion. No publication-date restriction was applied; relevance to the domain and data-source criteria took precedence over recency. The resulting corpus is accordingly a purposive, non-exhaustive mapping of the field rather than a census. Therefore, entries marked "None identified" in Table 2 record what this search located, not a claim that no relevant activity exists.

4.Alternative data in practice: Care

4.1 Process-mediated data

The strongest existing evidence for long-run ECEC returns comes from population registers, therefore from process-mediated data. Norwegian population registers provide a prominent example. Havnes and Mogstad (2011, 2015) exploit the expansion of subsidised childcare and use population-register outcomes to trace effects on education, earnings and welfare dependence over long horizons. An important adjacent example is Carneiro, Løken and Salvanes (2015), who exploit a sharp discontinuity in Norwegian maternity-leave entitlements around 1 July 1977 to study a maternity-leave reform rather than childcare, tracing this early-life policy exposure through population registers into outcomes, such as school dropout and wages, observed decades later. The Norwegian setting illustrates the central advantage of linked registers for social investment: the treatment occurs in early childhood while some outcomes are observed decades later.

Germany provides evidence using administrative school-entry examination records as outcome data: Cornelissen et al. (2018) exploit a staggered childcare expansion to estimate heterogeneous returns to early childcare attendance, finding that children from disadvantaged backgrounds benefit most despite being less likely to attend. Italy provides population-level educational outcome data through INVALSI, used in studies linking early childcare to both maternal labour-market outcomes and children's developmental outcomes (Brilli, Del Boca and Pronzato, 2016; Carta and Rizzica, 2018). The sources differ across these studies; Carta and Rizzica (2018), for instance, combine the Italian Labour Force Survey, INVALSI data and administrative records from the Ministry of Education.

Long-term care presents the same structure, with the difference that the relevant outcomes are health-system rather than labour-market ones. Spain offers the most developed European example among the partner countries. Serrano-Alarcón et al. (2022) use administrative records for the universe of applicants to the Catalan long-term care system, linked to primary and secondary healthcare records, and identify effects from plausibly exogenous variation in the leniency of the needs assessment. The Netherlands provides a complementary application. Bakx et al. (2020) exploit the random assignment of assessors, and the variation in their tendency to grant eligibility for nursing-home admission, to estimate the effect of eligibility itself on healthcare costs, hospitalisations and survival, finding that medical care use falls when eligibility is granted while total healthcare spending is unaffected.

4.2 Human-sourced data

Human-sourced data have also been used to estimate some outcomes of ECEC investments, but mainly in the US. The reference study is Herbst et al. (2020), who analyse Yelp reviews of 9,761 childcare businesses across the forty largest US cities, using text classification and sentiment analysis to measure perceived quality.

They find that satisfaction is substantially higher in higher-income markets; that safety, the learning environment and child–teacher interaction are the most frequently evaluated attributes; and that lower-income parents comment more on price and accessibility while higher-income parents comment more on the learning environment.

Online job postings have been used as a real-time measure of the childcare market itself. Ali, Herbst and Makridis (2021) combine the online job-postings database with Google Trends search intensity to show that early-education job postings fell by 13–16% after US stay-at-home orders in 2020, concentrated among private providers, while household search behaviour barely moved. Internet search data have been proposed as a proxy for childcare demand where conventional data are unavailable. Cooper and Reid (2026) examine the spatiotemporal properties of Google Trends childcare queries and conclude that the source is informative but strongly scale-dependent, becoming unstable at fine geographies.

4.3 Machine-generated data

Machine-generated data enter the ECEC literature principally to assess the accessibility of care centres and children's well-being. Accessibility matters for social-investment evaluation because it conditions the conversion of expenditure into participation: two areas with the same spending, or the same number of places, can offer very different effective access, and the return to an investment is mediated by also that difference. Fransen et al. (2015) use data on daycare centers, data on commuting patterns from the Flanders Multimodal Model (MMM) and data about the road network from TomTomMultinet to construct a measure for daycare accessibility in Flanders. Pennerstorfer and Pennerstorfer (2021) compute accessibility on a 250-metre grid for all childcare centres in Vienna. In a selection of OECD countries, Almeida et al. (2024) combine geolocated facility registers with gridded population and transport networks to compute harmonised travel-time accessibility to public employment services, primary schools and ECEC.

Accessibility measures have been constructed for care services for older people along the same lines. Guida, Carpentieri and Masoumi (2022) build a multimodal network for Milan, combining pedestrian, bus, tram and metro layers to measure accessibility to the health-protection agency services responsible for older residents.

Accelerometer studies of preschool physical activity, sedentary time and sleep have been conducted in Belgium and across six European countries: the ToyBox study measured 472 preschoolers in 27 Flemish kindergartens over six days (De Craemer et al., 2014) and later 2,438 children aged four to six across Belgium, Bulgaria, Germany, Greece, Poland and Spain (De Craemer et al., 2017). To our knowledge, however, no European study to date computes a travel-time or catchment-based accessibility measure for childcare and merges it onto individual records. Place-based measures do exist in this literature, since European identification strategies typically use municipal or provincial coverage rates, as in Brilli, Del Boca and Pronzato (2016), but coverage in an administrative area is a different quantity from the accessibility of a service to a household, and the second has not yet been constructed.

5.Alternative data in practice: ALMP and life-long learning

ALMP and life-long learning evaluation is the most mature field for process-mediated data. European employment agencies routinely generate records on programme participation, unemployment spells, job search and subsequent employment. Combined with appropriate identification strategies, these support counterfactual evaluation and, increasingly, heterogeneous-treatment-effect analysis.

Two uses of alternative data should be kept apart in this domain. Some sources measure the effectiveness of an ALMP directly, by recording participation and subsequent outcomes, or by serving as the channel through which an intervention is delivered and its take-up observed. Others measure the labour market in which programmes operate (labour demand, search intensity, job loss) and are relevant to evaluation because the effect of a programme depends on that context, but they do not by themselves measure a return to an investment in ALMP. Process-mediated sources belong overwhelmingly to the first group. Most human-sourced and machine-generated applications reviewed below belong to the second, and are discussed here for what they suggest about how such measures could be attached to evaluation designs.

5.1 Process-mediated data

In Germany, Goller et al. (2025) use large administrative data on long-term unemployed people to evaluate job-search assistance and training with causal machine learning (a family of methods that use machine-learning algorithms to estimate treatment effects under standard identification assumptions, and in particular to estimate how those effects vary across individuals), finding positive effects across programmes and substantial heterogeneity associated with local labour-market conditions and participant characteristics. The same German infrastructure supports evaluation of training for those already in work. Dauth (2020) uses the Integrated Employment Biographies to evaluate WeGebAU, a subsidy for further training of low-skilled employed workers, exploiting cross-regional variation in the policy styles of local employment agencies for identification. The design illustrates a point: because participation is voluntary and concentrated among those already better placed, identification depends on finding administrative variation in how the entitlement is administered rather than in who is entitled.

Belgium provides another example. Cockx, Lechner and Bollens (2023) use administrative data on unemployed people and causal machine learning to evaluate training programmes, finding positive post-lock-in effects (that is, effects measured after the lock-in period, during which participants in training search less intensively and their employment is temporarily reduced) and substantial heterogeneity across programmes and unemployed workers, with particular attention to unemployed immigrants. The study illustrates how administrative data can move beyond average treatment effects toward rules for allocating programmes to people according to estimated gains. A further Belgian example is provided by Albanese, Cockx and Dejemeppe (2024), who use administrative data to estimate the short- and long-term effects of a hiring subsidy for low-educated unemployed youths. The study follows employment effects over several years, illustrating how administrative records can capture longer-term labour-market returns alongside immediate programme impacts.

Sweden offers a further example, focused on the labour-market integration of immigrants. Using register-based longitudinal data, Bonfanti and Nordlund (2012) assess the impact of *Swedish for Immigrants* (SFI) language training on migrants' subsequent position in the labour market. Khaef (2026) extends this line of research methodologically: drawing on longitudinal register data covering immigrants who arrived in Sweden between 2000 and 2016, the study traces enrolment timing and subsequent participation patterns in SFI over a ten-year follow-up, documenting substantial selectivity in who remains enrolled and for how long. Norwegian registers support the same kind of question for later-life learning: Midtsundstad and Nielsen (2019) use individual register data from Statistics Norway to relate adult education completed after age 40 to subsequent participation in employment among older workers.

Ireland provides another strong example of the use of linked administrative data for ALMP evaluation. OECD, Department of Social Protection and European Commission/JRC (2024) link Department of Social Protection records on unemployment, benefit receipt and participation in the Tús and Community Employment employment-support programmes with Revenue data on earnings and employment. This allows researchers to reconstruct individuals' labour-market histories and examine outcomes following programme participation.

In Spain, Álvarez de Toledo Saavedra et al. (2017) use job-placement information in the MCVL (Muestra Continua de Vidas Laborales) to study worker–job matching and labour-market segmentation; the study does not estimate an ALMP effect but shows how records containing information on both workers and jobs generate evidence unavailable from household surveys. In Hungary, the country's Admin4 database covers a representative 50% sample of the population for 2003–2021, combining employment and firm information with administrative data on unemployment, social transfers, education, healthcare use and pensions, and permitting labour-market transitions to be followed across employment and non-employment states.

