transformative social innovation theory

A Review of Evaluation Methods Relevant for Social Innovation

with Suggestions for their Use and Development

TRANSIT Working Paper #14, July 2017

By Paul M. Weaver and René Kemp

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 613169

About TRANSIT:

TRANSIT is an international research project that develops a theory of Transformative Social Innovation that is useful to both research and practice. It is co-funded by the European Commission and runs for four years, from 2014 until 2017. The TRANSIT consortium consists of 12 partners across Europe and Latin America. For more information, please visit our website: http://www.transitsocialinnovation.eu/.

About the TRANSIT working papers series:

The TRANSIT working paper series aims to accelerate the public availability of research undertaken by TRANSIT researchers. It presents (intermediate) research results that in whole or part are suitable for submission to a scientific journal or book. It also considers those articles, which are appropriate for submission to (scientific) conferences, workshops or symposia. Our intention is to provide early access to TRANSIT research. All working papers can be found online: http://www.transitsocialinnovation.eu/working-papers.

About this TRANSIT working paper:

This paper draws on and extends an earlier paper: Weaver, P.M. and Kemp. R. (2014) A theoretical dialogue on monitoring and evaluation approaches, social capital, and the sociology of economics, a slightly modified version of which informed the TRANSIT workshop on resourcing and monitoring held February 16-17, 2017 in Maastricht (NL). The working paper in its present form draws additionally on discussions at that workshop and includes materials from interviews by Kay Hoffmeister with social organisations based in Berlin.

Suggested citation: Weaver, P. M. and Kemp, R. (2017) A review of evaluation methods relevant for social innovation : with suggestions for their use and development (TRANSIT working paper # 14, July 2017), TRANSIT: EU SSH.2013.3.2-1 Grant agreement no: 613169.

Date: July 2017

Authors: Paul M. Weaver and René Kemp

Contact: Paul M. Weaver | e-mail: pweaver.groundswell@gmail.com

Online link: http://www.transitsocialinnovation.eu/working-papers

A Review of Evaluation Methods Relevant for Social Innovation with Suggestions for their Use and Development

Authors: Paul M. Weaver and René Kemp

Abstract

Different forms of monitoring and evaluating social innovation initiatives (SII) are needed to respond to the concerns and questions of different stakeholders and to meet needs that arise at different stages in the development of SII. The established social innovation measurement paradigm, which is based on positivism and is grounded in economics-based methods, responds to some but not to all of these needs. It is ill-suited to explore, account for, or to support potentially-transformative social innovation. Furthermore, issues of causality and attribution are especially problematic in the case of societally transformative social innovation, which engages with complex systems and involves lines of influence that cut across levels of scale. There can also be incompatibilities between the asset-based cultures of SII and the use of deficit-based evaluation approaches. These concerns have led to calls to improve existing methods, but also to develop new evaluation frameworks that would address the limitations of conventional approaches (e.g. Antadze and Westley 2012). We propose that social return on investment (SROI) analysis pays more attention to the stories of people involved; e.g. those helped by a SII and involved professionals in respect to their co-production experiences working with SII. This is needed to explain the social innovation to interested outsiders, including what it is intended to achieve, its way of working, and its influences on outcomes. Focussing on only those factors that can be measured may keep from view essential factors and processes in influencing or achieving outcomes. For fostering innovation and improvement of social innovation, Developmental Evaluation (Patton, 2011) and Dynamic Evaluation (Kieboom and Vahl, 2014) are useful new methods. Since there are many potential purposes to be served and these change between stages in the development of SII, using several different evaluation methods can add value. The design and choice of evaluation method should be fit for purpose and context in each application. Action research can help in evaluating SII, as recent experience shows (e.g. Hobson et al., 2016).

Keywords

Monitoring, Social impact measurement, Social return on investment (SROI), Formative evaluation, Summative evaluation, Developmental Evaluation, Innovation

Research Highlights

- Different perspectives on social innovation are reflected in different perspectives on evaluation systems and tools
- Evaluation needs differ between organisations and stakeholders and are likely to evolve requiring different evaluation focuses and methods
- Work on evaluation frameworks is based on generalizable principles and protocols that seek to combine consistent guidelines with flexibility over indicators and metrics.
- Provides an overview of different evaluation approaches.

1. Introduction

Measuring and evaluating the impact of social innovation initiatives is mostly done as an informal, qualitative, ad hoc activity, if done at all. Many social organisations and societal initiatives monitor their activity levels, but they have difficulty measuring and evaluating their social impact and would like to do this in a better way. An exemplary statement is:

> *"The initiatives [of the Migration Hub] are growing; more and more people want their [our] services. We are having a great impact, but we don't have the tool to show the amount of impact we are having. How do we do this?" (Hoffmeister, 2016)*

An example of a monitoring and evaluating activity is a website that registers activities and comments (used by Civocracy, an online platform offering people the possibility to vote online on things that affect their community):

> *"on our website, we have (…) a progress box. In that box you can see the number of people participated, the number of comments that were forwarded by the community, and the number of comments (…). And then a last badge about policy making (…) [but] that is all we are tracking. [we] do not really get important data for the real social impact measurement. The demographic data is also missing. We are very biased on our work. So we say that our social value generating is amazing! But I can't show it. However we need to try as much as possible to unbias ourselves, and to elaborate on our impact. But that is a thing we haven't tackled yet." (Hoffmeister, 2016)*

Graefewirtschaft, a Berlin-based social enterprise that employs migrants and asylum seekers, audits the people they employ in the businesses it runs, which include a restaurant called "Die Weltküche", kitchens catering for schools, kindergartens and day-care centres, and household services and care for the elderly:

