Chapter 1 – Assessing the impact of private-sector investment funds in developing countries[1]

Authors: Shahrokh Fardoust (Research Professor, Global Research Institute, College of William and Mary, United States), **Pierre Jacquet** (President of the Global Development Network, member of the French *Cercle des Economistes*) and **Matt Ripley** (The Good Economy)

JEL classification: A11, C10, D61, F63, G24, H43, M14, O16

Keywords: impact, impact investment, impact evaluation, impact measurement, impact management, theory of change, causality, ESG, data collection, quantitative and qualitative methods, deep dives

Email: sfardoust@WM.edu (S. Fardoust); pjacquet@gdn.int (P. Jacquet); matt@thegoodeconomy.co.uk (M. Ripley)

[1] Research assistance from Alexis Atlani and Brendan Harnoys-Vannier is gratefully acknowledged.

Summary and key messages

This chapter examines impact investing – investment in private companies to generate socially desirable results in addition to financial returns – and the assessment of its effectiveness. Impact investing is growing fast, reflecting rising interest from private investors in their contribution to societal goals. Impact investments have become increasingly important both in responding to societal demands and in contributing to the investments needed for meeting the Sustainable Development Goals (SDGs), when other sources of financing fall short of the massive financing required.

The growth of impact investing is focusing attention on impact measurement and the need for credible impact studies. Generating impact mobilises a chain of stakeholders – funders, intermediaries and investees – and knowledge brokers who monitor, measure and guide processes. Each stakeholder has their own definition of, interest in, and contribution to impact. The private sector, which focuses on profitability, would also typically consider potential trade-offs between financial and societal returns. The empirical jury is still out on the exact nature of such trade-offs. However, there are various types of investor within the private sector, not all of which focus on competitive financial returns. This leads to the possibility of blended finance in which, for example, public money motivates private investors, under the promise of a commercial return, towards producing public goods.

Monitoring and evaluating the impact of impact investing involves significant challenges in defining a logical chain from inputs in an activity to outputs, outcomes and impacts. Measuring impact is complex and multidimensional, and discussing causality and attribution requires time, resources and capacity. There is no single optimal method: the choice of an appropriate method depends on resources, capacity and data availability. In practice, impact investors use a variety of approaches, try to improve clarity and credibility by adopting common standards and frameworks, and increasingly manage impact. However, there remains a research gap across investment studies.

European Investment Bank (EIB) and Global Development Network (GDN) deep dives are impact studies that try to make research-based approaches more palatable for all stakeholders and fill a gap between impact reports and fully fledged scientific impact evaluations. They combine methodological rigour with the time and operational constraints of the impact investing stakeholders, especially investees who produce goods and services generating impact.

Three main messages emerge from this chapter. First, impact studies mean different things to different stakeholders. There is a gap between the private value of impact studies (which focus on accountability, communication and targeted business-oriented learning) and their social value (which relies on the knowledge created about impact and how it can be generated). Bridging that gap requires analytical efforts, external incentives and resources. Second, rigorous academic research is important at all stages of the investment process. Evaluations should continue to consider a range of approaches and models, and evaluators should cultivate stronger connections with academia. Third, increasing cooperation with academia for impact assessments has a cost that is unlikely to be covered by impact investors themselves. To support the market for impact investing and social returns, efforts are needed to promote research-based evaluations that insist on rigour and control of biases, while considering private stakeholders' interests and motivations.

This chapter examines impact investing – investment in private companies intended to generate socially desirable results in addition to financial returns (Box 1) – and the assessment of its effectiveness. Such investment reflects rising interest on the part of private investors about their contribution to societal goals[2]. As such, impact investments have become increasingly important both in responding to societal demands and in contributing to the investments needed for meeting the Sustainable Development Goals, when other sources of financing fall short of the massive financing required (United Nations IATF, 2020).

Section 1 of this chapter briefly sets the context. Section 2 reviews some issues and challenges pertaining to impact investing. Section 3 discusses the measurement of impact. Section 4 introduces the rationale for the EIB-GDN deep dive programme.

[2] We use the word "societal" as a shortcut to encompass social and environmental dimensions.

1. The rapid rise of impact investing

Achieving the Sustainable Development Goals will require massive investment. In 2014, the United Nations Conference on Trade and Development (UNCTAD) estimated that the total investment needs of emerging market and developing countries could be as high as $3.9 trillion a year, only about $1.4 trillion of which could be financed from domestic revenue sources, official external aid, and private-sector financial flows and investments (UNCTAD, 2020). Given the magnitude of the gap, increases in official financial flows – which account for less than 5% of financing needs – must be supplemented by increases in developing countries' own domestic resources and by a substantial increase in the private sector's funding of and investments in the SDGs.

The situation has become even more challenging since those estimates were made. As a result of the COVID-19 shock to the world economy, financial flows to emerging markets and developing countries may decline by as much as $700 billion in 2020 compared to the level in 2019 (OECD, 2020). The negative impact of the pandemic on the financial flow to developing countries has been at least 60% larger than the impact of the global financial crisis in 2008 to 2009.

Box 1: The origins of impact investment

Recognition of the role of the private sector with respect to social and environmental goals came gradually. A first step was the observation that production processes as well as the final good or service affect society, and that invested money should not support activities that society considers immoral or unethical. Avoiding unethical or welfare-reducing activities was the approach of the first socially responsible investing mutual funds and the establishment of development finance institution financing of the private sector in developing countries – for example, by the CDC Group or the International Finance Corporation (IFC) in the 1950s (O'Donohue et al., 2010).

The environmental movement in the 1970s and the movement for divesting from apartheid South Africa in the 1980s followed a similar logic (Bugg-Levine and Emerson, 2011; IFC, 2019). This evolution towards corporate social responsibility was concerned more with ethics and avoiding negative impact than with actively pursuing specific social and environmental objectives.

Barman (2015) dates the "invention" of impact investing to about 2007, when the Rockefeller Foundation noted the gap between the trillions of dollars of social and environmental needs around the world and the resources committed by global philanthropy, which amounted to only about $500 million at the time. Several market-based solutions – including microfinance, community development and clean technology (Rockefeller Philanthropy Advisors, 2009; Monitor Institute, 2009; O'Donohue et al., 2010) – were being promoted at the time; however, these initiatives were not coordinated around an objective of societal impact, and investors did not include the pursuit of such impact in their objective functions. The Rockefeller Foundation brought together experts and mainstream investors to explore the conditions of their involvement. One conclusion reached was that such involvement needed to be based on evidence. From the start, the requirement of assessing impact was an intrinsic part of the impact investing concept.

The importance of assessing investment activities that involve both financial and social dimensions is reflected in the core characteristics that define impact investing (GIIN, 2019):

- Intentionally contribute to social and environmental impact and earn a financial return
- Use evidence and impact data to design the investment
- Manage impact performance
- Contribute to the growth of impact investing

Recognising this need, international donors have increasingly focused on new and innovative financial instruments to support private-sector investments in developing countries. The rationale – that there are market failures that prevent efficient private investment – is twofold. First, there may be information failures that increase the perception of risk in SDG-related investments in developing countries. Second, there are missing markets for social and environmental benefits, and public money may be needed to orient investor choices. A key challenge is that of "additionality," or the idea that donor intervention should not be a substitute for private choices that would have been made otherwise. In recent years, impact investing has emerged as a potential bridge between the financing requirements of the SDGs and official development assistance.

However, impact investing is also under a market-driven dynamic supported by increasing demand for societal benefits from citizens, consumers, workers, investors, advocacy groups and governments. A new class of investors has emerged who jointly focus on financial profitability or sustainability and a positive contribution to society. Impact investing thus appears both as a market-led trend on international financial markets and as a channel to stimulate international donors' efforts, which often reinforce each other.

These investors cover a spectrum of organisations from family offices and foundations, to large banks and investment houses that are increasingly entering into the space. The term "investor" may include asset owners (for example, pension savers), asset allocators (for example, financial advisors) and asset managers (those managing investment funds and client accounts). A simple distinction can be made between organisations like commercial entities, which invest with impact (where the primary focus is on financial rate of return alongside social impact) and those investing for impact (where the primary focus is on achieving social impact) (West, 2019). Major players in the investing for impact category are development banks that specialise in providing risk capital and supporting private-sector projects in developing countries. They can be either national institutions that are wholly or partly owned by governments, known as bilateral development financel institutions (DFIs), or private-sector arms of multilateral development banks (MDBs) that have been established by more than one country. These multilateral development banks – or dedicated organisations within them, for example the International Finance Corporation (IFC) as part of the World Bank Group – extend loans and equity investments to private firms in developing countries. They do this by investing in fund managers (so-called "intermediated" investments) or by providing capital "directly" to companies.

Impact investing has proven to be fairly resilient to the effects of the COVID-19 shock, particularly compared with foreign direct investment and portfolio investment inflows into developing countries (Kharas, 2020). It currently accounts for less than 10% of total net flows to developing countries. Size estimates vary among sources. The Organisation for Economic Co-operation and Development (OECD) analysed a portfolio of $70 trillion assets under management (OECD, 2020); around $17.5 trillion met the environmental, social and governance (ESG) criteria, while only a small fraction of these assets (around $444 billion) was labelled as impact investing. Using another methodology and database of

1 720 impact investors, the Global Impact Investing Network (GIIN) estimated the size of the impact investing market in 2020 at $715 billion (GIIN, 2020a).