5.2 Human-sourced data

Where a public employment service operates its own online portal, the resulting search logs constitute human-sourced data. Marinescu and Skandalis (2021) link French Pôle emploi registers to the platform's search and application logs for more than 400,000 unemployed workers, showing that search effort rises by at least 50% in the year before benefit exhaustion while targeted wages fall. Hensvik et al. (2021) use more than 180 million clicks on approximately 1.2 million advertisements on Platsbanken, the Swedish public employment service platform, to show that during the 2020 crisis job seekers reduced search intensity and reallocated search toward less-affected occupations beyond what vacancy shifts alone would predict.

A small number of experiments deliver an ALMP intervention through a digital platform and measure outcomes in registers. Belot et al. (2019) build a custom search interface on the UK Universal Jobmatch database and show that suggesting alternative occupations broadens search and raises interview rates, particularly for those searching narrowly. Belot et al. (2025) scale the design through the Dutch public employment service: among 10,738 jobseekers targeting structurally weak occupations, employment, hours and labour income rise 5–6% at eighteen months. Altmann et al. (2022) run an experiment on the Danish portal covering approximately 92,098 unemployment-insurance recipients. Their results show that online job search advice is highly effective when the share of treated workers is relatively low: in regions where less than 50% of job seekers are exposed to the treatment, working hours and earnings of treated job seekers increase by 8.5–9.5% in the year after the intervention.

Human-sourced data can also measure the labour market in which such policies operate rather than the policies themselves, the level and composition of labour demand, the intensity of job search, the incidence of job loss. In the terms presented in the Introduction, these are measures of the situational mechanism: the opportunity structure into which a programme places a participant, and on which its effect depends. For instance, web-scraped vacancy postings can be used to analyse the labour market demand. Cedefop's Skills-OVATE and Eurostat's Web Intelligence Hub collect and classify advertisements from national and European portals and for instance Lovaglio and Riussi (2025) combine the Cedefop database with EU-LFS regional data to document demand-side job polarisation across European regions after COVID-19. Štefánik, Lyócsa and Bilka (2023) show that Slovak online postings have genuine predictive power for official employment and unemployment indicators. However, one should be aware of these data's caveats: Napierala et al. (2022) benchmark the Web Intelligence Hub database against the EU Labour Force Survey and the Job Vacancy Survey and find systematic over-representation of high-skilled occupations and capital regions, with under-representation of public administration and of middle- and low-skilled roles, varying by Member State.

There are also studies that use search queries to nowcast labour-market conditions in Europe. From the work of Askitas and Zimmermann (2009), who relate German unemployment-related keyword volumes to official monthly unemployment, the approach has since been applied across Europe: Fondeur and Karamé (2013) for French youth unemployment, Naccarato et al. (2018) for Italy using the query "offerte di lavoro", Mulero and García-Hiernaux (2021) for Spain. The most methodologically developed European application is Caperna et al. (2022), who construct a cross-country comparable index from more than 2,000 unemployment-related queries across all EU27 Member States using machine-learning feature selection, and show that unemployment-related searches rose roughly 30% after COVID-19 lockdowns and remained elevated for more than five weeks. Similarly, but with social media data, Antenucci et al. (2014) for the US process 19.3 billion

tweets to construct indexes of job loss, job search and job posting from phrase matching, reducing ten job-loss signals to a single factor by principal components.

Using Twitter data, Llorente et al. (2015) relate Twitter-derived behavioural fingerprints, including mobility, activity patterns and social diversity, to unemployment across Spanish regions. In Italy, ISTAT publishes a Social Mood on Economy Index as an official experimental statistic: a daily index of economic sentiment computed from Italian-language tweets on X. Boss, Longo and Onorante (2025) extend the approach to the euro area, using Reddit content and large language models to nowcast inflation and unemployment.

5.3 Machine-generated and sensor data

Machine-generated data in the ALMP domain are used to detect labour-market shocks and to measure the spatial structure of labour markets. Toole et al. (2015), using call detail records from two industrialised European countries, analyse the closure of an automotive-parts plant that cost approximately 1,100 jobs in a town of 15,000. Using the closure of a large manufacturing plant as a case study, they first describe a structural break model to correctly detect the date of a mass layoff and estimate its size. Then they use a Bayesian classification model to identify affected individuals by observing changes in calling behaviour following the plant's closure and for these affected individuals they observe significant declines in social behaviour and mobility following job loss. Substantively, their study detects both the aggregate event and the individual exposure without any administrative record. Pintér and Felde (2022) derive Budapest metropolitan commuting patterns from 955 million call detail records covering 1.63 million SIM cards and validate them against the 2011 census with correlations between 0.85 and 0.97, establishing mobile network data as a fast substitute for census commuting statistics.

The three families play distinct roles in this domain. Process-mediated records are where ALMP effects have actually been estimated: participation, spells and subsequent earnings are recorded as a condition of administration, and the causal machine learning literature has recently used them to move from average effects towards allocation rules. Human-sourced and machine-generated sources appear in two different capacities - as the medium through which an intervention is delivered and its take-up observed, as in the digital job-search experiments, and as instruments for measuring the labour market into which a programme places a participant. The first is genuinely the evaluation of a policy; the second describes its context. No study reviewed here uses an online vacancy, a search query or a mobility trace to estimate the return to an active labour market policy.

6.Alternative data in practice: Health

Health occupies a distinctive position within the social-investment framework. Rather than treating health solely as an outcome of welfare-state policy, Goijaerts, Van der Zwan and Bussemaker (2023) argue that health can also be understood as an input into the stock, flow and buffer functions of social investment. Health affects the capacity to participate in education, work and social life, while employment and social protection can themselves affect health. This bidirectionality makes health particularly difficult to evaluate using expenditure and survey data alone.

The health literature also illustrates why the joining of different data sources matters. A health shock can be observed in hospital or claims data, while its economic consequences can be observed in employment, social-security and tax records. The resulting chain is closer to the social-investment logic than a simple regression of health expenditure on life expectancy.+

6.1 Process-mediated data

The most developed strand of this literature traces a health event through to its labour-market consequences. Simonetti, Belloni, Farina and Zantomio (2022) use WHIP-Salute, an Italian administrative dataset linking INPS work and social-security histories (1990–2012) to hospital discharge records from all public and private hospitals to study employment histories, social-security information and hospitalisation records. They find substantial and persistent reductions in employment and labour income following acute cardiovascular shocks, with important differences by labour-market institutions and worker characteristics. This provides a direct empirical building block for estimating the economic value of avoiding or mitigating health deterioration. Lazuka (2023) pursues the same question with longitudinal Swedish administrative data, asking whether medical innovations mitigate the economic consequences of health shocks; the study reports that innovation reduces negative economic consequences for affected individuals and generates spillovers to relatives, with substantial heterogeneity across diagnoses.

For Hungary, Bíró, Boza, Gyetvai and Prinz (2024) use matched employer–employee administrative data linked to individual health records. They show that workers experiencing health shocks have different labour-market and disability-insurance responses depending on firm quality. Hungary also provides a direct example of linking healthcare spending, income, geography and mortality. Bíró and Prinz (2020) use administrative data on a random 50% of the Hungarian population for 2003–2011 to study geographic and income-related inequality in healthcare spending and mortality. This is closer to the distributional side of the social-investment question: the data allow healthcare spending and health outcomes to be examined jointly by income and geography.

Spain provides a complementary example in the health domain. Muñoyerro-Muñiz et al. (2020) describe the Andalusian Health Population Database, which links pseudonymised population information with electronic health records and multiple healthcare information systems, including hospital activity, mental health, laboratory and imaging data, vaccination, renal care and pharmacy information. This is an example of extensive linkage within the health system rather than a demonstrated linkage between health and employment records, but it illustrates the level of longitudinal and multi-source health information that can be generated by administrative infrastructures.

Belgium illustrates a different kind of linkage, between survey and administrative health data. Berete et al. (2023) describe the HISlink project, which linked Belgian Health Interview Survey data to Belgian compulsory health-insurance records using the national register number. Successful linkage was achieved for 92.3% of participants in the 2013 wave and 94.2% in 2018. The resulting linked data combine information on health, socioeconomic characteristics and health behaviours from the survey with routinely collected administrative information on healthcare utilisation. This example is particularly relevant for SWINS because it demonstrates how administrative data can complement, rather than simply replace, survey information by adding detailed observed healthcare use to richer individual-level characteristics.

Electronic health records (EHR) and administrative health records (AHR) constitute the largest process-mediated resource in this domain. AHR are best understood as a subset of EHR: whereas EHR comprises electronic clinical data generated at the point of care for treatment purposes, AHR refers specifically to the subset collected for administrative purposes - reimbursement, eligibility or registry enrolment. Caruana et al. (2023) systematically review 70 such studies and identify twelve machine learning techniques and eight application areas directly relevant to a life-course perspective, since both exploit the temporal structure of linked administrative records. The review also identifies fragmentation, the difficulty of linking patients across disparate single-source administrative datasets, as one of the most significant limitations facing AHR-based research. For SWINS, EHR provide detailed clinical trajectories whose value for social-investment evaluation increases substantially when they can be joined to employment, income or social-protection records.