> *"Generally speaking we use the social auditing procedure. A main indicator for us is the amount of people we get into jobs and apprenticeships".*

In the Netherlands, a platform for social enterprise and citizen initiative called "Kracht in NL" created a metric system for social impact called MAEX.[1] The MAEX consists of 8 elements which are scored on a scale from 0 to 10: smart use of existing resources; self-reliability; social cohesion; cooperation for a better environment; (social) safety; sustenance support; leisure; education/development. The scoring is done by the initiatives themselves and reflects their own subjective assessment of impact. The MAEX improves the visibility of initiatives that occur around the Netherlands (1091 in total), signalling to others what they are doing, which social impacts are being created and what resources they need (specific expertise, money, materials etc.). For those interested in it (volunteers, local government, and businesses), it offers a portal for participation. For government and research it offers data on SIIs on the nature of activities, types of social impact that are being created, the average amount of volunteering time per week and amount of self-earnings.

In Figure 1 the MAEX scores are given for 4 initiatives in the Netherlands: Zelfregie-centrum Venray, through which experience experts help people with psychological problems get a grip on their lives (especially those who are not eligible to official care because of budget cuts and

[1] The name MAEX refers to the AEX the Amsterdam Exchange Index, the official index of the stock exchange in the Netherlands. The term M stands for Maatschappelijk (*Societal*).

illness requirements); Stichting Goed ontmoet, which is a food bank; Hilverzon Duurzame energie cooperatie, a renewable energy cooperation; and, Senioren kennis netwerk Maastricht, through which professionals-in-residence offer knowledge services to societal initiatives and local organisations at the interface of government and society.

Figure 1: MAEX scores for 4 initiatives in the Netherlands

Legend (from top clockwise): Levensondershoud (sustenance support); Sociale veiligheid (social) safety; Sociale Cohesie (social cohesion); Samenwerking, participatie (cooperation for a better social environment, participation); Slim gebruik/ duurzaamheid (smart use of existing resources and sustainability); Vrijetijdsbesteding (leisure); Ontwikkeling (education/development). The blue line refers to the local social context and the orange line to the target group of the initiative

Source: www.MAEX.nl

The field of evaluating social innovations is developing rapidly in response to a widening range of questions being asked about social innovation, its outcomes, its impacts and the contexts within which it operates and with which it interacts. The paper surveys selected literature on evaluation theory, methods and practice in this field. The materials reviewed are selected for their relevance to the concern of TRANSIT for transformative social innovation.

In section 2 we explore what is driving developments in this field by looking especially at the changing context for social innovation and at the expanding range of evaluation interests, foci and questions of different stakeholders. Section 3 provides a review of currently dominant evaluation methods and tools, their strengths and their weaknesses. This draws on existing surveys of the field. It is found that the dominant tools, indicators and metrics within the paradigm are based mostly on standard economic and accounting approaches, which is related to the need for summative evaluation by funding agencies. Section 4 describes the most prominent of the approaches: Social Return on Investment (SROI). Section 5 presents and discusses an exemplary analysis of SROI, as an example of summative evaluation for the case of Foster Parents in the Netherlands. Section 6 contains the recommendations from a group of experts on social entrepreneurship (GECES) tasked by the European Commission provides to offer recommendations on impact measurement. Section 7 outlines a contingency approach to impact measurement (developed by Alex Nicholls in the CRESSI project), which starts from the question how can organizations chose an approach that is appropriate to their concerns and context? Section 8 presents an innovation-oriented form of evaluation called developmental evaluation (DE), developed by Michael Quinn Patton. Rooted in case studies of (social) innovation processes, DE seeks answers to questions that are relevant to innovation, by helping social innovation actors to take a broader systems perspective and help them navigate (inherently uncertain and judgment-based) processes of change, by making them reflect on their assets, their theory of change and the opportunities and dangers afforded by a changing context. It contends that the measuring needs at each stage in social innovation processes are different and that the measuring approaches and tools used, such as indicators and metrics, will also need to change from one stage to the next. Section 9 discusses evaluation anxiety and discusses the role of action research. Section 10 discusses the importance of the monitoring of context (for helping SI initiatives find suitable partners and strategies). Finally, in Section 11, we draw conclusions from the overall discussion on monitoring, measuring and evaluation for the development of a theory of transformative social innovation (TSI) and the development of supporting tools and methods in the TRANSIT project. A table with monitoring elements of TRANSIT cases is provided in Appendix A.

2. Diverse demands for monitoring and evaluation

The 'evaluation' issue is not so simple or straightforward as it might at first appear. There are many different kinds of evaluation question that can be asked about social innovation. Different stakeholders, with different concerns and interests, have different evaluation needs and therefore pose different evaluation questions. Also the evaluation needs of specific stakeholders, and especially those of social innovators and social organisations, will change depending on the stage of the social innovation process and the context for the innovation. What is being evaluated and in respect to which impacts and which targets differs from one evaluation question to another. Even the levels of scale at which impacts manifest are potentially different. At one extreme, some evaluation questions concern outcomes and impacts experienced at the scale of individuals. At the other extreme are much broader changes that manifest at higher

levels of scale. The evaluation question that underlies and motivates the TRANSIT project is especially relevant here since it concerns impacts of social innovation processes that could manifest at the societal level through broad, lasting (and therefore transformative) changes in social relations, institutions, constructs and behaviours. In this perspective the 'targets' of interest are the social relations, institutions, constructs and behaviours manifested by and in society and aspects of these that are relevant to important qualities of society, such as its cohesiveness, greenness and resilience. As well as positive impacts there is also scope for social innovation to have negative impacts, which are also important to be included in evaluations. The 'content' of evaluation is therefore also a relevant aspect of evaluation design.