The successful expansion of the private impact investing market appears to be a crucial component of the SDGs, particularly in developing countries. However, it fuels questions about how impact should be assessed and draws increased attention to impact measurement within the development finance community and the group of impact investors (see, for example, GIIN 2019). Even among international development organisations that have spent decades grappling with the issue of impact, according to a report by the Overseas Development Institute, "there is too much ambiguity and confusion about what 'impact' is, how it should be defined, how to measure it and what kind of measurement is sufficient" (Hearn and Buffardi, 2016).

2. Pursuing impact: Concepts, stakeholders and motivations

2.1. What is impact?

Just as investors have diverse financial return expectations, different definitions of impact shape diverse attitudes to seeing evidence of impact "returns." The various ways in which investors conceptualise impact give "different weight to the importance of causality between an intervention and its effects and the measurability, breadth and timeframe of those effects" (IMP, 2020a). These expectations and definitions shape what investors mean by measuring impact. In practice, however, impact investors use the term "impact" in one of three ways[3].

First, impact refers to any type of change. Under this definition, most investments have direct and indirect, intended and unintended impacts on consumers and users, the environment and society at large (OECD, 2019). Impact is a synonym for any kind of positive or negative change, referring both to specific "outputs" (such as a company's products, services and practices) as well as "outcomes" (the effect of those products, services and practices on people and planet). This is the most general use of the term impact – a catch-all that many investors have now added to their financial lexicon to signify that a non-financial element is under consideration; for example, the "impact of a real estate investment" (IMP, 2020b) or the impact of a microfinance institution on a village in India.

Second, impact as contributions to higher-level change. In this definition, impact "refers to change occurring in communities or systems" (So and Staskevicius, 2015). Impact is the intended changes at the highest level of a theory of change or the top of a results framework. As defined by the International Initiative for Impact Evaluation (3ie), impact is "how an intervention alters the state of the world … [so] impact evaluations typically focus on the effect of the intervention on the outcome for the beneficiary population" (White, 2013). This framing is often used by practitioners and project managers emerging from the world of international development, and wider social science. In other words, it describes how projects are contributing to longer-term outcomes concerning poverty, prosperity and well-being, for example the impact of a health fund on mortality and morbidity.

Third, impact as causal change. In this definition, impact is the increase in the "quantity or quality of the enterprise's social outcomes beyond what would otherwise have occurred" (Brest and Born, 2013). This can be termed the more academic definition of impact, which is at the heart of the growing field of impact evaluations that seek to assess the changes that can be attributed to a particular project,

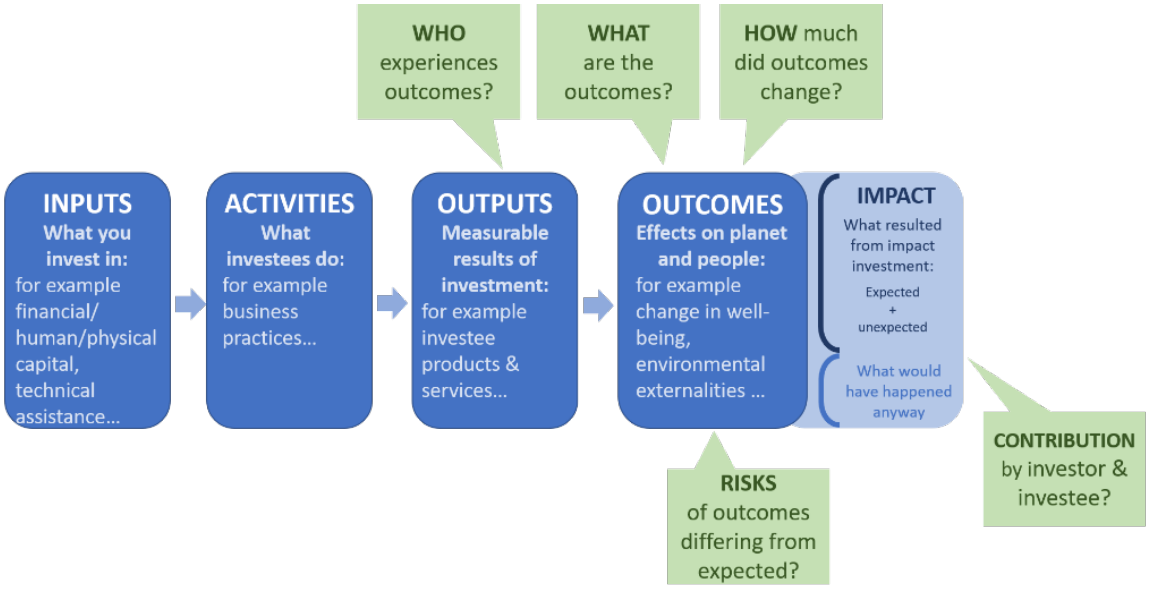
[3] For more on this discussion see Hearn and Buffardi (2016).

programme or policy[4]. Such evaluations seek to predict the outcome had the investor and enterprise not done what they did (the "counterfactual" scenario)[5]. Impact is therefore a specific change that is caused by the investment; for example, the additional profit accruing to small enterprises as a result of accessing a microfinance loan.

Many of the fundamental challenges – and confusions – about impact measurement can be traced to the difficulty of articulating what "impact" means and the different ways in which the term is defined (Thornley and Locascio, 2018). One investor may use the term in a narrow sense while others may use a broader definition. This can cause misunderstandings about the purpose and scope of impact assessments and evaluations. Even among multilateral development banks there are a range of uses: some see impact as a synonym for "results," while others see it as an academic term for counterfactual scenarios, or for the indirect and induced effects beyond outcomes (EIB, 2019).

To move beyond definitional debates and examine what impact looks like in practice, a collaborative effort by over 700 organisations under the auspices of the Impact Management Project (IMP) produced a set of shared fundamentals for measuring, managing and communicating impact. This deconstructs impact into five dimensions: understanding what outcomes are contributed to; who experiences those outcomes; how much those outcomes affect people or planet; what the contribution was to impact; and what risks exist to impact being different than expected. These IMP dimensions are set out against a simplified results chain (Figure 1)[6].

Figure 1: The impact value chain and the dimensions of impact



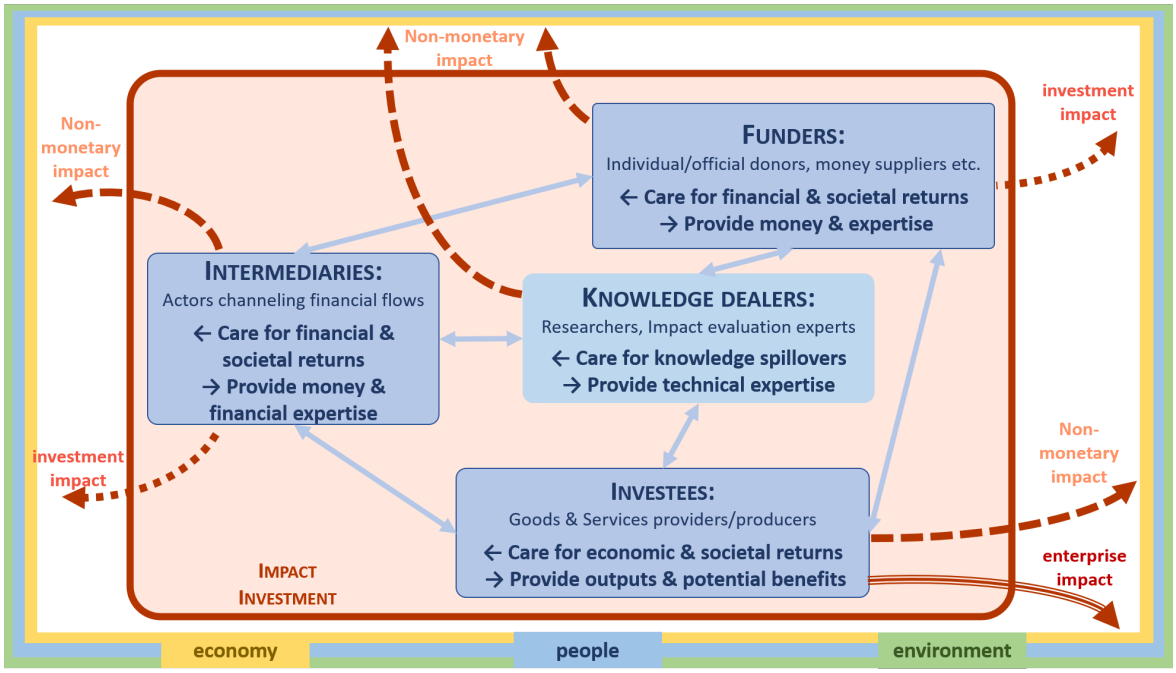
Source: adapted from Clark et al. (2004) and IMP (2019)

[4] World Bank.
[5] IMP contribution.
[6] In the W.K. Kellogg Foundation 2004 "logic model," outcomes refer to effects on programme participants, while impacts describe intended or unintended changes beyond participants in the society at large within seven to ten years as a result of programme activities.

2.2. Impact from what or whom?

Who creates impact and through what actions? The generation of impact involves a chain of stakeholders (Figure 2). Funders including individual and official donors as well as asset holders provide financial resources either directly to investees who provide impact-producing goods and services, or to financial intermediaries who reach out to investees. Brest and Born (2013) distinguish three categories of impact generation: enterprise impact, investment impact and non-monetary impact (Figure 2).

Figure 2: Stakeholders in impact investment



Source: prepared by the authors

Enterprise impact, the most straightforward (once a credible counterfactual is considered), is "the social value of the goods, services or other benefits provided by the investee enterprise." This is the category of impact that evaluations mostly consider (and this is the focus of the EIB-GDN deep dives discussed later).