While the studies above trace the economic consequences of health shocks, a smaller set of cases evaluates a deliberately implemented health-related service using alternative process-mediated data rather than health expenditure alone, though the two examples differ in design. The Swedish Salut Programme is a quasi-experimental case: municipalities adopting the programme after 2006 were compared with municipalities that had not yet adopted it, using routinely collected linked register data, finding improved birth outcomes and changes in downstream healthcare costs (Sampaio et al., 2023). A second example is a randomised controlled trial of an extended early-intervention service for first-episode psychosis in Quebec, where treatment assignment was randomised but costs were derived from administrative hospitalisation, medical service and prescription records (Groff et al., 2021). These two cases demonstrate the feasibility of using administrative data for service evaluation, though through different routes: one observational, one experimental. They do not, however, yet constitute a consolidated European literature comparable with ALMP evaluation.

6.2 Human-sourced digital traces

Human-sourced data enter the health domain principally through unsolicited online feedback: ratings and free-text reviews that patients write about providers on public platforms. Their attraction is that they capture an attribute registers do not record - how care was experienced - at a frequency and cost that official patient surveys cannot match. For the Netherlands, Hendrikx et al. (2019) apply text and sentiment analysis to unsolicited ratings and comments on ZorgkaartNederland, the national provider-review platform, to characterise care experience at regional level. In England, Boylan, Turk, van Velthoven and Powell (2020) analyse 1,396 public reviews covering all 70 general practices in Oxfordshire and correlate them against both the GP Patient Survey and the Friends and Family Test, finding significant moderate correlations with each, alongside qualitative evidence that reviews contain actionable content on quality and safety. In Germany, Emmert et al. (2015) study 65 practices on the physician-rating platforms jameda and Weisse Liste between 2011 and 2013, finding significant association with offline patient-satisfaction surveys but weak and inconsistent association with clinical indicators for diabetes, asthma and coronary heart disease. Across these studies the pattern is consistent: online patient feedback correlates moderately with subjective official instruments, and weakly or not at all with objective clinical quality. It measures experience and should be read as such.

6.3 Machine-generated data

Machine-generated data enter the health domain along two paths: one measures where services are and who can reach them; the other measures the physical environment to which people are exposed.

Geospatial data (e.g., population grids, road and public-transport networks, and geolocated facility registers) are used to measure the accessibility of services. At EU level, Eurostat's GISCO programme now publishes geographic accessibility datasets computing driving time from the Census 2021 population grid to the nearest healthcare facility across member states. In Italy, Pecoraro et al. (2024) decompose intra- and inter-regional components of accessibility using hospital discharge data and ISTAT travel-time matrices. In Finland, Kotavaara et al. (2021) assess multimodal accessibility to primary health care on a 250-metre register-based population grid combining car and scheduled public transport.

Mobile-phone and GPS data can measure mobility and effective access directly. In China, Xia et al. (2026) analyse 96 million de-identified mobile-phone users across 11 Chinese cities to study hospital utilisation and hospital-bypass behaviour, finding spatial and socioeconomic differences in how people use hospitals. It does not estimate the return to healthcare investment, but it demonstrates a measurement capability directly relevant to such a framework: effective accessibility and observed utilisation rather than simple geographic distance. Mobile-phone data have also been used to construct high-frequency proxies of territorial socioeconomic development and wellbeing. Pappalardo et al. (2016) develop an analytical framework using mobility and social measures extracted from mobile-phone data to estimate indicators of socioeconomic development and wellbeing.

Where the object of interest is the environment to which people are exposed, environmental and spatial machine-generated data are the relevant source, and attaching a place-based measurement to an individual through their residential address has become an established routine. Satellite observations, air-quality measurements, temperature and other environmental data can measure exposures that are absent from administrative records and two studies in particular reach outcomes that belong to the social-investment framework. For instance, Lim et al. (2024) assess the relationship between 16 years of exposure to $PM_{2.5}$, PM_{10}, and gaseous pollutants and Grade Point Averages (GPA) in exit examinations for 785,312 Danish children, finding a measurable reduction in grade point average associated with early-life particulate exposure. Van den Brekel et al. (2024) join address-level pollutant concentrations to demographic and socioeconomic registers for 17.3 million Dutch residents to document ethnic and socioeconomic inequalities in exposure.

Health is the domain in which the constituent data are richest and the estimated returns fewest. The process-mediated infrastructures reviewed here are among the most sophisticated in this report, but the studies that follow a health investment through to an economic or fiscal return remain few. The dominant design traces a health shock rather than a health service. This is a natural consequence of how the data are generated, since shocks produce sharp, well-dated administrative records and services often do not, but it means the literature currently answers a question adjacent to the one social-investment evaluation poses.

6.4 Looking ahead: the European Health Data Space

The fragmentation problem identified above is a structural constraint on the extent to which health-related data can be linked, at scale, to the kind of social-investment evaluation this work is concerned with. At the EU level, this constraint is beginning to be addressed directly through policy rather than through individual research infrastructure investments alone. The European Health Data Space (EHDS) is the clearest instance: the first EU-wide attempt to establish a legal and technical pathway for the cross-border data reuse that fragmented national systems have made difficult.

The Regulation entered into force on 26 March 2025. According to the European Commission, secondary use can cover electronic health records, administrative data such as claims and hospital discharge reports, research data, and data from wearables and mobile applications. The main rules on secondary use for many EHR categories are scheduled to start applying from March 2029, with further categories following in 2031. The EHDS therefore does not mean that harmonised European individual-level health data are already freely available; rather, it establishes a framework and infrastructure for controlled secondary use, including national Health Data Access Bodies and HealthData@EU.

For SWINS, the practical implication is that current national health registers should be treated as the principal empirical data sources, while EHDS should be treated as an emerging European infrastructure that may facilitate discoverability, access and cross-border reuse in the future.

7.Cross-country mapping for the SWINS consortium

This section maps the data infrastructures available across the eight SWINS partner countries. Its inclusion criterion is institutional: Table 2 below lists resources produced or maintained by a public body as a standing asset, independently of any single research project. Datasets assembled for one study, and the empirical findings obtained from them, are not listed here. The mapping distinguishes between the existence of a data infrastructure, its potential for cross-domain linkage, and evidence that the data have already been used in empirical research. The presence of an infrastructure should therefore not be interpreted as implying that individual-level linkage is currently available or that researchers have unrestricted access.

Table 2: Main alternative data sources by SWINS partner country

PARTNER COUNTRY	ALTERNATIVE-DATA INFRASTRUCTURE	RESEARCH POTENTIAL / LINKAGE	EVIDENCE
Belgium	Datawarehouse Labour Market &Social Protection (DWH LM&SP), managed by the Crossroads Bank for Social Security; Belgian Compulsory Health Insurance (BCHI) administrative data; Health Data Agency (HDA), which facilitates secure access and reuse of health and health-related datasets.	The DWH LM&SP links socioeconomic information from Belgian social-security institutions and other public bodies and supports both research data requests and policy evaluation. It provides a strong infrastructure for studying employment, unemployment, social protection and labour-market trajectories. Administrative data have also been linked to regional childcare records to study the relationship between childcare use and parental employment. In the ALMP domain, Belgian administrative unemployment data have been used to evaluate training programmes and hiring subsidies and to estimate heterogeneous and longer-term employment effects. In health, the HISlink project links Belgian Health Interview Survey data to compulsory health-insurance records, combining individual health and socioeconomic information with administrative healthcare utilisation.	**Official data sources:** Crossroads Bank for Social Security / DWH LM&SP; IMA/BCHI; Health Data Agency; Agency for Growing Up. **Key empirical evidence:** Cockx, Lechner and Bollens; Albanese, Cockx and Dejemeppe; Goffin et al./IDEA Consult childcare study; Berete et al. (2023).
Germany	IAB/FDZ administrative labour-market data, including SIAB; NEPS-ADIAB linking National	Detailed longitudinal employment, earnings and unemployment histories. NEPS-ADIAB combines educational trajectories,	**Official data sources:** IAB/FDZ; SIAB; NEPS-ADIAB. **Key empirical**

PARTNER COUNTRY	ALTERNATIVE-DATA INFRASTRUCTURE	RESEARCH POTENTIAL / LINKAGE	EVIDENCE
	Educational Panel Study survey/test data with IAB administrative employment histories.	competencies and other survey information with administrative employment histories, supporting research on education-to-employment transitions and long-term returns to education. German administrative data have also been harmonised with French DADS-EDP for comparative research. Cross-domain linkage with health data is more constrained and should be assessed separately.	**evidence:** Bachbauer, Wolf and Fuß (2022); Cornelissen et al. (2018); Goller et al. (2025); Filser et al. (2026)
Hungary	**Linked Administrative Panel Database (Admin4),** a linked employer–employee administrative database covering a representative 50% sample of the Hungarian population, complemented by information on unemployment, social transfers, education, healthcare, pensions, family links and firm characteristics.	Longitudinal individual- and firm-level information on employment, earnings, unemployment, transfers, education, healthcare use and pensions. The database can support analysis of labour-market and social-protection trajectories; separate research has also demonstrated linkage between employer–employee records and individual health records.	**Official/research data infrastructure:** KRTK Databank / Admin4. Boza, Pető and Tir (2026) **Key empirical evidence:** Bíró, Boza, Gyetvai and Prinz (2024); Bíró and Prinz (2020)
Ireland	CSO Administrative Data Centre and Health Research Data Centre; Educational Longitudinal Database (ELD); Department of Social Protection and Revenue administrative records; Data Access, Storage, Sharing and Linkage (DASSL**)**	CSO's PIK-based infrastructure supports secure linkage of pseudonymised administrative records. The ELD links education records to subsequent employment/self-employment and social-protection outcomes. Linked DSP–Revenue data have been used to evaluate ALMPs and follow individual labour-market trajectories. Health data can be accessed through controlled research arrangements, while DASSL demonstrates the technical feasibility of linking multiple health and social-data sources in a secure environment. DASSL should not be presented as a fully operational national linked database.	**Official data sources:** CSO; Department of Social Protection; Revenue. **Key empirical evidence:** CSO ELD; OECD/DSP/EC-JRC (2024) on ALMPs; Fennelly et al. (2022).