Overall, different evaluation purposes and questions call for different types of evaluation and for different evaluation approaches, methods and tools. Similar arguments apply to the range of different social innovations, stages in the social innovation process and implementation contexts. They apply also to different definitions and perspectives on social innovation and, especially, to whether social innovation is defined mostly or exclusively by its content (as products, services, activities, actions, etc.) or as a process. This wide diversity means that there is a need for different types of evaluation and for evaluations to be designed and implemented that are 'fit for purpose' and 'fit for context'. At the same time, there are clearly commonalities in evaluation processes that imply the possibilities of some generalizable principles and guidelines. Also some needs for information and other resources recur across different kinds of evaluation questions. There is a strong interest in capitalising on these commonalities to provide for comparability across evaluation and to help in reducing monitoring, measuring and evaluating burdens and increasing efficiency.

Furthermore, since social innovations are (by definition) innovative, and therefore are likely to be evolving through the stages of their invention, experimentation, proving and upscaling, they are a moving target as objects of evaluation. From a management perspective this is important because monitoring, evaluation and comparison of variants of the 'basic model' of the social innovation is an important part of learning about a social innovation and perfecting its design. The kinds of evaluation questions important to social innovators to support learning and continuous improvement are therefore likely to involve comparison of the impacts and outcomes of different design variants and their comparative advantages, disadvantages and effectiveness.

Against this backdrop it is useful to list some of the main distinctions that are relevant in mapping the field of social innovation evaluation. An important distinction is between internal stakeholders in social innovation and their concerns versus external stakeholders and their concerns. The distinction is important in part because the values sought through social innovation may be different for these different groups. Internal stakeholders, such as mission-oriented social organisation actors need to know what impacts their activities are having and how effective these are in achieving the outcomes and impacts they seek. They are especially interested in the technical effectiveness of their actions and activities in relation to achieving their social goals and objectives. They are likely to want to use evaluation as a management tool to help inform their decision making at different stages in the social innovation process. Some external stakeholders may be more interested in the social impacts actually produced and less concerned for how these are produced; for example interest organisations may want to know how social innovations affect particular individuals and groups of special concern to them.

When social organisations receive funding from public, private, philanthropic or blended sources there is a need both for the social organisation to demonstrate that the funds it receives

are making a difference and for the funders to demonstrate that grants, loans and investments in social innovation organisations and activities are productive and efficient. If social organisations take over or complement roles and functions taken by the state (for example in areas of welfare delivery) and receive income in return, this also generates a need to measure financial performance and added value for reasons of transparency and accountability. When the financial instruments used to finance activities take the form of performance assurance contracts, as applies to Social Impact Bonds, the very viability of the funding instruments depends on developing and agreeing ways to measure outcomes and impacts. In general a capacity to demonstrative effective and productive use of funds is especially important in the context of a more challenging financial context characterised by greater competition for funds.

What is at 'stake' and is, therefore, of monitoring and measuring interest for these different actors and stakeholders can include, inter alia: the range of outcomes and impacts produced, positive and negative; the activities through which outcomes and impacts (positive and negative) are produced; the nature of the mechanisms through which impacts and outcomes are produced; the value-added to different affected parties by the activities of the social organisation, including to beneficiaries in groups of special interest (such as the vulnerable, excluded, unemployed, or elderly); the relative and absolute 'technical' effectiveness of the social innovation; the financial effectiveness of investments in the social innovation; specific aspects of the social innovation (such as its acceptability to regulatory authorities, its need for finance, its possibilities to generate financial returns as well as social returns, the possibilities for it to take over roles from the state that the state might be happy to offload, the safeguarding and governance issues surrounding the innovation, etc.).

Evaluation questions can therefore be addressed toward the social innovations of interest, but they can equally be addressed toward the organisations that promote them, the resources (such as finance) that are applied to them, the strategies, activities and actions that they entail, or the contextual conditions they encounter. They can be directed toward specific outcomes and impacts that might be sought or be more open and designed to explore different outcomes and impacts, both positive and negative. They can be directed on different targets: individuals, groups, communities, sectors, society as a whole, etc. One question that has become very important in the current context, especially to policy makers, is the evaluation question that the TRANSIT project is asked to explore concerning the societally-transforming potential of social innovations.

At different stages in social innovation processes, evaluation may play very different roles. Formative evaluation is useful for gathering information about the effects of actions and activities, positive and negative, when little is known about these. Summative evaluation is useful when more is known about impacts and interest lies in fine tuning innovations or selecting among different variants prior to scaling up. As we discuss later in this paper in more detail, a third form of evaluation, developmental evaluation, is, in principle, very important for the purpose of increasing positive impacts (Box 1).

Against this backdrop there is a growing demand for evaluation systems and metrics to measure social impact and outcomes, but also a diversifying demand, since the demands are arising from different sources and these reflect different perspectives, purposes and needs. Different stakeholders have different evaluation foci and needs, which manifest as differences in the kinds of evaluation questions they seek to answer including differences at the very fundamental level of what is being evaluated; i.e., the impact of what on what? Furthermore, evaluation needs and questions are likely to change through different stages in the development of a social innovation

as the social innovation develops, evolves, diffuses and goes to scale. The dominant evaluation paradigm is based on positivism and involves a strongly linear model of evaluation that conceptualises clear cause-effect links and seeks to explore these. This paradigm focuses on social innovation as defined by content (i.e. social innovation perceived as innovative actions, activities, products or services) rather than social innovation defined as a process interacting with complex systems. In section 8, we offer a deeper discussion of developmental evaluation.