Investment impact is "a particular investor's financial contribution to the social value" created by the enterprise in which the investment takes place (Brest and Born, 2013). There are several routes through which an investor may generate impact: he or she may provide additional capital, offer a lesser cost (through concessional financing or better terms over longer time horizons with greater flexibility), or present different sets of incentives. These may be organised through different financial instruments (debt, equity and guarantees) suggesting that financial structuring matters for impact and is not neutral.

Non-monetary impact "reflects the various contributions, besides financial resources, that investors, fund managers and other parties may make to the enterprise's social value" (Brest and Born, 2013). This can be interpreted as the additional impact unlocked through explicit means, such as the provision of technical assistance, and implicitly through values and demonstration effects. A 2016 study funded

by the Danish government identified the following types of non-monetary impact (Koening and Jackson, 2016):

- Signalling – providing a stamp of approval, providing credibility, attracting other investors, acting as an honest broker.
- Knowledge – strengthening the quality of the investment model and technology, sharing knowledge, building the capacity of local partners, facilitating technology transfer, publicly sharing experiences and learning (beyond project boundaries).
- Demonstration – supporting innovative pacesetters to de-risk new business models, attracting capital in lower income countries and frontier markets that are not (yet) able to attract significant levels of commercial capital.
- Aggregation – supporting projects at regional or global level for aggregation of opportunities, diversification of risk and cross-boundary sharing of experience.
- Standards – promoting high environmental, social and governance standards in investee companies, financial institutions, funds and at industry level.

Non-monetary impact can have "systemic effects" beyond the individual investee companies to influence wider sector and market performance. As Desai et al. (2017) note in their research, "individual firm activity can make markets in developing countries work better through improved competition, demonstration effects of a profitable business model that may not have been tried before in that context, or by spurring sector wide policy, legal, or regulatory reforms."

Figure 2 adds another category of stakeholders, "knowledge brokers," to recognise that producers of impact-related knowledge also belong to the impact generation chain. Impact assessments inform project design, implementation and results. The knowledge they create can inform project implementation as well as future investments, and can lead to better impacts over time and a better understanding of the relationship between activity and impacts. Evaluative research should be considered part of impact investing through the mobilisation of measurement systems, and the use of research to document externalities and inform the complex network of economic, social and environmental impacts. This is one of the rationales behind the EIB-GDN deep dives.

2.3. Does impact imply below-market rates of return?

Brest and Born (2013) ask whether investors can "both make money and make a difference." One argument is that without making enough money to cover costs and invest in development, any difference they might have is likely to be unsustainable. The issue is the kind of financial return that impact investing can achieve, and the compatibility between maximising shareholder value through financial returns (Friedman, 1970) and targeting stakeholder value. Decisions to invest for impact are supposed to take place in private markets where investments respond to and compete on profitability requirements, and private returns need to be made commensurable. This relates to a central dilemma in welfare economics: perfect markets lead to pareto optima but generally not to social optima (and imperfect markets lead to neither). Could impact investing bridge the two[7]?

From a conceptual perspective, this may seem doubtful. Given the historical philanthropic inspiration, expectations of potential impact investors about financial outcomes may be less than for regular

[7] Pareto optima deal with effectiveness: it is not possible to improve the situation of one actor without deteriorating that of another one. This clearly does not imply a socially desirable (not even mentioning optimum) allocation of resources.

investments (for example buying stocks or bonds to pursue profits). Why would adding objectives that are difficult to capture privately and that may incur additional costs not affect financial returns? Answers include benefits in terms of image, satisfaction and fidelity of the consumer base; returns in terms of capacity to innovate in new and promising market segments; and discovery of profitable investments that would have been overlooked. Visionary impact investors can then allocate resources to companies and push them toward impact-creating decisions[8]. Accordingly, "impact investing" would be good, competitive investing. Fundamentally, however, most environmental and social benefits have characteristics of public goods, however imperfect, and are therefore underproduced through market forces alone. Producing them should not be expected to systematically generate commercial returns.

Causality between commercial and societal returns can go both ways: environmental and social performance may contribute to profitability, and profitable firms may be better able to engage in ESG or impact investing (Desai et al., 2017). Ultimately, any trade-off between financial and non-monetary returns is an empirical question. However, the empirical jury is still largely out (Box 2).

If and when markets do not assign financial values to societal impacts, there must be some outside willingness to "pay" for them. The financial return of an impact investment can be abstractly thought of as an algebraic combination of a market return (for a typical investment) and the financial value assigned to the expected societal impact. Impact investments can be seen as synthetic vehicles that combine a traditional investment, producing market returns, and a commitment to societal impact, which has a market value that the investor or a third party may be willing to pay for. This provides an original model for thinking of impact investments as a virtual combination of public or private philanthropy – or socially oriented investors – and market-based investment decisions. This addresses the potential disconnect between financial returns and societal impacts by allocating them to different types of investor. Chowdhry et al. (2019) studied the joint financing between profit-motivated and socially motivated (impact) investors and argued that the latter's financial claims counterbalance the former's focus on profits. They describe a continuum of impact investments, from pure non-profit status for higher valued social projects, to pure commercial investment (when impact is fully compatible with market returns); typical impact investment thus becomes a contingent social contract between socially motivated and commercially motivated investors. For such a model to make sense, however, more knowledge is needed about societal impact and how to measure it and its value. Blended finance experiments have taken place in which public money (or money from philanthropic organisations, development finance institutions or individuals) nudges private investors, under the promise of a commercial return, towards producing public goods. This combination of differently motivated investors might create opportunities for new and promising models of public-private partnerships.

[8] The example of "fair trade" labels shows that companies or consumers are willing to buy their inputs and products from providers who have fair labour practices, environmental safeguards, and so on, and that willingness can be reflected in accepting a significant premium. A Swedish study, for example, claims that the premium could be up to one-third of the price of a good (Schollenberg, 2011).

Box 2: Is there a trade-off between financial and societal returns?

Limitations of empirical studies include the following: uncertainty about what is covered by, and measured as, impact; endogeneity; selection biases; different characteristics of investment (in terms of sector, geography, timing, and so on); and diverging investors' expectations and objectives (between development finance institutions and private impact funds, for example). The dataset and sample of investors may be incomplete; it may include types of investment that belong to "softer" categories of sustainable, responsible or ESG investments; it may be biased by the fact that only willing respondents take part; it may fail to differentiate between investors looking for market returns and those satisfied with below-market returns; and there may also be omitted variables, such as the quality of fund managers.

Results on impact investment returns are mixed: some impact investors have achieved commercial returns, others have not. Desai et al. (2017) show that financial returns depend on the sector rather than geography and country income level. Rangarajan (2019) shows that public market investments yield different returns across different "impact" sectors. There may not be a general relationship between financial and social returns. Another empirical question relates to the sustainability of societal and financial returns. For example, the growth of societal benefits (through impact investing or ESG) may lead to declining societal productivity, which might translate into lower financial returns and remove any competitive advantage from impact investing.

Reviewing more than 2 000 empirical studies, Friede et al. (2015) find that roughly 90% reveal a non-negative relationship between non-financial (as measured by ESG) and financial performance, and that the large majority find a positive relationship. Moreover, the positive relationship appears to be stable over time. On average, the IFC's equity projects in impact investing have achieved financial results in line with or better than the evolution of the Morgan Stanley Capital International (MSCI) Emerging Market Index over 1988-2016, proving the possibility of being financially sustainable over long periods while investing for impact (IFC, 2019). These results suggest that it may be possible to invest for social impact while achieving reasonable financial returns. In the GIIN's 2019 annual impact survey, Mudaliar et al. (2019) find that most investors' financial performance was in line with expectations (14% over-performing, 77% performing as expected, and 9% underperforming). Using survey data, Castellas, Findlay and Addis (2016) find that financial returns met respondents' expectations (but an important caveat is that expectations are declared subsequently and are self-reported; investors may try to improve their image). Other studies, summarised by the GIIN, report average returns for impact funds seeking market rates and show that market rate-seeking impact investments can earn market-like returns in certain instances. Nofsinger and Varma (2014) find that socially responsible mutual funds outperform the market during crises and underperform during non-crisis periods.

However, Desai et al. (2017) find that that there is no causal link between ESG and lower profits or any evidence that better ESG performance leads to higher financial performance. Using industry data and controlling for vintage year, fund size, sequence and geography, Barber et al. (2019) find that the internal rate of return of impact funds was 4.7% lower than that of all venture capital. They infer that investors in impact funds are willing to accept lower returns than traditional investors.

Kovner and Lerner (2015) find that community development venture capital funds are less likely to go public or to be acquired than traditional venture capital funds. A sample of social investments by the UK government yielded a -9.2% financial return (SIRC, 2015).

3. Assessing the impact of impact investing: Issues, methods and practice

3.1. How to assess the impact of impact investing

A hallmark of impact investing is the commitment of the investor to measure and report the social and environmental performance of their underlying investments (GIIN, 2020b). Unlike purely financial investments which have outcomes that are more easily assessed relative to markets, impact investments have benchmarks on which there is less consensus. Monitoring and evaluating the results of projects has been a core business activity of development banks for decades, drawing on well-established methods in the fields of development finance and international development more widely. These activities are usually housed in an evaluation function that is independent of the line management.