PARTNER COUNTRY	ALTERNATIVE-DATA INFRASTRUCTURE	RESEARCH POTENTIAL / LINKAGE	EVIDENCE
Italy	INPS social-security and labour-market records; Ministry of Health administrative data incl. Hospital Discharge Records (SDO) and the developing SIISI; INVALSI assessment data; Istat Sistema Integrato dei Registri; WHIP-Salute (National Statistical Plan, code SAL-00054), linking INPS histories 1990–2012 to hospital discharge records on a 7% sample, ISTAT Social Mood on Economy Index	These sources provide information on employment, earnings, social-security histories, educational outcomes and healthcare utilisation. The feasibility of linking sources varies by dataset and research setting. A concrete example is provided by Simonetti et al. (2022), who link employment, Social Security and hospitalisation records to study the long-term labour-market consequences of acute health shocks.	**Data sources:** INPS; Ministry of Health; INVALSI; ISTAT **Key empirical evidence:** Simonetti et al. (2022).
Norway	Statistics Norway (SSB) register-based microdata, covering population, education, employment, income and welfare; national health registries managed by the Norwegian Institute of Public Health and other authorities; municipal health and care registers; national kindergarten and education registers.	Norway provides extensive individual-level administrative data that can be linked across domains and over time, subject to research-access requirements. The infrastructure supports analysis of education, employment, earnings, welfare, family and health trajectories. National health registries provide information on hospital care, mortality, prescriptions, births, chronic conditions and other health events, while municipal registers add information on primary and long-term care. The childcare infrastructure includes national kindergarten registers and provider-level information.	**Official data sources:** Statistics Norway; Norwegian Registry for Primary Health Care (KPR); Norwegian Institute of Public Health; Directorate for Education and Training. **Key empirical evidence:** Havnes and Mogstad (2011); Havnes and Mogstad (2015); Carneiro, Løken and Salvanes (2015).
Serbia	Statistical Office of Serbia administrative sources; health and social-insurance administrative systems; national open-data infrastructure; University Clinical Center of Serbia EHR data	Administrative information on employment, pensions/disability insurance, social insurance and health can potentially support analysis across social-investment domains. However, evidence for established individual-level cross-domain research linkage is more limited than in the other partner countries, particularly for health. Interoperability and access to	**Official data sources:** Statistical Office of the Republic of Serbia; Republic of Serbia Open Data Portal; University Clinical Center of Serbia / HMA-EMA real-world-data catalogue. **Key empirical / methodological evidence:** No strong peer-reviewed example

PARTNER COUNTRY	ALTERNATIVE-DATA INFRASTRUCTURE	RESEARCH POTENTIAL / LINKAGE	EVIDENCE
		depersonalised health data remain important constraints.	of individual-level cross-domain administrative linkage was identified.
Spain	MCVL (Continuous Working Life Sample), based on Social Security administrative records and supplemented with population-register and, in its fiscal version, personal income-tax data; Andalusian Health Population Database.	MCVL allows individual working-life trajectories to be constructed from repeated annual data on employment, earnings and Social Security histories, subject to access and confidentiality requirements. The WORKss cohort demonstrates its use for longitudinal research on employment trajectories, sickness absence, disability and mortality. The Andalusian database provides extensive linked health information across EHR, hospital, mental-health, laboratory/imaging, vaccination, renal and pharmacy systems. These are complementary infrastructures; no single national MCVL–EHR database is documented.	**Official data sources:** Spanish Social Security / MCVL; Andalusian Health Population Database. **Key empirical evidence:** López et al. (2014); López Gómez et al. (2016, WORKss); Muñoyerro-Muñiz et al. (2020).

Source: Authors' elaboration

Belgium

Belgium has a well-developed system of administrative and linked data that is relevant across several areas of social investment. The Datawarehouse Labour Market and Social Protection (DWH LM&SP), managed by the Crossroads Bank for Social Security, provides an important infrastructure for combining administrative information on employment, unemployment and social protection. It can also be linked to data from other domains: for example, administrative childcare records in Flanders have been linked to the DWH LM&SP to study parental employment, while Belgian administrative labour-market data have been used to evaluate training programmes and hiring subsidies (Cockx, Lechner and Bollens, 2023; Albanese, Cockx and Dejemeppe, 2024).

In health, the Health Data Agency provides a national framework for the secure sharing and reuse of health data, while Belgian Compulsory Health Insurance (BCHI) provides routinely collected information on healthcare utilisation. The HISlink project demonstrates the feasibility of linking BCHI data with the Belgian Health Interview Survey, with successful linkage rates above 90% in both waves studied (Berete et al., 2023). Overall, Belgium provides documented examples of administrative data being linked across labour-market, childcare and health domains, although these linkages are implemented in specific research settings rather than through a single integrated cross-domain database.

Germany

Germany has one of the strongest administrative-data infrastructures among the SWINS partner countries, particularly for longitudinal labour-market and education-to-employment research. The Institute for Employment Research (IAB) and its Research Data Centre (FDZ) provide access to detailed administrative labour-market data, including the SIAB, which contains longitudinal information on employment, earnings and unemployment based on social-security records. A particularly valuable infrastructure is NEPS-ADIAB, which

links survey and competence data from the National Educational Panel Study (NEPS) at the individual level to detailed IAB employment histories. The administrative component extends the observation period substantially, with employment data in the NEPS-ADIAB products reaching back to 1975. This combination allows researchers to study educational trajectories, competencies and decisions together with subsequent employment, earnings and career outcomes (Bachbauer, Wolf and Fuß, 2022).

The German infrastructure has also been used directly to evaluate social-investment policies. In the ECEC domain, Cornelissen et al. (2018) use administrative school-entry examination records to study heterogeneous effects of an expansion of early childcare, finding particularly large benefits for disadvantaged children. In the ALMP domain, Goller et al. (2025) use large administrative data on long-term unemployed individuals to evaluate job-search assistance and training programmes, including heterogeneous treatment effects using causal machine learning. These applications illustrate the ability of German administrative data to capture both long-term outcomes and treatment-effect heterogeneity.

Germany also provides evidence that administrative datasets can be harmonised for comparative research. Filser et al. (2026) harmonise the German SIAB with the French DADS-EDP, both based on social-security records, to construct comparable data on maternal employment and demonstrate the feasibility of cross-national analysis using administrative employment histories. For health, however, the situation is less straightforward: the availability of rich labour-market registers should not be interpreted as evidence of equally mature cross-domain linkage with individual health data. Legal, identifier and interoperability requirements create additional barriers. For SWINS, Germany therefore represents a particularly strong case for longitudinal analysis of education, employment and labour-market policy, while health–employment linkage requires separate assessment.

Hungary

The Hungarian infrastructure of linked employer–employee administrative data infrastructure. The Admin4 database covers a representative 50% sample of the Hungarian population for 2003–2021 and combines longitudinal employment and firm information with administrative information on unemployment, social transfers, education, healthcare use and pensions. Its broad coverage allows labour-market and social-protection trajectories to be followed over time. The database therefore provides a particularly relevant infrastructure for cross-domain social-investment research, while specific linkages with health records depend on the research setting. That linkage has in fact been demonstrated: Bíró, Boza, Gyetvai and Prinz (2024) join matched employer–employee records to individual health records to study labour-market and disability-insurance responses to health shocks, and Bíró and Prinz (2020) use administrative data on half the Hungarian population to examine geographic and income-related inequality in healthcare spending and mortality jointly. Among the partner countries, Hungary therefore offers one of the few documented instances of employment and health records being analysed together at individual level.

Ireland

Ireland has a developing but increasingly structured infrastructure for linking administrative data across several areas relevant to social-investment research. The Central Statistics Office (CSO) operates an Administrative Data Centre in which administrative records are processed and pseudonymised using Protected Identifier Keys (PIKs), allowing approved statistical linkage across data sources. A particularly relevant example is the Educational Longitudinal Database (ELD), which matches education records with subsequent employment and self-employment data from Revenue and benefits data from the Department of Social Protection, providing a basis for analysing education-to-employment and social-protection outcomes over time.

Ireland also provides documented examples of administrative-data use for labour-market policy evaluation. An OECD–Department of Social Protection–European Commission/JRC evaluation of Tús and Community Employment links administrative information on programme participation and social-protection histories with Revenue employment and earnings data, allowing longer-term labour-market outcomes to be examined. In health, the CSO Health Research Data Centre, launched in 2024, provides approved researchers with controlled access to pseudonymised individual-level health and related administrative and survey data through a secure research environment. The DASSL project provides an additional proof-of-concept for linking health and social data, including administrative records, patient registries and electronic health records, but should not be interpreted as an operational national linked database.