Box 1: Developmental evaluation

"Developmental evaluation refers to long-term, partnering relationships between evaluators and those engaged in innovative initiatives and development. Developmental evaluation processes include asking evaluative questions and gathering information to provide feedback and support developmental decision-making and course corrections along the emergent path. The evaluator is part of a team whose members collaborate to conceptualize, design and test new approaches in a long-term, on-going process of continuous improvement, adaptation, and intentional change. The evaluator's primary function in the team is to elucidate team discussions with evaluative questions, data and logic, and to facilitate data-based assessments and decision-making in the unfolding and developmental processes of innovation."(Patton, 2008).

In Table 1, the results of a developmental evaluation exercise are given for the case of homeless people in Canada, showing the elements of DE, the translation of it for the case of homeless day laborers and the ways in which they were helped with securing housing and achieving better income.

Table 1. Results from a Developmental Evaluation exercise: Experimenting with innovative ways to help homeless day laborers secure housing and better income in Canada

What was developed through developmental evaluation?	What this means	Examples
1. Understanding the challenges of innovation and systems change	The effort to tackle a complex problem may generate new and/or deeper insights about the nature of the challenge being addressed and/or the context in which it is being addressed.	The innovators realized the importance of social supports in the "homelessness puzzle", once some of the clients who secured housing were drawn back to the streets to regain the friendship and company of their previous network.
2. Theory-of-change elaboration	The innovators may have new ideas about how they might address the challenge and/or the kinds of results they might expect from their efforts.	The innovators expanded from their strategy focused primarily on housing and employment income to one that included education, social networks, and mental and emotional help.
3. Change mechanisms	The establishment of concrete mechanism (e.g., practices, regulations, relationships, policies) that have an influence on the challenge being addressed may represent the most tangible development of the innovation.	The innovators established (a) a protocol with local credit unions to provide clients with access to bank accounts, even before they had permanent addresses; and (b) an arrangement where laborer could bypass predatory, temporary job agencies (which took 50% of their wages) and use a nonprofit intermediary that allowed them to retain all their employment earnings.
4. Capacity development of social indicators	Developments that relate to the capacity and morale of the innovators and affect how they think and pursue their innovation (e.g., skills, resources, membership).	The trust between previously disconnected service agency leaders increased after these early successes and allowed them to open up their work to discussing the deeper reasons why they found it difficult to integrate their services more closely (e.g., competition for resources).
5. Deepening understanding of context	Developments that are not under the complete control of innovators but in which what happens (emerges) contextually shapes the goals, design, delivery, and results of the innovation (e.g., economy, demographics, key events). All developments are important to track and assess in DE Whereas the previous four types in this exhibit refer to the development of the innovations, this fifth one (the context) is equally important because innovation does not emerge in a vacuum, but instead is highly influenced by the context in which it is unfolding	A slowdown in the construction industry (the major employer form homeless day laborers) required the innovators to develop relationships with different types of employers and adjust their expansion plans.

Source: Patton (2016)

3. Evaluation paradigms and tools

There is a limited body of past research on the monitoring of social innovation and the evaluation of social impact. Albeit this is now a fast-developing area, it is widely acknowledged in the evaluation literature that the area of social impact measurement has been under-conceptualised, under-theorised and under-researched. A specific observation is that research on metrics for social innovation is scarce and that there has been very little work to develop tools and methods to evaluate social innovation and social impacts specifically (e.g. Ebrahim and Rangan 2010).

On this basis, most currently-used tools in social impact measurement practice were not developed expressly for social impact assessment. Instead they are based on standard economic methods and tools of financial accounting and reporting. These tools therefore have closer affinity with perspectives and needs arising in social finance, for example needs to measure investment efficiency and productivity or to optimise investment portfolios, rather than with needs arising from mission-oriented social organisations concerned to improve or to track the effectiveness of their activities and actions.[2]

Against this backdrop, there have been several review studies already of approaches to measuring social impact (e.g. Mulgan, 2010; Nichols, 2015; Antadze and Westley, 2012). Included in most such reviews are:

- Cost-benefit and cost-effectiveness analysis: these are widely used tools based on expressing the costs and benefits of interventions in money terms, often applying a discount rate and often using a 'costs-saved' approach to place money values on social benefits; e.g. the saved or avoided healthcare costs of an intervention that promotes healthier lifestyles. CBA and CEA are often used for large programs in areas of public provision of social and welfare services: health care, care of the elderly/young/vulnerable, rehabilitation of ex-offenders, etc.

- Stated preference: the approach is based on expressing benefits in money terms using a willingness-to-pay approach through which beneficiaries or potential beneficiaries of an intervention are asked to estimate how much they would be prepared to pay for the benefits.

- Revealed preferences: compares options and infers the value of benefits from the choices people actually make; e.g. using relative costs of similar homes in different locations to estimate difference in local amenity value.

- Social impact assessment and social return on investment: estimates direct costs of an action/intervention, the probability of it working and the likely change in future outcomes (sometimes with a discount rate). This is a broad family of tools used to support philanthropy and impact investment decisions, for example to compare alternative grant or investment options. Many (sometimes fund-specific) variants exist; e.g. Best Available Charitable Option (Acumen Fund).

[2] Antadze and Westley (2012) note that for social organisations productivity of finance is a means to an end, not an end of itself.

- Public value assessment: seeks to assess how much the public values a service, such as public-service broadcasting

- Value-added assessment: seeks to assess the quality added through a service, such as education, by comparing entry- with exit- level characteristics, such as educational-levels (rather than by using only qualifications attained, which cannot account for differences in school intake).