What final outcomes should be considered and measured in determining whether an investment has impact and how should they be judged? The objective behind impact investing seems to point towards effects that are beneficial to groups of individuals or to society. However, the notion of benefits is multipronged and subjective: one classically distinguishes social, economic and environmental benefits (see, for example, Bugg-Levine and Emerson, 2011); there may be debate on whether a given effect is a benefit (and for whom); there may be tensions and contradictions between outcomes in terms of positive value for society; and the overall effect (Bugg-Levine and Emerson's "blended value"), which is the net result of a variety of partial specific effects along various dimensions, cannot be directly observed and must be broken down into elementary effects that can be measured and interpreted within given value scales. How can these elementary effects be combined to get an overall sense? Such aggregation implies a scale of values that may differ among individuals, groups and societies, and be unstable over time. This value scale is not explicit at any given time. The aggregation problem involves questions of summation (how to add disparate elementary effects) and of scale of priorities (how to consider "positive" effects along some dimensions and "negative" effects across others). Impact assessments may be thought of as providing packages of disparate pieces of information that only ongoing public debate can sort out and transform into a tentatively, and possibly temporarily, shared perception of any overall "impact."

Shared perceptions may evolve over time, not maturing into any stable consensus. This might translate into difference and even incompatibility between the prior expectations of the benefits of the intended impact and the subsequent perception of their value, even when the intended impact has been achieved, because the focus and value perspectives may differ before and after. The aggregation challenge introduces the inevitability of a time consistency problem in impact assessment.

Impact assessments should document specific impacts – negative or positive – and avoid providing aggregate synthetic judgments. Ruff and Olsen (2018) argue that impact commensuration should be separated from the description of what happened. Impact assessments require further interpretation and judgment to evaluate whether impact investing has benefited society.

3.2. Broad challenges

Recent initiatives to increase the evidence base about impact often use experimental research techniques to isolate precise effects on incomes, livelihoods, health and well-being outcomes. Efforts have remained limited in scale and coordination. "Much of the industry's reporting on impact has relied heavily on individual stories of successful entrepreneurs, enterprises, or employees, accompanied by input and output data (e.g., dollars invested, numbers of investee businesses, number of jobs created, etc.)," writes Jackson (2013a, p. 610). What is the benchmark for assessing impact? Reeder and Colantonio (2013) list nine challenges that evaluations should address[9]. For simplicity, they are reinterpreted here and grouped under three headings.

Identifying outcomes and moving to impacts

How does one move from outcomes to impacts? Development banks and impact-oriented private investors have arguably focused on tracking output-level changes such as the number of jobs created or microfinance loans disbursed: outputs are easier to measure directly and are often assumed to proxy potential positive impact creation. The links among inputs, actions and results have both prior and subsequent dimensions. Advance logical reasoning (also known as the "theory of change") proposes a model of causal links between activities and outcomes (Hehenberger and Harling, 2018) using science- or evidence-based arguments that explain how activities will lead to outputs that lead to outcomes. A theory of change should be built in advance and encourage thinking about all potential (negative and positive) effects. It conceptualises an investment and inspires the choice of relevant indicators.

Jackson (2013b) notes the benefits of using a theory of change to evaluate impact investing:

- It is a cost-effective way to undertake a continuous systematic analysis.
- It helps investors understand their impact and adjust their strategies.
- If public, it can help engage stakeholders and sustain a virtuous circle of feedback, adjustments and impact.
- It is compatible with other dimensions and methodologies of evaluation.
- It helps shape the contours of subsequent impact assessments, which may be partly inspired by the original theory.

Reisman and Olazabal (2016) add that a theory of change increases the visibility of change processes and helps define the investment's underlying assumptions that need to be tested.

Determining causality

The International Association for Impact Assessment (IAIA, 2009) defines the impact of an action as the difference between what would happen with it and without it. To have impact, an action must create additional effects over the counterfactual. This definition is conceptually clear and convincing but difficult to put into practice. The counterfactual is often implicitly described as the absence of investment, and impact is measured as the difference between investing or not investing in a project or company. This fails to recognise potential alternative uses of the invested money or that another investor might invest in the same project or company. The definition of a relevant, credible and valid

[9] Their work builds on insights by Barrow (1997), Coccossis and Parpairis (1992), Hughes (2002), Vanclay (2003) and Hehenberger et al. (2013).

counterfactual is crucial to impact assessment, requiring research-based approaches, and sophisticated approaches and tools.

Establishing causality is central to impact assessment. It requires constant awareness that "correlation does not imply causality." To establish causality, the assessor must control the biases that affect impact analysis, including the following:

- Selection biases (systematic differences between treatment and control groups that explain some of the observed differences in results).
- Contamination (exposure to factors other than the intervention under study).
- Attrition (withdrawal of people included in the groups).

Assessing value

Discussion about the overall impact and its value for society is generally incomplete. Communication of impact often focuses on an investment's positive effects, ignoring potential conflicts with other sectors or the fact that an investment may benefit some groups but hurt others. Focusing on positive effects alone is inappropriate: the objective of impact investment is to have a positive impact on society as a whole. Assessments of impact investments are necessary inputs into a deliberation process that leads society to decide, over time and after potentially conflictual debates, in which direction to move.

Complete impact reporting relies on many impact-related indicators, the choice of which is an important part of the assessment. For example, "balanced scorecards" – a management tool frequently used by impact investors – assess a company's performance along several dimensions, based on a variety of indicators for each. Analysis leads to the attribution of a synthetic rating for each dimension, which echoes the aggregation problem. The standard economic approach to facilitate aggregation relies on the monetisation of performance along each dimension; however, this alone cannot determine the choice of appropriate weights for each monetised indicator.

Beyond the value judgments attached to the synthetic rating, using multiple indicators faces two additional issues. The first is the choice of indicators. Spiess-Knafl and Scheck (2018) report broad agreement on the specific, measurable, attainable, relevant and time-sensitive (SMART) principles. Concerns should be over quality (the nature of the change), quantity (the scope of the change) and time (by when the desired change should have taken place). The second is the potential trade-off between clarity and precision. A small number of aggregated indicators is desirable for clarity, whereas a battery is preferable for precision. Grabenwarter and Liechtenstein (2011) discuss the limitations and shortcomings of one commonly used indicator – CO_2 emissions.

3.3. Who does what?

Approaches to measuring and managing impact within the impact investing market range from reports that describe observed changes without formally relating them to specific actions, to demanding scientific impact evaluations that detail causality and attribution.

Spiess-Knafl and Scheck (2017) report that more than 150 tools and methods are used for assessing social impact, with no consensus to guide choice among them. They argue that four factors determine the choice. Reeder and Colantonio (2013) draw a typology by distinguishing between two different cultures in the use of instruments: system builders are more likely to use statistical tools such as regression analysis, whereas case-by-case supporters tend to focus more on qualitative information.

The two cultures have different purposes and are based on different forms of human relationships. According to them, most impact investors belong to the case-by-case culture. Sophisticated analytical techniques are too technocratic to reflect the complexity and the special nature of the project being assessed, and of the link between investment and results (for example, the possibility that results may depend on an influential fund manager and not be reproducible).

In a landscape study carried out on behalf of the UK government, The Good Economy (2018) adopts a different typology of use and distinguishes between organisational-level and project-level approaches (which may co-exist within a single institution). At the organisational level, the general approach is to go beyond financial metrics and link organisational effectiveness indicators to project-specific results (EIB, 2019). Most investors have a results framework (for development finance institutions) or impact theses (for fund managers) in place which contain or are linked to a set of indicators reflecting desired changes occurring at the output and outcome level. Capturing these changes requires a degree of standardisation – of indicators, scorecards, processes and timelines – that allow investors to "roll up" data from individual investments to the portfolio/corporate level, which in turn allows them to report to the ultimate asset owners and in the public domain. Most multilateral development banks try to strike a balance between standardisation and flexibility (EIB, 2019): some use highly standardised indicators, allow for highly tailored indicators, or try to blend both approaches to fulfil the demand for consistent and comparable reporting. The EIB uses a results measurement framework, based on a logical framework approach, as part of project appraisal and to enhance the Bank's ability to monitor and report on the actual results achieved in its operations outside the European Union (EIB, 2017).

At the project level, anticipated impacts are calculated by investors in order to make investment decisions, such as in investment committees or during due diligence, often using data provided by fund managers, investees, secondary data and assumptions, or sometimes by collecting primary data through site visits. Investors report their results based on data provided by their investees. There is a "value chain" of impact data flowing up from companies through to funds to the ultimate asset owners.

Reeder et al. (2014) identify two distinct approaches. "Evidence followers" rely on existing evaluation evidence of some particular links in their evidence chain; for example, "an investor in girls' education would not need to conduct their own rigorous evaluation to establish the relationship between education and improved economic opportunities for women" (IFC, 2019). "Evidence generators" actively seek data and evidence that outcomes are occurring as expected. Most investors exhibit both behaviours, following evidence on some links of their evidence chain and generating it on others.

There are numerous ways in which investee businesses measure their impact. However, the Rockefeller Philanthropy Advisers (2020) note a "lack of transparency […] on impact performance across the industry" as a key constraint to developing the field of impact measurement and management. There is an urgent need to improve the quality of impact data. However, resource-constrained companies often lack the time, expertise and budget to collect meaningful data on their non-financial performance. This leads to impact data of varying quality, "often either missing altogether, pulled together from third party research (of varying relevance), or based on poor proxies pulled from operational financial data such as sales figures" (IMP, 2019). Self-reported output data suffer from numerous biases, which range from over-inflating positive impacts to "halo effects" that underplay negative effects. The strength of the impact data chain would be improved by more reliance

on independent third-party data collection, notwithstanding biases always present in any measurement, notably in the selection and interpretation of indicators.