Overall, Ireland therefore offers concrete examples of administrative data being used across education, employment/social protection and health, with particularly strong potential for linking education and labour-market outcomes. However, the maturity, scope and access arrangements differ across data infrastructures and should be assessed separately.

Italy

Italy has a broad set of administrative and register-based data sources with potential for social-investment research. These include Istat's register-based statistical infrastructure, INPS Social Security and labour-market records, INVALSI assessment data, and Ministry of Health administrative data, including Hospital Discharge Records (SDO) and the developing SIISI infrastructure for integrating individual health-service events. These sources are complementary rather than part of a single, routinely accessible cross-domain database, and the feasibility of individual-level linkage varies across data sources and research settings.

The principal exception is WHIP-Salute: originating in a 2007 Ministry of Health surveillance project and part of the National Statistical Plan since 2013 under code SAL-00054, it links INPS work and social-security histories for the period 1990–2012. Concrete evidence of the research potential of linked administrative data is provided by Simonetti et al. (2022), who use it to combine employment, Social Security and hospitalisation records to examine the long-term labour-market consequences of acute health shocks. This study is particularly relevant to the social-investment framework because it follows part of the pathway from health to economic outcomes, linking a health shock to subsequent employment trajectories.

Norway

Register-based data in Norway cover population, education, employment, income, welfare and health, and Statistics Norway provides access to microdata for approved research projects. These registers can be combined across domains and followed over time, creating strong potential for life-course analysis of social-investment policies. Norway also has a comprehensive system of national health registries, covering areas such as hospital care, mortality, prescriptions, births and other health events, complemented by municipal health and care registers. For instance, for long-term care, Norway has the Municipal Patient and User Registry (KPR), which incorporates the earlier IPLOS nursing and care statistics, records individual-level information on all recipients of publicly financed care, from in-home safety alarms to full-time institutional care, together with assessed physical and cognitive functioning, and is linkable through the personal identification number to the Norwegian Patient Registry and to Statistics Norway's sociodemographic files. In education and childcare, national education and kindergarten registers provide additional administrative information on educational trajectories and childcare provision.

The research literature illustrates the value of these data for studying long-term returns to social investment. Havnes and Mogstad (2011), for example, use administrative data to evaluate a large expansion of subsidised childcare and find effects on children's educational attainment, labour-market participation and welfare dependency. Carneiro, Løken and Salvanes (2015) similarly use Norwegian administrative data to examine the long-term effects of expanded maternity-leave entitlements, finding effects on children's educational attainment and adult wages. Together, these studies demonstrate the ability of Norwegian administrative data to follow individuals from early-life and family policies through education and into adult labour-market outcomes.

Serbia

Serbia should be treated separately from EU Member States in the mapping. The Statistical Office reports access to administrative sources from tax, pension/disability insurance, employment and social-insurance institutions for official statistics. The national open-data portal contains health datasets, and the University Clinical Center of Serbia is listed in the European HMA/EMA real-world-data catalogue as a source of hospital inpatient, outpatient and electronic health-record data. At the same time, a recent assessment of the Serbian health-information system reports limited interoperability and the absence of clear procedures for access to depersonalised health data. The Serbian case therefore combines substantial data generation with more pronounced linkage and interoperability constraints.

Spain

Spain provides two complementary examples of the potential of administrative data for social-investment research. The Continuous Working Life Sample (MCVL) is based on Social Security administrative records and is supplemented with population-register information and, in its fiscal version, personal income-tax data. Its repeated annual editions allow researchers to construct individual working-life trajectories covering employment, earnings and Social Security histories, subject to access and confidentiality requirements. The MCVL has also been used as the basis for the Spanish WORKing life Social Security (WORKss) cohort, which provides longitudinal information on employment trajectories, sickness absence, permanent disability and mortality (López et al., 2014; López Gómez et al., 2016). At the regional level, the Andalusian Health Population Database provides a complementary health-data infrastructure, linking pseudonymised population information with electronic health records and multiple healthcare information systems, including hospital activity, mental health, laboratory and imaging data, vaccination, renal care and pharmacy information (Muñoyerro-Muñiz et al., 2020). These are complementary infrastructures rather than a single integrated employment–health database: WORKss does not contain comprehensive individual health data, while the Andalusian database is primarily a health-sector linkage infrastructure.

8.Research potential and limitations

8.1 What can realistically be measured?

The evidence reviewed suggests that alternative data can support several distinct components of the social-investment return framework, but the three generative families differ in the way they can contribute.

Process-mediated sources are best positioned to measure long-run individual outcomes such as employment, earnings, benefit receipt, education and, where available, healthcare use. Human-sourced and machine-generated sources are more promising for the intermediate mechanisms that registers do not contain - mobility, effective accessibility, environmental exposure, perceived service quality, and behavioural responses at high frequency. Here, however, the review found a consistent asymmetry: with very few exceptions, these families have not been used to estimate the return to a social investment.

The exceptions to this trend, where a human-sourced or machine-generated measure has been attached to a process-mediated backbone, are in our opinion the ones that define the frontier, and several of them are still working papers. For instance, vacancy text has been linked to matched employer–employee registers in Norway (Audoly et al., 2024), job posts have been matched to labour force survey respondents in Denmark (Daly et al., 2025), platform search logs have been joined to unemployment registers in France (Marinescu and Skandalis, 2021).

The table below summarises the combinations that follow from this.

Table 3: Research potential by policy domain and data source

POLICY DOMAIN	CORE PROCESS-MEDIATED SOURCE	COMPLEMENTARY HUMAN-SOURCED AND MACHINE-GENERATED SOURCES	POTENTIAL RETURN PATHWAY	STATUS
Health	EHR, hospital and claims data + social security and tax records	Environmental exposure; mobility; service accessibility; patient reviews	Health → employment → earnings → fiscal effects	Health shock → employment demonstrated; service → return largely not
ECEC	Childcare and education registers + employment and tax records	Service-location and accessibility data; provider reviews; childcare-market job postings	ECEC → child outcomes and parental employment → earnings	Register-based returns demonstrated; complementary sources not yet linked
ALMP and life-long learning	PES and unemployment records / further-training and adult-education records + employment and earnings data	Online vacancies; search queries; platform search logs; mobility data	Programme → matching and employment → earnings → fiscal effects	Register-based returns demonstrated; platform logs and digital delivery emerging

Source: authors' elaboration

From the review we identify four potential research streams, that can be explored exploiting also alternative data:

1. **From service receipt to access:** Administrative registers record that a service was received; they do not record whether it could realistically have been reached. This is where machine-generated data can contribute. Eurostat's GISCO datasets now provide the supply-side layer for all Member States and can be used to compute a catchment-based accessibility measure for childcare, healthcare and employment services, which can then be merged onto individual register records through the residential address. The binding constraint is the spatial resolution at which personal records are geocoded and released.

What this would yield is an observable conversion factor. The capability approach holds that a resource becomes a real freedom only through conversion, and that conversion depends on factors operating at individual, social and territorial levels. Where two individuals hold the same entitlement but reach different outcomes, that difference is currently absorbed into unobserved heterogeneity; an accessibility measure attached to the individual moves one territorial conversion factor out of the residual and into the model. It also separates two explanations that administrative data conflate: that an eligible person did not use a service because they did not want it, or because they could not reach it.

2. **Non-take-up:** The non-take-up of a service, which is often invisible in administrative data, can be estimated by combining conventional and administrative sources. Eligibility can be simulated through tax-benefit microsimulation of the EUROMOD type; actual receipt is observable in benefit and service registers; the difference between the two is the non-take-up gap. Human-sourced data can add a third layer, as proxies for latent demand: search queries, portal traffic and search logs, and helpline records. The platform-log studies reviewed above show that such data exist within public systems and can be analysed at individual level.
   Those least likely to take up a service, however, may also be least likely to search for it online, so digital traces will systematically under-detect the non-take-up that matters most.

3. **The eco-social linkage:** European environmental epidemiology assigns modelled exposure to residential addresses at population scale and links it to health, education and socioeconomic registers. What we did not identify is any European register-based study linking environmental exposure to labour-market outcomes at individual address level. Studies in this sense could answer questions like: do social services attenuate the consequences of environmental exposure, and do they do so differentially by socioeconomic group? This is also where satellite and sensor data have their defensible use, as an exposure layer attached to individuals whose service receipt and outcomes are observed in registers, and not as a generic proxy for regional economic activity, which measures no service and no return.

4. **Observing policy bundles:** The evaluation literature very often assesses one programme at a time, and so it estimates the average effect of an instrument across people who are receiving very different combinations of everything else. Multi-domain administrative infrastructures make the alternative feasible. Where a single database records employment, unemployment, transfers, education, healthcare use and pensions for the same individuals - as Hungary's Admin4 does for half the population, and as the Nordic registers do more generally - the portfolio of services and transfers an individual actually received can be observed.

8.2 Cross-country comparability

The main comparative challenge is whether the concept measured by a given register is comparable across countries, and not merely whether the country has such a register. A social-security employment record, a health insurance claim, or an EHR may have different definitions, coverage and institutional incentives. The recent harmonisation of French and German administrative data illustrates both the feasibility and the work required to create comparable datasets. Cross-country projects should therefore define common concepts and outcome variables before attempting to pool national microdata.