- Quality-adjusted or disability-adjusted life years: seeks to estimate the cost-effectiveness of health-care treatments by combining objective and subjective experiences (extension of life and experienced life quality).

- Life satisfaction: assesses social projects and programs in terms of the extra income beneficiaries would need in order to achieve an equal gain in their life satisfaction. The Life Satisfaction Approach is potentially interesting for evaluating social innovation impacts in contexts where there is a demand for money valuations. The Life Satisfaction Approach seeks to value non-market impacts. It uses econometric methods to estimate the life satisfaction provided by non-market goods and converts this into a monetary figure by also estimating the effect of income on life satisfaction. The approach therefore seeks to assess impact in terms of "how people think and feel about their lives as a whole, instead of assessing impact based on what people say they want and what they choose" (Fujiwara and Campbell, 2011).

- Government accounting: some national governments (e.g. France, Italy) use standard sets of indicators to monitor government spending and its societal effects

This listing confirms that many of the most widely-used methods in current use stem from conventional accounting practices, so they are not designed from first principles for capturing social impact. CBA, CEA and related assessment methods are further limited because the indicators can measure only single aspects of performance and each impact requires a 'tailor-made' indicator. These assessment methods are therefore unable (or unlikely) to reflect the full value of the social impacts. The most widely-used model of social impact measurement currently – the Social Return on Investment (SROI) model – is also based on conventional economic and accounting principles and on cost-benefit thinking. The approach seeks to establish a ratio of returns (economic, environmental and social) to the activities of an organisation. The model reviews the inputs, outputs, outcomes and impacts made and experienced by stakeholders in relation to the activities of an organisation, putting money values on all costs and benefits created by the organisation (economic, social, environmental) and expressing these in terms of a productivity ratio (Arvidson et al 2010).

Not being developed specifically for this field and for the needs within it, currently-available methods and tools are therefore not always fully appropriate for the specifics of particular evaluation tasks or for fulfilling specific evaluation functions. Needs within the field are diverse already but also are still diversifying, so the challenge of developing new and more appropriate social impact evaluation approaches is becoming more urgent.

Antadze and Westley (2012) state that the situation calls for the creation of new evaluation models that incorporate not only financial but also environmental and social considerations and

that provide mechanisms for determining the scale, impact and durability of social innovations. But they note, also, that there are special challenges in evaluating social innovations, which is intrinsically more difficult than evaluating technical innovation.[3] Of considerable significance in this regard is their observation that the dynamics of social innovations and the challenges they address are nonlinear, uncertain and unpredictable, which implies that "a positivist approach to measuring social impact is insufficient" (Antadze and Westley 2012, p.134). This is significant because the established paradigm of monitoring and measuring social innovation is based on positivism. However, this observation holds added significance for the TRANSIT project since a positivist paradigm is intrinsically unsuited for exploring the wider impact of social innovations in respect to broad-scale processes of societal transformation. We return to this point later in the present paper.

4. Social return on investment

There is currently a strong government interest in incentivising or requiring social organisations to use the Social Return on Investment (SROI) model. Wood and Leighton (2010), in a report on behalf of DEMOS, suggest this is because public services are facing a period of unprecedented cuts as efforts are made to recover the economy and public finances in the wake of the economic and financial crisis and policymakers are seeking both to harness social organisations in the delivery of services and to ensure value for money. In the UK, for example, social organisations have increasingly been harnessed in delivering public services leading, since 1997, to "an unprecedented shift toward plurality in public services."[4] Thus, Wood and Leighton observe: "a recent surge of interest in social reporting has seen SROI becoming the tool promoted by government, thanks to its unique feature of attributing monetary values to 'soft' outcomes" (Wood and Leighton, 2010, p.14).

The surge of interest by policymakers in social reporting using SROI has led to concern over both the appropriateness of making SROI a dominant approach in measuring and reporting social impact and the practical feasibility of this in the short term. Concerns are expressed also for social organisations to be treated fairly, so that lack of capacity to use SROI should not be an impediment to their receiving income. In a survey of social organisations, Wood and Leighton (on behalf of DEMOS) found that very few organisations are implementing SROI as yet and that the majority are not 'SROI-ready'.[5] The DEMOS study concludes that "although SROI may be neither practicable not desirable for all organisations, the basic concepts of outcomes evaluation that it encourages are important for all organisations to achieve." Furthermore, the DEMOS study accepts that there will be continuing pressure on social organisations to monitor and report SROI.

Against this backdrop, the DEMOS report makes a set of recommendations, including that there is a need first to set a more achievable social value measurement target for the whole sector. This could be "a universal benchmark" established as "a stretch target" to help improve evaluation in the sector. It therefore needs to embody good practice in outcomes measurement and evaluation and to be accompanied by investment in training and practical guidance. Also,

[3] They state, for example, that unlike technical innovation, the impact and outcomes of social innovations cannot, at least initially, be judged by growth in market share, profitability, or even consumer satisfaction.

[4] Wood and Leighton report that by 2010 the UK government accounted for one-third of the total income of social organisations and that, by then, around 27,000 charities (25% of the total number of registered charities) relied on government for over three-quarters of their funding.

[5] Wood and Leighton define SROI-readiness as involving the capacity "to identify and measure organisational outcomes adequately in a quantitative way."

social organisations should be encouraged and incentivised to work toward it by commissioners and funders. The DEMOS study further suggest that any such benchmark should be underpinned by three principles: proportionality, so that the burden of evaluation is in line with the scale and nature of the organisation undertaking it; comparability, so that even with a range of flexible frameworks, organisations can still produce outputs based on comparable principles and terms of reference; and, standardisation, so that there are tools and data available to remove the need to evaluate outcomes from scratch and reduce the burden on organisations.