While impact measurement is a defining feature of impact investments, "the reality is that defining what to measure and, subsequently, defining how to collect the appropriate data are questions that many impact investors struggle with" (CDC, 2019a). Half of respondents in the GIIN 2020 Annual Impact Investor Survey view the sophistication of impact measurement and management as a "significant challenge" for impacting investing over the next five years (GIIN, 2020a). Such opinions are reflected in investor practices in the same GIIN survey. While 91% of investor respondents measured what outputs occur, and 78% measured outcomes, just 32% measured "contribution to the effect beyond what would have happened anyway," introducing an element of impact in the more academic sense. These practices shape the choice of data collection instruments and methods most commonly used. According to a 2017 survey by the GIIN, specific to the field of impact measurement and management (IMM), 60% of respondents use surveys to collect impact data, 57% use interviews, but only 6% have attempted more experimental methods (GIIN, 2017).

Empirical studies (Loveridge, 2016; Edens and Lall, 2014) confirm that impact measurement practices hardly focus on "additionality" (in the sense of specific contribution by an investor). The IMP (2020b) also reports that "Having a strong focus on attribution favours rigorous […] quantitative methods that […] may limit the scope of an impact evaluation to only what can be measured under special conditions; are not always feasible or cost-effective given the complexity of the real world."

Stakeholders in the impact investing ecosystem are increasingly asking for better evidence of impact (EIB, 2019). Development finance institutions have been under sustained pressure to demonstrate results and value for money, given their use of taxpayer money (Attridge et al., 2019). This requires better techniques as well as data provision to demonstrate to stakeholders that investors are managing (and delivering) for positive impact, to refute accusations of "impact washing" and to generate trust (IFC, 2019). Techniques that can objectivise outcomes – and ideally estimate attribution – help resolve the central "paradox within impact investing": the prioritisation of "social impact" without prioritising "impact evidence" (O'Flynn and Barnett, 2017).

3.4. Methods

Research-based impact assessments should start with a comprehensive literature review to identify areas and questions for further investigation and build on existing knowledge. Researchers should then choose a metric to describe outputs or outcomes and to consider how to access relevant data. Depending on time and resources, they may use secondary (existing) data or collect primary data themselves. Quantitative data are based on numbers whereas qualitative data are based on words or are non-numerical. Data collection methods can be qualitative or quantitative, depending on how they are implemented (for example, closed vs. open questions in surveys). Some data collection methods have a quantitative or qualitative "flavour." Interviews, open question surveys and focus groups are more appropriate for qualitative data collection. Quantitative approaches typically rely on direct observation and measurement, value estimates (derived objectively through economic arguments or subjectively through questionnaires), experiments and surveys.

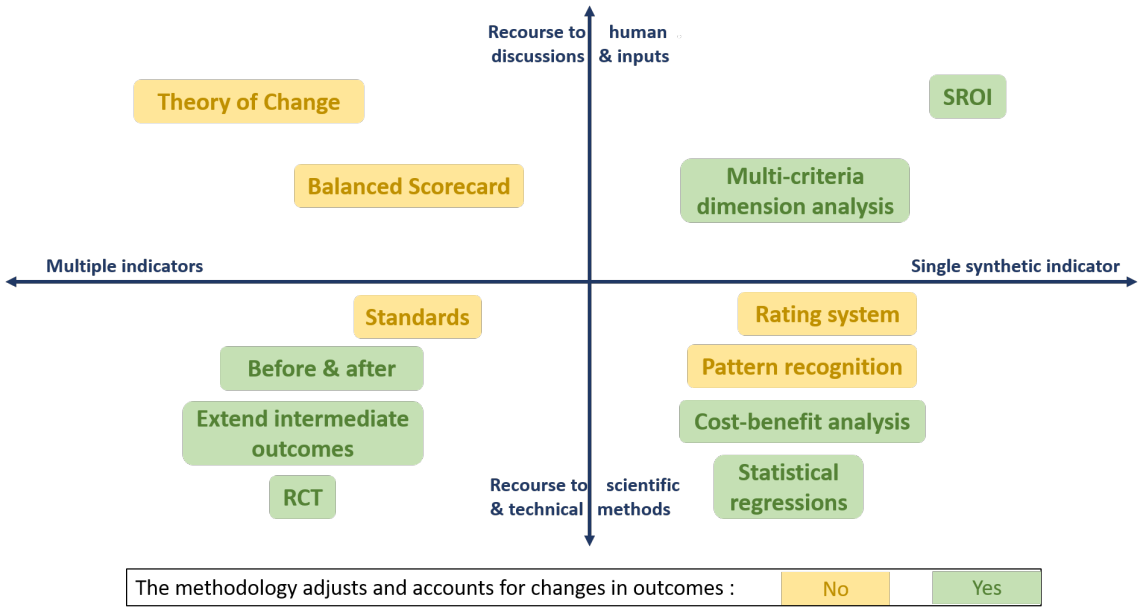
Part of the data collection challenge in assessing societal impact is valuing societal benefits that have no or highly distorted market prices. This leads to different approaches depending on the degree of objectivity and economic foundation sought: at one extreme, rigorous cost-benefit analysis estimates

shadow prices to reflect real resource costs accounting for opportunity costs (under alternative uses); at the other, more subjective and participatory methods may be used to derive the value given to goods, services and societal benefits by interviewees. For example, beneficiaries may monetise the value of the good or service provided. Expert opinions captured with techniques such as the Delphi method – in which a group provides structured estimates of likely changes in outcomes (Reeder and Colantonio, 2013) – are useful to assess an investment's impact.

These quantitative and qualitative methods have different purposes and results. They require different time, financial and human resources. Choosing among them depends on what an impact assessment aims to achieve. The metric used to describe outputs or outcomes can allow before/after comparisons, and qualitative or quantitative comparisons of the measured impacts with a benchmark or with similar initiatives. The Best Available Charity Option (BACO), for example, compares the impact of one initiative with all options offered by other organisations (particularly in the non-profit and public sectors) on the same social issue (Spiess-Knafl and Scheck, 2017). The Acumen Fund, another example, compares the costs per unit to produce and distribute insecticide-treated nets across similar initiatives (Acumen Fund, 2015).

Another dimension concerns the issue of causality (between an activity and the observed outcome and impact) and attribution. Not all impact assessment approaches and methods allow for or involve causality and attribution. Reeder and Colantonio (2013) draw a typology of assessment methods along two clustered dimensions (Figure 3). The vertical axis depicts the degree to which a given technique relies on objective vs. subjective data. The horizontal axis shows whether measurement leads to a battery of indicators ("multiple indicators") or a synthetic figure ("single synthetic indicator"). The assessment methods on the grid are further distinguished as relating outcomes to the intervention being assessed and identifying the changes in outcomes associated with the intervention (the coloured boxes in the figure indicate whether a method adjusts and accounts for changes in outcomes).

Figure 3: Assessing impact: diverse methods



Source: Reproduced from Reeder and Colantonio (2013, p. 23). RCT, random control trials; SROI, social return on investment.

To assess impact with attribution and causality, a range of quantitative impact evaluation methods as part of a well-specified theory of change can be used to set out causal pathways by which a programme or project influences final outcomes. Widely used quantitative methods to assess impact with causality include instrumental variables, regression discontinuity design, difference-in-differences, propensity score matching, randomised assignment and mixed methods (a combination of qualitative and quantitative approaches).

Causality can be determined using natural or randomised experiments, or quasi-experimental techniques. Such techniques enable the control of biases, notably selection biases, through randomisation or the careful consideration, or estimation, of systematic differences in the samples. One lesson from academic evaluation research is that there is no "ideal" method. Randomised control trials (RCTs) have been described as the "gold standard" of experimental approaches, because they provide unbiased estimators. However, they may lack precision (Deaton and Cartwright, 2018; Deaton, 2020) and relevance. Randomised control trials arguably rank at the top of available methods in terms of internal validity (that is, rigour in establishing causal relationships); however, they are less highly rated for external validity (that is, confidence that results can be generalised). Attempts at standardising impact assessment methods are likely to be self-defeating: choosing the appropriate evaluation method to measure the impact of a project is part of the process, and depends on data availability, quality and context[10].

One ambitious project combining these approaches introduced the notion of a social return on investment (SROI), which aims to produce a quantitative (monetary) value. A consensus group[11] of stakeholders is asked: (a) what would have happened without the company's actions (the

[10] Akerlof (2020) describes the perverse effects of an obsession with methodological purity within the economics profession.

[11] Instead of voting to determine the majority opinion, a consensus group is committed to finding solutions that everyone can agree to: no decision is made against the will of any individual or minority.

counterfactual); (b) what did happen (what was added in terms of goods, services or other value for society); and (c) what was produced in monetary terms[12]. Future impact is discounted to define the net present value (SROI Network, 2012); the result is divided by the costs to provide a monetary ratio. This subjective method is arguably a less rigorous cost-benefit/effectiveness analysis. SROI informs causality (at least perceived causality) but may lack precision (Reeder and Colantonio, 2013); the extent of perceived causality may be exaggerated (Clifford et al., 2013). The approach has been criticised for being too complicated, because it may be difficult to account for disagreements among stakeholders (Arvidson et al., 2010) and the process is resource intensive[13].

There is an ongoing debate about the relationship (and compatibility) of social cost-benefit analysis with economic impact assessments. Recent research on environmental issues (for example, Joseph et al. 2020), indicates that benefits estimated through economic impact analysis using dynamic computable general equilibrium (DCGE) models are amenable to subsequent analysis in a social cost-benefit analysis framework, and combining the two approaches provides additional insights. The authors recommend that the environmental assessment process is reformed to include cost-benefit analysis complemented by economic impact analysis to provide decision-makers with more complete information. The relationship between social cost-benefit analysis and DCGE might have a broader relevance. Many approaches using rigorous impact evaluation techniques give insufficient attention to costs and cost effectiveness. This is important for guiding investment choices and identifying sustainable ways to produce impact. There is a good case for using a social cost-benefit analysis framework for evaluating impact investment.