8.3 Access, privacy and representativeness

Administrative data reduce sampling error but can be inaccessible, institutionally fragmented or limited to events that generate an administrative record. Digital traces provide high-frequency information but may systematically exclude people who do not use the relevant technology or platform. Also, human-sourced and machine-generated data are, with some exceptions, not in public hands and their availability depends on commercial agreements. Health data are especially sensitive and subject to strict governance requirements. The EHDS establishes a European framework for controlled secondary use, but national access conditions remain important, and the implementation of the new framework is gradual.

The limitations of human-sourced and machine-generated data are, moreover, not only a matter of access. What these sources record is frequently a proxy whose relation to the quantity of interest has to be established rather than assumed. The presence of a mobile phone in the vicinity of a healthcare facility indicates proximity, not use: the device may belong to a visitor, an employee or a passer-by, and utilisation can only be inferred through behavioural rules, such as dwell time or repeated visits, that require validation against administrative records. Representativeness is a second concern, since phone ownership, operator market shares and internet use are unevenly distributed, and the groups with limited digital access, including many older people, are both the least likely to generate digital traces and among the main users of the services considered here.

9.Conclusions

The review indicates that the main opportunity offered by alternative data is the combination of complementary sources, which allows social-investment pathways to be observed over longer periods, across domains and, where possible, at greater spatial and temporal resolution. Process-mediated sources are particularly valuable for extending population coverage and follow-up, and for measuring outcomes such as employment, earnings, benefit receipt, education and healthcare utilisation. Human-sourced and machine-generated sources can complement them by capturing the intermediate mechanisms and exposures that neither registers nor surveys observe well: effective accessibility, mobility, environmental exposure, perceived service quality, and behavioural responses at high frequency.

However, with our review we find that process-mediated sources are where returns have actually been estimated. Human-sourced and machine-generated sources have, with very few exceptions, been used to measure something better rather than to estimate a return: online advertisements measure labour demand, search queries and social-media content nowcast unemployment, call detail records detect layoffs, floating catchment methods measure accessibility, patient reviews measure experience. In our opinion the frontier lies in those exceptions, where a human-sourced or machine-generated measure has been attached to a process-mediated backbone - vacancy text to employer–employee registers, platform search logs to unemployment registers, digitally delivered interventions to outcome registers.

Table 4 summarises, for each domain, the main message of the existing evidence, the data that are missing, and the avenue that appears most promising for further work.

Table 4: Evidence, gaps and priorities by policy domain

POLICY DOMAIN	MAIN MESSAGE ON EXISTING EVIDENCE	KEY MISSING DATA	MOST PROMISING AVENUE
ECEC and long-term care	Long-run returns demonstrated with linked registers (Norway, Germany, Italy, Spain); non-administrative sources measure perceived quality and accessibility.	Individual-level accessibility attached to personal records; individual participation records where provision is private; measures of service quality.	Catchment-based accessibility computed from GISCO and merged onto register records through the residential address, as an observable territorial conversion factor.
ALMP and life-long learning	Register-based evaluation; digitally delivered interventions evaluated in registers are the current frontier.	No reviewed study uses online vacancies, search queries or mobility traces to estimate a return; context measures and evaluation designs remain separate literatures.	Vacancy and platform-log measures of local labour demand and search behaviour used as conditioning information within register-based evaluations.
Health	The dominant design traces a health shock into employment and income rather than a health service into a return.	Health-service interventions with administratively dated introduction, coverage and eligibility rules, linked to employment, income and fiscal records.	Identification of an administratively dated service intervention in Hungary, Italy or Belgium, to which existing linked registers can be attached; EHDS over the medium term.

Source: authors' elaboration

The empirical evidence is most developed for care and ALMP, where process-mediated and linked data have been used to estimate long-term and heterogeneous effects of childcare provision, training programmes and employment-support interventions. Health presents a different picture. Rich health registers, claims and linked records exist in several consortium countries, and individual studies connect health events to subsequent employment, income and social-protection outcomes. But the dominant design traces a health shock rather than a health service.

Connecting health-service information to employment, income and fiscal data is therefore the clearest opportunity for further SWINS work. The country mapping outlined in section 7 identifies the regions where this initiative holds the greatest potential for implementation. First, Hungary, where the Admin4 system records healthcare utilization in conjunction with employment, transfers, education and pensions for half the population, and the linkage between employer-employee records and individual health records has already been demonstrated (Bíró, Boza, Gyetvai and Prinz, 2024). Then, Italy offers a comparable individual-level join in WHIP-Salute, though on the shock side rather than the service side, while Belgium holds the two halves of the design in separate infrastructures, with HISlink linking survey and health-insurance data at rates above 90% and the DWH LM&SP covering labour-market and social-protection trajectories.

In all three cases, the challenge is to identify a health service for which we can construct the counterfactual. The Salut Programme (Sampaio et al., 2023) is informative because the staggered municipal adoption provided the date and the boundary that the individual service record lacked. The primary step for SWINS is therefore the identification, within one of these countries, of a health-service intervention whose introduction, coverage and eligibility rules are administratively dated, to which the extant linked registers can then be attached.

The consortium countries also demonstrate substantial institutional diversity in the availability, accessibility and maturity of alternative-data infrastructures. Norway and Germany provide particularly mature environments for longitudinal register-based research. Belgium, Ireland and Hungary provide documented examples of cross-domain administrative linkage, while Spain and Italy offer important sectoral and linked-data infrastructures with different degrees of integration across domains. Serbia represents a different situation, with substantial administrative and digital health data but more pronounced constraints relating to interoperability, linkage and access. This heterogeneity creates an opportunity to identify which components of the social-investment return framework can be measured consistently across countries and which require country-specific data, linkage strategies or modelling assumptions.

The central implication is therefore to move from a search for the “best alternative dataset” towards a focus on the “best data architecture”. For each policy domain, this architecture should combine information on the policy intervention and service received with individual-level outcomes and, where possible, employment, income and fiscal information. The most promising designs combine administrative data with complementary survey, geospatial, mobility, environmental or other high-frequency sources. The European Health Data Space may increasingly facilitate such data reuse and linkage in health, but it should be regarded as an emerging European framework rather than as a currently harmonised individual-level database.

The final implication is addressed beyond the project. Much of the comparative weakness documented in this report is not a property of administrative data as such but of the uncoordinated way in which they are constituted. National registers are products of national administrative systems and the concepts they encode diverge in definition, coverage and recording practice in ways that no amount of ex-post harmonisation fully repairs. Yet the precedents assembled in this review suggest that coordination is achievable when it is attempted at the point of collection rather than afterwards: Cedefop's Skills-OVATE and Eurostat's Web Intelligence Hub have made online vacancy data comparable across Member States, GISCO supplies a common supply-side geography, and the European Health Data Space is now attempting the same for health. What is missing is an equivalent effort on the social-security and social-service side. Agreement among national social-security and employment institutions on a minimum common core — a small set of variables defined and recorded identically, with a shared reference period and a documented linkage key — would not require any institution to abandon its own administrative categories. It would, however, convert a landscape of nationally idiosyncratic records into one on which comparative evidence can accumulate.

References

Aaberge, R., Langørgen, A., and Lindgren, P. (2013). The distributional impact of public services in European countries (Discussion Paper No. 746). Statistics Norway.

Albanese, A., Cockx, B., and Dejemeppe, M. (2024). Long-term effects of hiring subsidies for low-educated unemployed youths. Journal of Public Economics, 235, 105137.

Ali, U., Herbst, C. M., and Makridis, C. A. (2021). The impact of COVID-19 on the U.S. child care market: Evidence from stay-at-home orders. Economics of Education Review, 82, 102094.

Almeida, V., Hoffmann, C., Königs, S., Moreno-Monroy, A. I., Salazar-Lozada, M., and Terrero-Dávila, J. (2024). Geographic inequalities in accessibility of essential services (OECD Social, Employment and Migration Working Papers No. 307). OECD Publishing.

Altmann, S., Glenny, A. M., Mahlstedt, R., and Sebald, A. (2022). The direct and indirect effects of online job search advice (IZA Discussion Paper No. 15830). Institute of Labor Economics.

Álvarez de Toledo Saavedra, P., Núñez Hernández, F., and Usabiaga Ibáñez, C. (2017). ¿Quién se empareja con quién en el mercado laboral español? Un análisis cluster basado en la Muestra Continua de Vidas Laborales. Investigación Económica, 76(299), 87–124.

Antenucci, D., Cafarella, M., Levenstein, M. C., Ré, C., and Shapiro, M. D. (2014). Using social media to measure labor market flows (NBER Working Paper No. 20010). National Bureau of Economic Research.

Askitas, N., and Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, 55(2), 107–120.

Audoly, R., Bhuller, M., and Reiremo, T. A. (2024). The pay and non-pay content of job ads (Federal Reserve Bank of New York Staff Reports No. 1124). https://doi.org/10.59576/sr.1124

Bachbauer, N., Wolf, C., and Fuß, D. (2022). Education and employment trajectories in NEPS-ADIAB: The survey data of the National Educational Panel Study linked to administrative data of the Institute for Employment Research. European Sociological Review, 38(4), 663–676.

Bakx, P., Wouterse, B., Van Doorslaer, E., & Wong, A. (2020). Better off at home? Effects of nursing home eligibility on costs, hospitalizations and survival. *Journal of Health Economics*, *73*, 102354.