The Calouste-Gulbenkian Foundation, which funded the DEMOS research, advocates that foundations work with grantees to face the challenges posed by measuring social returns and, in support of this, makes the following suggestions to funders:

- Include a budget for measurement of outcomes (and then social returns) in your grants and set measures for yourself as well so you learn in parallel with your grantees.
- If you hit difficulties in your discussions with grantees come back to the common goal: social impact. Discuss this with your grantees: many of the problems of measurement turn out to be problems of lack of agreement on goals.
- Even if you do not seek to quantify the financial impact of your outcomes do think about how this financial value will be delivered as it will refine your view of what is truly valuable.
- Monitor implementation around outcomes, not outputs, as this is where the impact is often assumed and not managed.
- Be persistent: this is going to be a long journey. Seek continuous improvement; one of the benefits of measures is that they enable us to continually ratchet up expectations (of ourselves and others).
- Share good practices between grantees and with other funders.

5. An illustrative example of SROI: The case of Foster Care in the Netherlands

In this section, we present an example of a social return on investment calculation, to give an idea of how it works.[6] The analysis is undertaken by Sinzer in the Netherlands, at the request of the sector organization of foster care homes for children who are unable to live with their parents. It concerns children in the up-to-18 age category for whom youth care professionals consider the home situation inappropriate for their socio-emotional, physical and cognitive development because of violence, sexual abuse and negligence. Foster care homes (gezinshuisen) are an alternative to state care institutions (residentiele instellingen). In the SROI the costs and benefits are determined for funders, municipalities, the children themselves and agencies responsible for finding places of care. The costs and benefits for the parents are not determined but their acceptance of the referral and their assessment of moments of contact is assessed (via scores on a scale from 0 to 5).

Effects for all people and organisations involved are determined via interviews with foster care providers, experts and, to a lesser extent, the children. For children the positive effects include: the creation of a prosocial network (based on relationships of care and trust), greater chances of

[6] This section is based on http://www.gezinspiratieplein.nl/lezen-weten/lijst-met-alle-publicaties/74-maatschappelijke-business-case-gezinshuizen/file

obtaining a school diploma, finding work, less debt and a greater sense of self-esteem and acceptance. For funders the benefits of foster care over residential home care include lower care costs (which amount to 11688 euro per child). For municipalities/government the benefits include: prevention of crime, less need for special education, and lower demands on social welfare arrangements.

Based on interviews with foster parents, for each effect category, the duration of the effect (1, or 3 years) is estimated together with the attributed chance of the effect occurring. Each effect is monetarized, based on cost and benefit information that is available. The cost information used gives a hard element to the monetarization but no attempt is made to personify the costs and benefits. The basis for calculating the benefits from reduced crime are the costs of custody. The avoided psychological damage to victims of crime is not included. The benefits of having a support network of 3 persons are calculated as 4500 based on the maximum payment of 1500 euro that volunteers in the Netherlands can receive for doing volunteering work. This translates into a benefit for the child of 714 euro a year based on a positive impact chance of 32% (the impacts chance stems from subjective guess by experts). The gains of a diploma are based on econometric studies that say that every extra year of schooling adds 5-15% income. The 5% number is used as this is considered most appropriate (the percentage increases with level of education) and for income, the average minimum wage for 18 to 20 year olds is used. Here the impact factor is estimated at 22%. Those assumptions appear more reasonable than those for estimating the benefits of having a social support network, which are not based on benefits for the recipient of care but based on the sum of money that government is prepared to pay for the work done by volunteers in general.

The analysis is an example of a summative evaluation. The value of the benefits for a foster care child is estimated at 1,727 euro, for municipalities at 4,106 euro and for agencies responsible for finding places of care it is 189 euro. The biggest benefit category is the saving in the costs of direct care by bringing children in foster care homes. The SROI (the quotient of overall benefits and costs) is estimated at 1.30 (91% of which stems from the lower care cost element), indicating a positive societal business case. To funders, the study showed a clear benefit which they already knew. For supporting innovation, the SROI offers little. The analysis does not provide any insights into conditions for success and processes behind achieving positive results for the children, the results of which could help foster care people to provide care in a better way and other organisations to take helpful measures. Human stories about life changing experiences are absent from the evaluation, a missed opportunity.

The limitations of a purely summative evaluation are recognized in a blog by Marlon van Dijk on theories of change, where she offers useful suggestions for increasing impact.[7] In her blog, she argues for a deeper investigation of the conditions for success and for using such knowledge for a more tailored approach of social care. The example she uses is that of an alcohol rehabilitation programme, where it was discovered that having a social support network of family and friends was a critical factor for success. It also was discovered that failures to get off alcohol resulted in depressions, feelings of failure and reduced motivations, as negative side-effects. This learning led to the introduction of buddies for people without social support and the decision to limit the programme to those with social support. Both choices greatly improved the effectiveness of the rehabilitation programme and helped to reduce the negative side-effects.

[7] http://blog.sinzer.org/author/marlon-van-dijk

6. The recommendations of the GECES sub-group

The DEMOS study, while supporting the idea of a universal reporting of social impact, nevertheless has concerns over the pre-occupation on SROI that represents the evaluation priority of funders, but does not necessarily respond to the evaluation needs and concerns of other stakeholders. As evaluation needs and questions of parties with other than only financial perspectives also need to be considered, more comprehensive approaches to monitoring and evaluation are needed. The DEMOS study therefore recommends the development of a more comprehensive and universally applicable evaluation framework. Progress in this direction is represented by the recommendations of another group – the GECES sub-group on Impact Measurement – set out in a report on 'Proposed Approaches to Social Impact Measurement (GECES, June 2014).