3.5. Common principles

Impact investment funds may have an incentive to use their own approaches and metrics if the market is not standardised. By doing so, they will not contribute to any dynamic of standardisation. The lack of convergence on a method, framework or metric for environmental and social benefits is one reason why a market for impact investing has yet to reach maturity.

The conceptual case for convergence is questionable. Convergence may bring clarity. Many authors criticise the confusion that reigns in the impact assessment market. Convergence may also bring transparency and credibility, because it provides a framework and ensures conformity with a set of rules that, however imperfect, impose a reference of quality. Convergence also provides some comparability (of results), which is important for evaluating and ranking investments.

There is tension between standardisation and flexibility – and between standardisation and effectiveness. Big investors are reassured by standards, whereas social organisations may prefer assistance in developing their own methodologies (Reeder and Colantonio, 2013; Ógáin, Lumley and Pritchard, 2012). Standardisation has drawbacks. Fornaziere (2012) argues that potential negative consequences of standardisation include the distortion of capital flows towards easily measured sectors; the distortion of social companies towards easy-to-serve populations to increase funding; and an increase in pressure on social businesses to comply.

[12] The seven principles of SROI are as follows: "1. Involve stakeholders; 2. Understand what changes (for those stakeholders); 3. Value what matters (the "monetization principle …); 4. Include only what is material; 5. Do not over-claim; 6. Be transparent; 7. Verify the result" (SROI Network, 2012).

[13] Reeder and Colantonio (2013, p. 28) note that "methods such as 'SROI-Lite' have been developed. Sponsored by Santa Clara University, SROI-Lite asks enterprise managers to define the most important output they create, its unit cost, and the ratio of cost to successful output."

Ruff and Olsen (2018) argue that there is no need for standard metrics to ensure comparability as they make things more rigid and complicated, and the absence of flexibility prevents comparability. They propose a "bounded flexibility" concept whereby for each issue area, companies can choose from a limited number of possible measures, enabling partial comparability. Full comparability can then be ensured by measuring success-to-target rates (a measurement tool also open to critique). They call for more professionalisation of impact assessments (of impact investing).

3.6. Measurement as management

Over the past decade, the evaluation conversation has shifted from debates about how to best measure impact to satisfy asset owners – an upwards accountability objective – to how to measure consistently to compare and benchmark across investments. This has included standardising elements of the results measurement infrastructure, such as developing ratings and assessment systems and catalogues of indicators like IRIS+ (a catalogue of metrics curated by the GIIN that investors can use to track results across portfolios and self-report on their achievements)[14]. The development finance institution community has made efforts to harmonise results indicators through the Harmonized Indicators for Private Sector Operations (HISPO) initiative, a collaboration of 25 development banks to set out 27 reporting indicators for international financial institutions' shared clients, drawn from over 400 metrics that were reviewed[15, 16].

Typically, investors have focused impact activities on anticipating potential social and environmental impacts, risks and financial returns when deciding which companies they should provide capital to (Loveridge, 2016). This reflects the fact that the most important decision – and potentially the greatest determinant of the depth and breadth of positive impact creation – can be which companies to invest in. After the investment decision, outputs are tracked as a proxy for outcome achievement. Contribution, if factored in, is mainly considered before an investment is made, including through the requirement of "additionality" as articulated by the multilateral development banks to ensure their activities add to what is available in the market, and do not "crowd out" the private sector (IFC, 2018). In practice, additionality has proved difficult to measure, both conceptually and operationally (Kenny and Moss, 2020).

Recently, the measurement conversation has shifted from a "rear-view mirror" approach to proving impact to improving and supporting future impact. This is based on the realisation that expected impacts are often not attained, particularly as investee business models can change – especially for early-stage enterprises backed by venture funds – and there are uncertainties in stimulating positive social change. Various initiatives have promoted operating principles that partly or fully promote this more integrated approach between measurement and management[17].

[14] See iris.thegiin.org.
[15] See indicators.ifipartnership.org.
[16] The following initiatives have established and promoted common operating principles (alongside the IMP): the UN Principles for Responsible Investing (PRI), which focus on ESG risk management (UN PRI, 2013); the Principles for Positive Impact Finance of the United Nations Environment Programme (UNEP), which pushes investors to adopt an impact-based approach; the Global Impact Investing Rating System (GIIRS), developed by B Lab, a non-profit organisation that administers the B Corps' fee-based certification of impact business models and performance; and the Operating Principles for Impact Management (OPIM) launched by the World Bank Group in April 2019, developed by the IFC jointly with asset and fund managers to promote transparency and discipline in the impact investing market as well as clarity about which investments can be called impact investment (IFC, 2019).
[17] In 2017, for example, the IFC developed its own proprietary framework (IFC, 2019), called the "anticipated impact measurement and monitoring system" (AIMM).

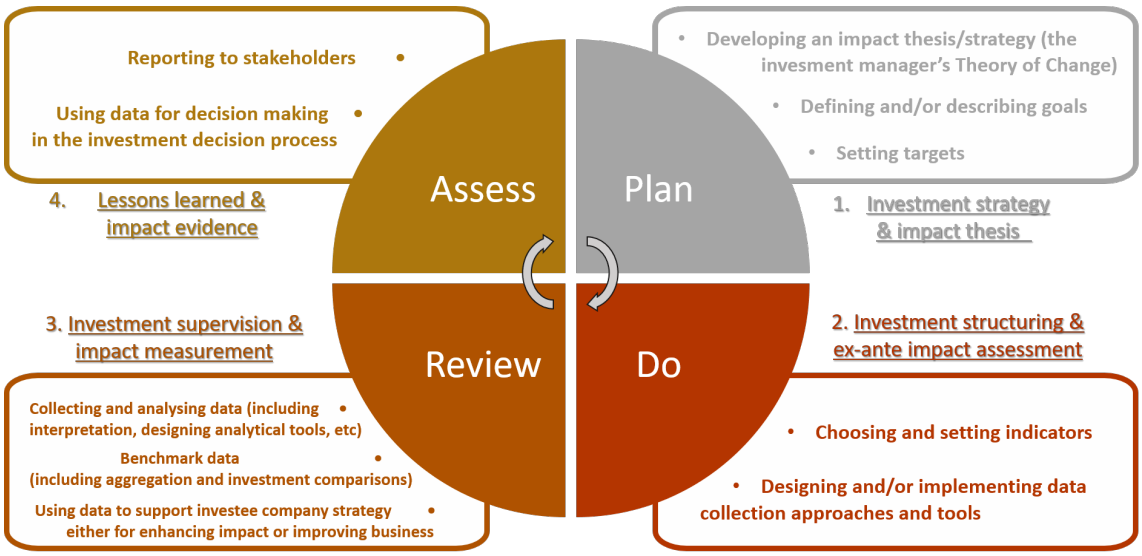
The term "impact measurement" is now generally used together with "impact management." This shift represents an increasing demand to integrate social and environmental performance into operational and strategic decisions. Impact management is framed through the lens of continuous improvement: from New Philanthropy Capital "using your data to improve what you are doing" (Rotheroe, 2017) to Social Value UK "creating a culture [of impact] within an organisation" (Carpenter, 2020). There is no formal definition of impact management, but it is best seen as the practice of integrating impact at each stage of the investment and business decision-making process.

Impact measurement and management itself is defined as the process of selecting and embedding social and environmental performance considerations into the investment cycle, collecting data, and using the information to drive decision-making (GIIN, 2020c). An impact measurement and management system is a set of activities that broadly covers the following:

- Selecting goals and indicators that are mission-aligned.
- Setting targets and strategies most likely to achieve and reflect these goals.
- Measuring and analysing metrics to understand what is happening in reality. This can lead to full "monetisation" – attributing a monetary value to achievements across various dimensions – which the IFC (2019) considers as a "gold standard" among the integrated impact assessment frameworks because it enables credible internal and external comparability.

These are commonly expanded on and turned into a measurement "cycle" with four key steps: plan, do, assess and review (Figure 4).

Figure 4: The Plan-Do-Assess-Review cycle

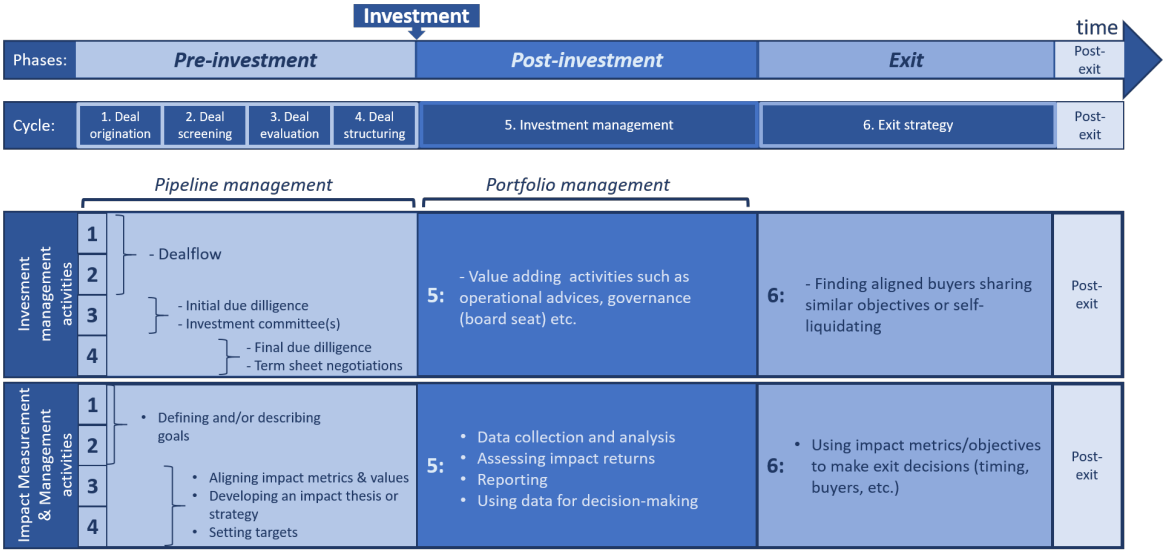


Source: Adapted from GIIN (2020a) and IFC (2019)

To be effective as a driver of continuous improvement, impact considerations need to be embedded within an impact investor's strategy, structuring, supervision and, where relevant, within investment exits (IFC, 2019). Impact management is a discipline much like financial management: policies, processes and procedures are necessary to ensure active management of an investment's performance.