Belot, M., de Koning, B. K., Fouarge, D., Kircher, P., Muller, P., and Phlippen, S. (2025). Advising job seekers in occupations with poor prospects: A field experiment (NBER Working Paper No. 33819). National Bureau of Economic Research.

Belot, M., Kircher, P., and Muller, P. (2019). Providing advice to jobseekers at low cost: An experimental study on online advice. The Review of Economic Studies, 86(4), 1411–1447. https://doi.org/10.1093/restud/rdy059

Berete, F., Demarest, S., Charafeddine, R., De Ridder, K. A., Van Oyen, H., Van Hoof, W., Bruyère, O., and Van der Heyden, J. (2023). Linking health survey data with health insurance data: Methodology, challenges, opportunities and recommendations for public health research. An experience from the HISlink project in Belgium. Archives of Public Health, 81, 198. https://doi.org/10.1186/s13690-023-01213-0

Biggeri, M., and Ferrannini, A. (2014). Sustainable human development: A new territorial and people-centred perspective. Palgrave Macmillan.

Bíró, A., Boza, I., Gyetvai, A., and Prinz, D. (2024). Health shocks, social insurance, and firms (IZA Discussion Paper No. 17069). Institute of Labor Economics.

Bíró, A., and Prinz, D. (2020). Healthcare spending inequality: Evidence from Hungarian administrative data. Health Policy, 124(3), 282–290.

Blazquez, D., and Domenech, J. M. (2018). Big Data sources, Big Data analytics and their impact on organisation performance. Information Systems Management, 35(2), 111–122.

Bonfanti, S., & Nordlund, M. (2012). *Does 'Swedish for Immigrants' (SFI) matter? A longitudinal assessment of the impact of SFI on migrants' position in the Swedish labour market* (Working Paper No. 24/2012). University of Florence, Department of Economics.

Boss, K., Longo, L., and Onorante, L. (2025). Nowcasting the euro area with social media data (arXiv:2506.10546). https://arxiv.org/abs/2506.10546

Boylan, A.-M., Turk, A., van Velthoven, M. H., and Powell, J. (2020). Online patient feedback as a measure of quality in primary care: A multimethod study using correlation and qualitative analysis. BMJ Open, 10(2), e031820. https://doi.org/10.1136/bmjopen-2019-031820

Brilli, Y., Del Boca, D., and Pronzato, C. D. (2016). Does child care availability play a role in maternal employment and children's development? Evidence from Italy. Review of Economics of the Household, 14(1), 27–51.

Bursens, F., Marchal, S., Verbist, G., (2026), "Review of macro-regularities about investment in social services. SWINS Working Paper n°5.1. SWINS – Sustainable Well-being through Investment in Social Services.

Caperna, G., Colagrossi, M., Geraci, A., and Mazzarella, G. (2022). A babel of web-searches: Googling unemployment during the pandemic. Labour Economics, 74, 102097. https://doi.org/10.1016/j.labeco.2021.102097

Carneiro, P., Løken, K. V., and Salvanes, K. G. (2015). A flying start? Maternity leave benefits and long-run outcomes of children. Journal of Political Economy, 123(2), 365–412.

Carta, F., and Rizzica, L. (2018). Early kindergarten, maternal labor supply and children's outcomes: Evidence from Italy. Journal of Public Economics, 158, 79–102.

Caruana, A., Bandara, M., Musial, K., Catchpoole, D., and Kennedy, P. J. (2023). Machine learning for administrative health records: A systematic review of techniques and applications. Artificial Intelligence in Medicine, 144, 102642.

Ciani, F., and Fattacciu, I. (2025). Theoretical framework for SWINS – Sustainable Well-being through Investment in Social Services (Deliverable D2.1). ARCO, research lab at PIN Foundation. www.swins.eu

Cockx, B., Lechner, M., and Bollens, J. (2023). Priority to unemployed immigrants? A causal machine learning evaluation of training in Belgium. Labour Economics, 80, 102306.

Coleman, J. S. (1990). Foundations of social theory. Harvard University Press.

Cooper, J. A., and Reid, S. C. (2026). Exploring the spatiotemporal dynamics of Google Trends data: An application in investigating internet search interest and child care demand. Spatial Demography, 14(1), Article 11. https://doi.org/10.1007/s40980-026-00164-8

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2018). Who benefits from universal child care? Estimating marginal returns to early child care attendance. Journal of Political Economy, 126(6), 2356–2409.

Daly, M., Groes, F., and Jensen, M. F. (2025). Skill demand versus skill use: Comparing job posts with individual skill use on the job. *Labour Economics*.

Dauth, C. (2020). Regional discontinuities and the effectiveness of further training subsidies for low-skilled employees. *ILR review*, *73*(5), 1147-1184.

De Craemer, M., De Decker, E., Verloigne, M., De Bourdeaudhuij, I., Manios, Y., and Cardon, G. (2014). The effect of a kindergarten-based, family-involved intervention on objectively measured physical activity in Belgian preschool boys and girls of high and low SES: The ToyBox-study. International Journal of Behavioral Nutrition and Physical Activity, 11, 38. https://doi.org/10.1186/1479-5868-11-38

De Craemer, M., Verloigne, M., De Bourdeaudhuij, I., Androutsos, O., Iotova, V., Moreno, L., Koletzko, B., Socha, P., Manios, Y., and Cardon, G. (2017). Effect and process evaluation of a kindergarten-based, family-involved cluster randomised controlled trial in six European countries on four- to six-year-old children's steps per day: The ToyBox-study. International Journal of Behavioral Nutrition and Physical Activity, 14, 116. https://doi.org/10.1186/s12966-017-0574-z

Emmert, M., Adelhardt, T., Sander, U., Wambach, V., and Lindenthal, J. (2015). A cross-sectional study assessing the association between online ratings and structural and quality of care measures: Results from two German physician rating websites. BMC Health Services Research, 15, 414. https://doi.org/10.1186/s12913-015-1051-5

Fennelly, O., Voisin, B., Olszewska, M., Corrigan, D., Moriarty, F., Fahey, T., Sarkar, D., Nagy, S., and Wong, S. (2022). DASSL "Data Access Sharing Storage and Linkage" proof-of-concept: Health and related data linkage in Ireland. International Journal of Population Data Science, 7(3). https://doi.org/10.23889/ijpds.v7i3.1908

Filser, A., Achard, P., Amend, I. M., Frodermann, C., Gaede, I., and Wagner, S. (2026). Harmonising French and German administrative DADS-EDP and SIAB data: A practical guide on how to create comparable data sets on maternal employment. Journal for Labour Market Research, 60(1). https://doi.org/10.1186/s12651-026-00430-8

Fondeur, Y., and Karamé, F. (2013). Can Google data help predict French youth unemployment? Economic Modelling, 30, 117–125. https://doi.org/10.1016/j.econmod.2012.07.017

Fransen, K., Neutens, T., De Maeyer, P., and Deruyter, G. (2015). A commuter-based two-step floating catchment area method for measuring spatial accessibility of daycare centers. Health and Place, 32, 65–73.

Gábos, A., Ciani, M., Tomka, B., and Biggeri, A. (2021). Reliability of EU-SILC and EU-LFS estimates underlying Social Scoreboard indicators.

Garritzmann, J. L., Häusermann, S., Palier, B., and Zollinger, C. (2021). The world politics of social investment: Volume I – Welfare states in the knowledge economy. Oxford University Press.

Goffin, K., Valsamis, D., Carpentier, M., Faes, K., and Verhaeghe, N. (2023). Onderzoek naar kinderopvang voor baby's en peuters en tewerkstelling in Vlaanderen: Eindrapport. IDEA Consult and HIVA–KU Leuven, in opdracht van het VIONA-onderzoeksprogramma, Departement Werk en Sociale Economie, Vlaamse overheid.

Goijaerts, J., van der Zwan, N., and Bussemaker, J. (2023). Health and the social investment state. Journal of European Public Policy, 30(5), 828–848.

Goller, D., Lechner, M., Pongratz, T., and Wolff, J. (2025). Active labor market policies for the long-term unemployed: New evidence from causal machine learning. Labour Economics, 94, 102729.

Groff, M., Latimer, E., Joober, R., Iyer, S. N., Schmitz, N., Abadi, S., Abdel-Baki, A., Casacalenda, N., Margolese, H. C., Jarvis, G. E., and Malla, A. (2021). Economic evaluation of extended early intervention service vs regular care following 2 years of early intervention: Secondary analysis of a randomized controlled trial. Schizophrenia Bulletin, 47(2), 465–473. https://doi.org/10.1093/schbul/sbaa130

Guida, C., Carpentieri, G., & Masoumi, H. (2022). Measuring spatial accessibility to urban services for older adults: An application to healthcare facilities in Milan. *European transport research review*, *14*(1), 23.

Havnes, T., and Mogstad, M. (2011). No child left behind: Subsidized child care and children's long-run outcomes. American Economic Journal: Economic Policy, 3(2), 97–129.

Havnes, T., and Mogstad, M. (2015). Is universal child care leveling the playing field? Journal of Public Economics, 127, 100–114.

Hedström, P., and Ylikoski, P. (2010). Causal mechanisms in the social sciences. Annual Review of Sociology, 36, 49–67. https://doi.org/10.1146/annurev.soc.012809.102632

Hemerijck, A. (Ed.). (2017). The uses of social investment. Oxford University Press.