GECES (Group of Experts of the Commission on Social Entrepreneurship) was established in the policy context of European Commission legislation and practice. The Single Market Act II states that "the Commission will develop a methodology to measure the socio-economic benefits created by social enterprises" and that "the development of rigorous and systematic measurements of the impacts of social enterprises on the community is essential for demonstrating that the money invested in social enterprises yields high savings and income". The GECES sub-group on Social Impact Measurement was therefore established "to agree upon a European methodology which could be applied across the European social economy" (GECES 2014, p.i).

The immediate need for a methodology relates to two funding instruments for social enterprises. The European Social Entrepreneurship Funds (EuSEFs) and the Programme for Employment and Social Innovation (EaSI).[8] Both programmes come under the Social Business Initiative (SBI) and are focused on supporting the development of Social Enterprise within EU Member States.[9] The measurement needs of these two instruments are different, however. For EuSEFs there is a need to create a standard for judging whether a social enterprise qualifies to receive financial support. For the EaSI programme, under which grants will be made available to social enterprises that are able to demonstrate "a measurable social impact", the need is for those managing the funds to report upon the extent to which the social impact targets of the whole fund are delivered (GECES 2014).

The GECES report makes clear, nevertheless, that the development of a standard for impact measurement goes beyond these immediate needs, pointing out that "nowhere in the world is there an agreed standard for social impact measurement" and that "to develop one would bring consistency to reporting" (GECES 2014, p.i). While the GECES standard is thus developed in relation to social enterprise and social entrepreneurship and is sensitive to the funding instruments it is intended to serve, the underlying intention to "measure social impact" is relevant more widely across the field of social innovation. The GECES report, like the DEMOS report, thus identifies the need for an agreed and universal measurement standard and seeks to

[8] Under EaSI, €86 million in grants, investments and guarantees will be made available to social enterprises in 2014-2020.

[9] In this context, a social enterprise is defined as an operator in the social economy whose main objective is to have a social impact rather than make a profit for their owners or shareholders. It is characterised by the dominance of a social objective of the common good, the reinvestment of most profit with a view to achieving the social objective, and by governance structures (e.g. systems of organisation, ownership, stakeholder participation, etc.) that reflect this mission

build foundations for this in the form of a set of principles and guidance for an approach that evaluates impact based on outcomes.

The main argument of the GECES report is that it is neither possible nor desirable to devise a rigid set of indicators in a top-down and 'one-size-fits-all' fashion to measure social impact in all cases.[10] The report warns that to impose a pre-determined, closed set of quantitative indicators "from the top" also risks being highly counterproductive, especially if funding decisions are based on performance against these indicators, since this introduces dangers of perverse incentives; for example risks of social organisations organising themselves so as to maximise their achievements against pre-set measures, rather than to achieve the greatest social impact in their own eyes (GECES, 2014, p. 11).

Rather the sub-group recognises that there exists a range of approaches to measuring social impact, which differ in the detail of indicators and metrics, but which show some convergence on the main steps in the process that constitutes the groundwork for any measurement of social impact. Broadly, these steps involve, "identifying clearly the social impact sought, the stakeholders impacted, a 'theory of change' for social impact, putting in place a precise and transparent procedure for measuring and reporting on inputs, outputs, outcomes and for assessing thereby the impact actually achieved, followed by a 'learning' step to improve impacts and refine the process" (GECES, 2014, p.10, highlights maintained as per original). This is recognised to be an iterative process.

Instead of proposing indicators and metrics of its own, therefore, the GECES sub-group proposes the development of a standardised methodological approach for developing customised and context-sensitive impact assessments. This is based upon a set of principles and guidelines concerning: setting objectives; analysing stakeholders; measuring results; verifying and valuing impact; and monitoring and reporting. As no single set of indicators can be devised top-down to measure social impact in all cases the approach proposed by the GECES is to develop a framework for indicators. It is suggested that this, "should provide a broad structure into which the majority of cases should fit, showing differences between different types of intervention but recognising that for each type indicators are likely to be selected from a range" (GECES 2014, p.11).

Concepts and terminology play a key role in the development of the GECES guidance. The literature review undertaken by GECES as part of its remit identifies five key terms that recur in social impact assessment studies. These are adopted in developing the approach that GECES proposes. The approach distinguishes:

- Inputs: Resources used in delivery of an intervention
- Activity: What is done with those resources (the intervention)
- Output: How that activity touches the intended beneficiaries
- Outcome: The change arising in the lives of the beneficiaries and others
- Impact: The extent to which that change arises from the intervention

[10] The report of the sub-group states that "there is a range of approaches to measuring social impact, each of which promotes particular kinds of indicators, but that none of these has yet reached the state of a gold standard". Further "it is unlikely that any of these will become a gold standard since diversity of social need, intervention, scale and stakeholder interest demand different information and presentation of it" (GECES, 2014, p.10).

In respect of the last of these several adjustments are to be taken into account:

- Deadweight: To account for changes that would have occurred anyway
- Alternative attribution: To discount change attributable to other activities and actions
- Drop-off: To account for the decreasing effect of interventions over time
- Displacement: To account for negative consequences.