Figure 4 shows how impact management is integrated throughout the investment process. Measurement takes place at two distinct stages: pre-investment (advance measurement), and post-investment (subsequent measurement). These are synonymous with the "plan-do" steps, where impact intentions are articulated, and the "assess-review" steps, where actual impact is tracked.

Figure 5: Impact measurement management across a typical investment cycle from deal screening to exit



Source: prepared by the authors

According to the GIIN's latest impact investor survey (GIIN, 2020a), a majority of the 278 respondents pay significant consideration to impact before investing – during due diligence (81%) and screening (77%). However, this drops to 62% during portfolio management. Only 36% collected impact data during exit, and 21% post-exit. Impact data are given more weight in pre-investment decision-making (articulating intended impacts) than in post-investment performance (measuring actual impact).

3.7. Who pays for evaluation?

Rigorous impact assessments are expensive because they require time, accurate data and expertise (Ruff and Olsen, 2018; Glänzel and Scheuerle, 2016). It is unclear who should pay (investors, investees, funders or a third party) (Harji and Jackson, 2018). There is a potential conflict of interest when a stakeholder of the impact value chain pays to assess the impact of their investment or production. The relevance of this conflict of interest varies depending on the objective of the assessment: it is clear for proving accountability; leads to questionable credibility for reporting impact. It is less problematic for organising the learning function within the organisation. Beyond the issue of conflict of interest, stakeholders may differ in their abilities to bear the costs, perceptions of the usefulness of impact assessments, internal capacity to conduct rigorous assessments and willingness to acquire them. Depending on the process, the learning gain may only profit investors, so investees may be reluctant to undertake an evaluation. Who pays has a bearing on the nature of the impact assessment, and determines its motives and incentives (for example, scientific rigour, communication and accountability) and its credibility.

It is unclear what the potential trade-offs among feasibility, cost, usefulness and robustness are. The ability to bear the costs of a comprehensive scientific evaluation, with or without subsidies from foundations or development agencies, influences the design of the evaluative effort. Spiess-Knafl and

Scheck (2017) argue that the importance of the evaluation (time spent and degree of accuracy sought) is likely to be proportionate to the organisation's size and suggest that companies should acquire internal expertise on impact assessments. In contrast, Clifford et al. (2013) note that such capability can be independent of an organisation's size and complexity. Reeder and Colantonio (2013) consider that "gold standard levels of robustness are impractical for [the] vast majority of enterprises."

Scarce information is available on the money or resources investors allocate to impact measurement. A study by ANDE – a network of organisations supporting small and growing businesses – found that investors' median spending on measurement (calculating staff time and consultant fees) as a percentage of the operating budget is 2.2% and ranges from 0% to 25%. The 50% smaller organisations in their sample spent a median of 8.1% of their budget on measurement (Edens and Lall, 2014). No figures for impact measurement resources as a percentage of assets under management have been disclosed.

Covering the costs of rigorous assessments remains an issue and is one dimension of the "research gap" that underlies the EIB-GDN deep dive initiative.

4. Deep dives as a "bridge"

4.1. The research gap

It is useful to consider the motives and demand for impact assessments. Part of the demand is institutional. Evaluations play a broad governance role. For public investments, they have become part of due process; policy innovations and development projects require evaluation (both before and after). Private-sector evaluations (of impact investments) have been recognised as a condition for the structuring and expansion of the impact investment market. Regardless of their institutional nature, and whether they take place before (through evidence-based assessments relating potential investments to impacts through a theory of change, simulations of impact, and so on), during or after implementation, evaluations use and produce knowledge. This production of knowledge is useful for three reasons that in turn characterise three different functions of evaluations:

1. The accountability function: evaluations are needed to demonstrate that funds are used in conformity with the original intention or mandate, based on the prescribed rules of engagement, and are efficiently and effectively directed towards the intended purposes. Evaluations may also demonstrate to the general public and policymakers that resources are properly used, that impact investment works, and that it should be allowed and encouraged to expand.
2. The learning function: evaluations aim to improve decision-making (and future impact) through learning how project design and implementation relate to impacts (a learning-by-doing function). They improve the likelihood that a given idea and objective of impact may translate into such results. Moreover, evaluations improve understanding and documentation of the links between specific actions and impacts (a knowledge creation function). Thus, they also fulfil a global enlightenment function that extends the learning function to society at large.
3. The promotion function: evaluations provide communication material that can promote a set of actions or the image and reputation of investors, intermediaries or companies.

These various functions of evaluations may require different products and approaches. Accountability requires independent, credible, unbiased evaluations to be undertaken from the

perspective of the individuals or groups to which the various actors need to be held accountable. Learning requires the involvement of the actors under evaluation, who will not learn as much if they are not associated with the evaluation exercise. Independence may matter less for the learning function and may even be counterproductive. Knowledge creation requires open, unbiased minds; a critical approach to assumptions and results; a long-term perspective; and the ability to apply knowledge and methods from the literature. Promotion requires specific qualities including salience and the ability to tell powerful and easily transmissible stories.

Different actors expect different benefits from impact assessments. We infer that the evaluations that impact investors undertake are mainly driven by the accountability requirements and the promotion motive. Learning is a recognised goal of evaluations by institutions, although it is complex to achieve because it implies a transformational use of acquired knowledge to change methods, practices and organisation. Chapter 3 discusses how the EIB-GDN served a learning function for investees. The wider part of the learning function – enlightenment for society as a whole – relates to the gains for the overall community beyond the payer or the institution under evaluation, and often comes as a by-product rather than a driver of impact assessments. However, it may be crucial for the development and effectiveness of impact investing, because its expansion and social role depend on sound, rigorous knowledge about impact and how it is created.

Rigorous impact assessments, notably focusing on causality, can serve all these functions. However, while desirable from a credibility and knowledge perspective, they are expensive and time-consuming, and not perceived in the private sector as essential for accountability or promotion motives. The field of development evaluation has broadened over the last decade to include more impact evaluations, using academic research techniques to learn "what works," including in the field of private-sector development (McKenzie, 2010). However, a typical impact evaluation by the World Bank or Inter-American Development Bank costs several hundred thousand US dollars, with some being as much as $1 million. They can also prompt subsequent assessments ("what worked or did not work" vs. "what is working or is not working") but these come too late to result in real-time learning.

Private investors have little incentive to call for full academic impact evaluations, because they do not correspond to their immediate needs. The challenge therefore is to find ways to encourage more academic research in the evaluation process. This is what we call the "research gap." We believe that there would be substantial positive returns to filling the gap. Independent research-based insights and evaluations using sound methods provide tractability and credibility. For example, evaluative research has improved the design, implementation and effectiveness of conditional cash transfer programmes since the early 1990s. Similarly, the US government's social experiments on housing assistance and negative income tax had major effects on social programmes. Using existing knowledge and methods, academic research introduces the possibility of rigorous, fact-based deliberation at all stages of the decision chain and related feedback loop, in the design and implementation of projects, and what can be learnt from their effects. This dynamic learning process is more important than the promotion of "results," which cannot encompass the entirety of impact and are by necessity incomplete and potentially self-serving. Introducing academic research into evaluations brings rigour to impact assessments, providing a systematic, methodical approach to deal with biases, to interpret results with precision, soundness and credibility, and to build global knowledge.

4.2. EIB-GDN deep dives

The EIB-GDN deep dives aim to address this gap – generating new research-based knowledge of how to conduct impact assessments that can better meet the evidence and learning needs of the impact investing community – and were shaped as a proof of concept. Deep dives can be defined as a dedicated evaluative research exercise to collect and analyse data on the outcomes resulting from an investment, as well as any causal impacts. They focus on individual projects – not the performance of whole portfolios – and can be used to complement existing approaches to organisation-wide evaluation and result reporting.

Deep dives occupy the middle ground between fully fledged, rigorous impact evaluations (that are often not feasible in an investment context) and inferences from self-reported monitoring data. Several institutions, including the EIB, have begun to test different approaches to carrying out impact studies, partnering with academia and specialised organisations[18]. Among multilateral development banks, for example, the Inter-American Development Bank has closely integrated impact studies into its project cycle. The IFC has carried out a number of "rapid evaluations" (Box 3). Other investors, for example the agribusiness lender Root Capital, use impact studies on a representative sample of their portfolio to validate their theory of change and inform course corrections, taking a mixed-methods and "client-centric" approach (McCreless, 2015).

Box 3: A rapid evaluation approach in the IFC

In 2011, only 3% of the female population of Pakistan had access to a bank account, and women-owned businesses faced a credit gap of $179 million. In 2015, the IFC made an equity investment and senior loan to Habib Bank Limited (HBL), the largest commercial bank in Pakistan. This was the first IFC investment made with the explicit intent to support finance for women-owned businesses. HBL has subsequently launched a sub-brand, HBL Nisa, to target women. HBL has also set up a women's business unit, increased key performance indicators for women's deposits, appointed a diversity manager, and trained staff in gender intelligence.