Hendrikx, R. J. P., Drewes, H. W., Spreeuwenberg, M., Ruwaard, D., and Baan, C. A. (2019). Measuring regional quality of health care using unsolicited online data: Text analysis study. JMIR Medical Informatics, 7(4), e13053. https://doi.org/10.2196/13053

Hensvik, L., Le Barbanchon, T., and Rathelot, R. (2021). Job search during the COVID-19 crisis. Journal of Public Economics, 194, 104349. https://doi.org/10.1016/j.jpubeco.2020.104349

Herbst, C. M., Desouza, K. C., Alashri, S., Kandala, S. S., Khullar, M., and Bajaj, V. (2020). What do parents value in a child care provider? Evidence from Yelp consumer reviews. Early Childhood Research Quarterly, 51, 288–306.

Informal Working Group on Social Investment. (2024). Social investments for resilient economies (Joint ECOFIN-EPSCO Technical Note). Council of the European Union.
Khaef, S. (2026). Enrollment and patterns of participation in Swedish for immigrants (SFI): Timing, duration, and selectivity. *Genus, 82*, 11. https://doi.org/10.1186/s41118-026-00288-1

Kotavaara, O., Nivala, A., Lankila, T., Huotari, T., Delmelle, E., and Antikainen, H. (2021). Geographical accessibility to primary health care in Finland: Grid-based multimodal assessment. Applied Geography, 136, 102583. https://doi.org/10.1016/j.apgeog.2021.102583

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. Science, 343(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Lazuka, V. (2023). Household and individual economic responses to different health shocks: The role of medical innovations (IZA Discussion Paper No. 16226). Institute of Labor Economics.

Lim, Y.-H., Bilsteen, J. F., Mortensen, L. H., Loft, S., Jørgensen, J. T., Cole-Hunter, T., Napolitano, G., So, R., Lund, K. L., Andersen, Z. J., and Amini, H. (2024). Lifetime exposure to air pollution and academic achievement: A nationwide cohort study in Denmark. Environment International, 185, 108500.

Llorente, A., Garcia-Herranz, M., Cebrián, M., and Moro, E. (2015). Social media fingerprints of unemployment. PLOS ONE, 10(5), e0128692. https://doi.org/10.1371/journal.pone.0128692

López, M. A., Benavides, F. G., Alonso, J., Espallargues, M., Durán, X., and Martínez, J. M. (2014). The value of using administrative data in public health research: The Continuous Working Life Sample. Gaceta Sanitaria, 28(4), 334–337. https://doi.org/10.1016/j.gaceta.2014.02.010

López Gómez, M. A., Durán, X., Zaballa, E., Sanchez-Niubo, A., Delclos, G. L., and Benavides, F. G. (2016). Cohort profile: The Spanish WORKing life Social Security (WORKss) cohort study. BMJ Open, 6, e008555. https://doi.org/10.1136/bmjopen-2015-008555

Lovaglio, P. G., and Riussi, M. (2025). How is labour demand changing across European regions in the post-COVID-19 era? International Regional Science Review, 48(5–6), 524–575. https://doi.org/10.1177/01600176251327985

Marical, F., d'Ercole, M. M., Vaalavuo, M., and Verbist, G. (2006). Publicly-provided services and the distribution of resources (OECD Social, Employment and Migration Working Papers No. 2006/14). OECD Publishing.

Marinescu, I., and Skandalis, D. (2021). Unemployment insurance and job search behavior. The Quarterly Journal of Economics, 136(2), 887–931. https://doi.org/10.1093/qje/qjaa037

Matsaganis, M., and Verbist, G. (2009). Distributional effects of publicly funded childcare. In European inequalities: Social inclusion and income distribution in the European Union (pp. 177–185). TÁRKI.

Mazzone, A., De Cian, E., Falchetta, G., Jani, A., Mistry, M., & Khosla, R. (2023). Understanding systemic cooling poverty. *Nature Sustainability*, *6*(12), 1533-1541.

Midtsundstad, T., and Nielsen, R. A. (2019). Lifelong learning and the continued participation of older Norwegian adults in employment. *European Journal of Education*, *54*(1), 48-59.

Mulero, R., and García-Hiernaux, A. (2021). Forecasting Spanish unemployment with Google Trends and dimension reduction techniques. SERIEs: Journal of the Spanish Economic Association, 12(3), 329–349. https://doi.org/10.1007/s13209-021-00231-x

Muñoyerro-Muñiz, D., Goicoechea-Salazar, J. A., García-León, F. J., Laguna-Téllez, A., Larrocha-Mata, D., and Cardero-Rivas, M. (2020). Health record linkage: Andalusian health population database. Gaceta Sanitaria, 34(2), 105–113.

Naccarato, A., Falorsi, S., Loriga, S., and Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. Technological Forecasting and Social Change, 130, 114–122. https://doi.org/10.1016/j.techfore.2017.11.022

Napierala, J., Kvetan, V., and Branka, J. (2022). Assessing the representativeness of online job advertisements (Cedefop Working Paper No. 17). Publications Office of the European Union. https://doi.org/10.2801/807500

OECD, Department of Social Protection Ireland, and European Commission Joint Research Centre. (2024). Impact evaluation of Ireland's active labour market policies. OECD Publishing. https://doi.org/10.1787/ec67dff2-en

Pappalardo, L., Vanhoof, M., Gabrielli, L., Smoreda, Z., Pedreschi, D., and Giannotti, F. (2016). An analytical framework to nowcast well-being using mobile phone data. International Journal of Data Science and Analytics, 2(1), 75–92.

Pecoraro, F., Cellini, M., Luzi, D., and Clemente, F. (2024). Intra- and interregional components of a spatial accessibility gravity model to assess patient mobility. BMC Health Services Research, 24, 973. https://doi.org/10.1186/s12913-024-11411-3

Pennerstorfer, A., and Pennerstorfer, D. (2021). Inequalities in spatial accessibility of childcare: The role of non-profit providers. Journal of Social Policy, 50(1), 122–147.

Pintér, G., and Felde, I. (2022). Commuting analysis of the Budapest metropolitan area using mobile network data. ISPRS International Journal of Geo-Information, 11(9), 466. https://doi.org/10.3390/ijgi11090466

Plavgo, I., and Hemerijck, A. (2021). The social investment litmus test: Family formation, employment and poverty. Journal of European Social Policy, 31(3), 282–296. https://doi.org/10.1177/0958928720950627

Sampaio, F., Häggström, J., Ssegonja, R., Eurenius, E., Ivarsson, A., Pulkki-Brännström, A.-M., and Feldman, I. (2023). Health and economic outcomes of a universal early intervention for parents and children from birth to age five: Evaluation of the Salut Programme using a natural experiment. Cost Effectiveness and Resource Allocation, 21, 33. https://doi.org/10.1186/s12962-023-00439-7

Serrano-Alarcón, M., Hernández-Pizarro, H., López-Casasnovas, G., and Nicodemo, C. (2022). Effects of long-term care benefits on healthcare utilization in Catalonia. *Journal of Health Economics*, *84*, 102645.

Simonetti, I., Belloni, M., Farina, E., and Zantomio, F. (2022). Labour market institutions and long term adjustments to health shocks: Evidence from Italian administrative records. Labour Economics, 79, 102277.

Štefánik, M., Lyócsa, Š., and Bilka, M. (2023). Using online job postings to predict key labour market indicators. Social Science Computer Review, 41(5), 1630–1649. https://doi.org/10.1177/08944393221085705

Toole, J. L., Lin, Y.-R., Muehlegger, E., Shoag, D., González, M. C., and Lazer, D. (2015). Tracking employment shocks using mobile phone data. Journal of the Royal Society Interface, 12(107), 20150185. https://doi.org/10.1098/rsif.2015.0185

UNECE. (2013). Big data and modernisation of official statistics. United Nations Economic Commission for Europe.

van den Brekel, L., Lenters, V., Mackenbach, J. D., Hoek, G., Wagtendonk, A., Lakerveld, J., Grobbee, D. E., and Vaartjes, I. (2024). Ethnic and socioeconomic inequalities in air pollution exposure: A cross-sectional analysis of nationwide individual-level data from the Netherlands. The Lancet Planetary Health, 8(1), e18–e29.

Van Havere, T., and Verbist, G. (2026). The value of social services (SWINS Working Paper No. 4.2).

Verbist, G., Förster, M., and Vaalavuo, M. (2012). The impact of publicly provided services on the distribution of resources: Review of new results and methods.

Verbist, G., and Matsaganis, M. (2014). The redistributive capacity of services in the European Union. In Reconciling work and poverty reduction: How successful are European welfare states? (pp. 185–211).

Xia, J., Ye, P., Pan, L., et al. (2026). Mobile phone data show spatial and socioeconomic inequalities in hospital utilization. Nature Health, 1, 760–769. https://doi.org/10.1038/s44360-026-00100-6

Evidence note

This deliverable deliberately distinguishes between (i) data sources documented by official institutions, (ii) data linkages demonstrated in published studies, and (iii) future research potential. Claims in the country mapping are not intended to imply that all listed datasets are openly accessible, nationally comprehensive, interoperable, or directly linkable to all other sources. Where the present evidence does not establish one of these properties, the text does not assume it.

SWINS Sustainable Well-being through INvestment in Social Services

SWINS
Sustainable Well-Being through Investments in Social Services

sustainablewellbeing.eu