The IFC used quasi-experimental methods to conduct a rapid evaluation of the cause and effect of the gender intelligence programme. The evaluation examined the differences between employees who had undergone training compared to untrained employees. Over 13 000 HBL employees were surveyed, with branch-level data analysed for the years 2014 to 2016. This study showed that branches whose managers have been trained in gender intelligence demonstrated a 10% increase in the volume of deposits from women-owned accounts when compared to branches with untrained branch managers.

Adapted from: Creating Impact, the Promise of Impact Investing (2019)

The closest comparison to the EIB-GDB programme, and also a source of information, is the series of deep dives run between 2016 and 2018 by the Impact Programme, a UK government-funded initiative seeking to build the market for impact investing in sub-Saharan Africa and South Asia.

[18] The EIB has concluded a study on the impact on central and eastern European small and medium-sized enterprises over a comparator group, based on EIB and ORBIS data. Efforts are ongoing to better coordinate and pool such activities among the multilateral development banks.

These studies focused on a selection of high-impact investments made by the CDC Group, the United Kingdom's development finance institution (Box 4). The Impact Programme deep dives focused on generating rapid insights into impact that had value for investors and investee companies alike – rather than comparing outcomes between a "treatment" and a "control" group, they prioritised "lean" methods that provided evidence of outcome and were useful for company decision-making. The deep dive approach has since been integrated into CDC's Rapid Insights Toolbox (CDC, 2019b).

Box 4: Impact Programme deep dives

The CDC Impact Fund invests in funds and other intermediated vehicles that deliver high development impact, particularly to underserved populations. In mid-2016, as part of the UK aid-funded Impact Programme, the fund began piloting a series of insight studies. These were called "deep dives" and aimed to complement existing data reporting and improve understanding of how people engaging with portfolio companies were experiencing change. Each deep dive cycle took two to three months. The aim was to deploy a measurement approach that was light-touch, in-depth and could generate value for investors, investees and end beneficiaries. In partnership with 60 Decibels, the studies used lean data principles, with each deep dive aiming to be:

- Bottom-up – helping companies listen to customers, communities, employees and suppliers to provide actionable insight on their needs and interests.
- Useful – yielding data of sufficient quality to support decision-making.
- Iterative – allowing learning, adaptation and replication.
- Light-touch – using low-cost tools and technologies that require a minimal investment of time and money.
- Dynamic – enabling rapid data collection within a fast-changing environment.

In the first year, deep dives into five portfolio companies took place in Ghana, Nigeria and Kenya, covering the agriculture, energy and information and communications technology sectors. The table below sets out the range of topics covered against the IMP's five dimensions of impact.

Dimension of impact	Example
Who experiences change, and how underserved are they?	A deep dive into an online news platform to understand the socioeconomic and demographic profile of readers
What outcomes does the change relate to, and how important are they to the people (or planet) experiencing it?	A deep dive into a low-cost urban Wi-Fi provider to understand the importance of in-home internet connectivity
How much of the change occurs in the time period in terms of depth, scale and duration?	A deep dive into a solar irrigation company to understand time and cost savings from moving away from manual/diesel pumps

Contribution. How does the change compare and contribute to what is likely to occur anyway?	A deep dive into an animal feed mill to understand how the products create superior value for customers compared to the alternatives used by non-customers
Risk. How likely is the outcome to differ from what is expected?	A deep dive into a company's shea nut suppliers to understand future selling intentions and workplace safety concerns

EIB-GDN deep dives push the boundary of impact studies further to address questions about causality. Deep dives expand the range of methods and approaches used to assess the impact of investments. A more academic approach can yield higher quality impact data by including methods that unpack the causal relationship between observed changes and a specific intervention or investment.

The underlying premise is that the potential to use techniques drawn from more rigorous research methods as part of impact measurement systems has not been fully explored, and "the vast amount of [impact evaluation] literature and experience that exists on government or grant-funded programmes has [so far] not been properly adapted to the realities of private sector-focused impact measurement, neither from a resource nor process perspective" (CDC, 2019a). However, improving the evidence base is not simply a matter of duplicating research methods and designs in the context of impact investing, as "measuring the impact of investments in a private-sector context is different from independent impact evaluations of government- or grant-funded programmes […] due to commercial, legal or logistical reasons" (CDC, 2019a).

EIB-GDN deep dives aim to generate knowledge on how to use impact research techniques that are rigorous but the "right size" for a private-sector context. The core challenge was to establish methods, techniques and processes to assess attribution with a "reasonable standard of robustness" (Reeder et al., 2014). Box 5 presents the main objectives and aspirations of the programme. Chapters 2 and 3 outline the lessons learnt on this journey and document the extent to which the initial goals were validated.

Box 5: EIB-GDN deep dives

The pilot programme initiated in 2016 seeks deeper insights into the profile of end beneficiaries and how they are affected by EIB operations, using primary data, in-depth analysis and academically rigorous methodologies. This box summarises the objectives and goals of the programme, which are detailed in Chapter 2. Some of these studies also address the question of causality – the extent to which any observed change in outcomes can be attributed to an operation under the EIB's Impact Financing Envelope for Africa, the Caribbean and the Pacific, rather than to any other social or economic changes – via impact evaluation methods. The programme is using and enhancing capacity for impact studies by training a set of talented young researchers from African, Caribbean and Pacific countries.

The three overarching objectives of the programme are:

1. To give the EIB deeper insights into the results and impact of its investments, which it can share with its stakeholders, for example the committee members (representing the EU Member States) responsible for the African, Caribbean and Pacific Investment Facility. The funding for the projects under study and the EIB-GDN programme comes from this facility.
2. To bring added value through the research for EIB client companies (direct and indirect clients, in the case of intermediated operations such as investments in funds).
3. To build capacity for local researchers and foster a local research community. The GDN recruited talented early career researchers from Africa, the Caribbean and the Pacific for the programme, who received support from five global expert advisors.

To achieve these objectives, the EIB assigned four aspirations for the deep dives:

***Aspiration 1: Answer questions that matter to EIB clients and other stakeholders*.**

The companies should ultimately drive the scope and focus of the deep dives: asking questions that they need to answer but have not been able to tackle using their own resources. This will be a two-way discussion: the researchers will work with the companies to identify which questions can feasibly be addressed. The findings should be relevant to other EIB stakeholders, particularly the committee responsible for the Investment Facility for Africa, the Caribbean and the Pacific, which funds the projects being studied and the deep dive programme.

Aspiration 2: Generate customer insights with commercial value.

The core of the value proposition is to improve understanding of customers/clients (and potential customers in some cases): who they are, what they want, how they use the investee's products and what motivates them. The specifics will vary by company, but better insights into customers should translate into a better understanding of how to strengthen a company's business model, particularly when the findings do not fit the company's pre-held assumptions. Insights into customers also translate into a better understanding of the company's social impact, which can be employed for marketing purposes with insights into the value added. This should be of interest for companies with an explicit social/environmental mission, and/or those aiming to attract resources from international financial institutions/multilateral development banks/donors.

***Aspiration 3: Use methods that are rigorous, timely and right-sized*.**

Once the questions are defined, the researchers will work to identify appropriate methods, drawing on their academic training and the support of the world-renowned experts mentoring them. The methods should be chosen bearing in mind the timescale of the programme and the data needs of the companies. The depth of the engagement should be tailored to the size and stage of the firm. The rigorous methodologies ensure that the evidence informing the firm's decisions is accurate, relevant and meaningful. The core aim of using cutting-edge techniques is to avoid basing decisions on misleading findings. The researchers will need to adapt these techniques to be appropriate for the questions, hence the studies will differ from academic work.

***Aspiration 4: Contribute to a new way of thinking about impact measurement*.**

The techniques used in the deep dives will generally not be new. However, it remains an innovative way of thinking about impact measurement – putting commercial value added via customer insights at the core, rather than starting with questions set by academics or donors. This can bring significant value to companies and is the only way to proceed – deep dives can succeed only if the company is fully engaged, and this will be achieved only if there is commercial benefit.

4.3. Concluding remarks

While remaining small relative to global volumes, the impact investing segment of financial markets is growing fast, focusing attention on measurement and assessment of results, including social returns. Three main messages emerge from this chapter.

First, impact assessments mean different things to different stakeholders. There is a gap between the private value of impact assessments (driven mainly by the desire to document and report success, and the requirements of accountability) and their social value (based on the extent to which they increase the conceptual and operational understanding of the economic, social and environmental nexuses). Bridging the gap requires analytical efforts, and external incentives and resources.

Second, rigorous academic evaluative research benefits all stages of the investment process (identification, conception, implementation, and assessment of results and impact). Substantive inputs cannot be reduced to measurement frameworks; evaluations should continue to consider several approaches and models and evaluators should cultivate strong connections with academia (Harji and Jackson, 2018; Ruff and Olsen, 2018). The EIB-GDN deep dive programme shows how rigorous research can provide insights into impact creation.

Third, the cost of increasing cooperation with academia for using more research-based methods to improve the quality of impact assessments is unlikely to be covered by private investors. This cost, which represents the public-goods dimension of impact assessments, is related to the "enlightenment" and "learning" functions of evaluations. Supporting the market for impact investing cannot rely only on investors' engagement and generosity. Research-based evaluations applying rigour and control of biases should be promoted, and the focus should be more on knowledge creation than on the promotion and reporting functions of evaluations, while considering private stakeholders' interests and motivations. It is hoped that the lessons learnt from EIB-GDN deep dives will lead to a new generation of research-based evaluations, thus filling the gap.