In coming to a standard capable of wide application, GECES draws a distinction between four elements in producing a meaningful measurement of social impact:

- Process: The series of steps and stages (mechanisms) by which the social innovators or funders investigate, understand and present how activities achieve change (outcomes) and impact in the lives of service-users and stakeholders.
- Framework: For each major area or sector in which social innovators intervene, a list of the most usual outcomes being targeted and, for each of these, a series of sub-outcomes that appear most regularly.
- Indicator: A particular way of attaching a value or measure to those outcomes and impacts.
- Characteristics: Qualities of the reported measurement of the outcomes and impacts that contribute to recognition of reliability, validity and robustness.

The GECES report also draws attentions to limitations that attach to any project seeking to establish a standard for measuring impacts of social innovation. It refers to: the intrinsic difficulty of capturing all impacts objectively, especially given the wide diversity of impacts; difficulties in capturing qualitative aspects of impacts, which can be underrepresented when using quantitative indicators; the need for measurement to be a proportionate activity that balances the wish for accuracy with the costs of measuring more precisely; the trade-off between comparability and context-sensitivity (the need for measurement and choice of indicators to be relevant in the specific case and context); and, the difficulty of sticking to any standard over a number of years in the fast-changing world of which social innovation is a part. The standard proposed by GECES therefore reflects "a balance between the needs of funders, investors and policy-makers for sound information on measurable social impacts with the need for proportionality and practicality" (GECES 2014, p.ii).

Within the limits implied by these caveats, the GECES sub-group develops a process to measure social impact, defined as "the social effect (change), both long-term and short-term, achieved for its target population as a result of its activity undertaken – taking into account both positive and negative changes, and adjusting for alternative attribution, deadweight, displacement and drop-off". This standard process involves five stages:

- Identifying objectives: This is concerned with both the objectives of the service being measured and the measurement objectives of the various parties.

- Identifying stakeholders: This is concerned with identifying who gains and who gives what and how.
- Setting relevant measurements: This is concerned with the theory of change that the social innovation uses to plan and implement its intervention; i.e. with how the activity is thought to achieve the outcomes and impacts most needed by beneficiaries and stakeholders. Measurements appropriate to explaining the link from activity to impact are set with input from major stakeholders.
- Measuring, validating and valuing: This is concerned with assessing whether – and to which extent – the targeted outcomes are achieved in practice.
- Reporting, learning and improving: This is concerned with using the measurements in regular reporting about services and their effectiveness to internal and external audiences.

The format for reporting is an integral part of the proposed GECES standard and is a means for quality assurance as well as of communication, so both structure and content are prescribed in terms of a set of points to be covered. These include:

1. An explanation of how the process has been applied
2. An explanation of the effects of an intervention in terms of outcomes, beneficiaries, and an account of attribution that considers deadweight, alternatives, drop-off, etc.)
3. The social innovator's logic model (theory or hypothesis) of change, suggesting how/why the activity caused or contributed to the outcomes and impacts
4. An identification of third parties having a role in the delivery of these outcomes and impacts, explaining how they contributed (which is important for alternative attribution)
5. An identification of those stakeholders whose interests are being measured and the nature of the gain to them (appropriately categorised)
6. A set of indicators for the identified impacts with explanation for the selection of these, how the indicator relates to the impact and the needs and interests of stakeholders and how these have been agreed with stakeholders.
7. A social and financial risk analysis covering the contingency that targeted social and financial outcomes are not delivered.

The standard for social impact assessment proposed by GECES marks some progress in that it provides for some consistency across assessment processes but proposes a flexible framework for selecting metrics and indicators so that these are contextually appropriate. It also continues in the (positive) direction of measurement based upon outcomes and impacts, rather than on outputs (see, also: Epstein and Yuthas 2014).[11]

[11] Epstein and Yuthas (2014) Measuring and Improving Social Impacts: A Guide for Non-Profits, Companies and Impact Investors, Greenleaf Publishing.

The standard is nevertheless narrower in scope than would be needed to cover a full range of possible evaluation questions. In part this is because it is based on a positivist approach to impact assessment, which assumes a greater clarity of means and goals than necessarily applies in all stages of a social innovation. An explicit argument in the GECES report, for example, is that "social enterprise needs to be defined and qualified by way of function, principle and primary purpose, and the impact measurement should be based upon and emerge from this" (GECES, 2014, p.7).

In part the scope is reduced because social innovators and their evaluation questions and needs are under-represented in the development of the standard relative to investors and their needs. While recognising that the context for and field of evaluation has been changing rapidly over recent years "to meet changing social, policy and investment needs", the proposed GECES standard heavily emphasises support for investment and financial decisions.[12] The report refers specifically to the global financial crisis as a motivation for measuring impact because of "the resulting heightened desire of funders and investors (public or private) to concentrate scarce resources on initiatives with an impact that can be demonstrated". Equivalently, the needs of service-providing organisations considered in developing the standard are mostly those relating to scrutiny and accountability; i.e. monitoring is designed to support transparency in explaining how funds are used and to demonstrate they are being used productively.[13]

More generally, in the GECES report the field of social impact measurement is considered to be a facet of the economic evaluation of impact, which has roots that can be traced back to the 18th Century. By implication, the proposed GECES standard is also developed from those roots and strongly reflects the approaches of economics, finance and accounting. In some reports there is recognition also that a 'perfect' system would be an illusory objective and that the aim should be more one of continuous improvement of schemes and metrics. It is suggested also that working to improve evaluation is an important part of the process of social innovation, since this imposes a discipline of conscious reflection and precision over the impacts that a social innovation seeks to have, the mechanisms through which impacts are created, and the strategies to be deployed for maximising the effectiveness of achieving impacts.

7. A contingency approach to impact measurement and evaluation

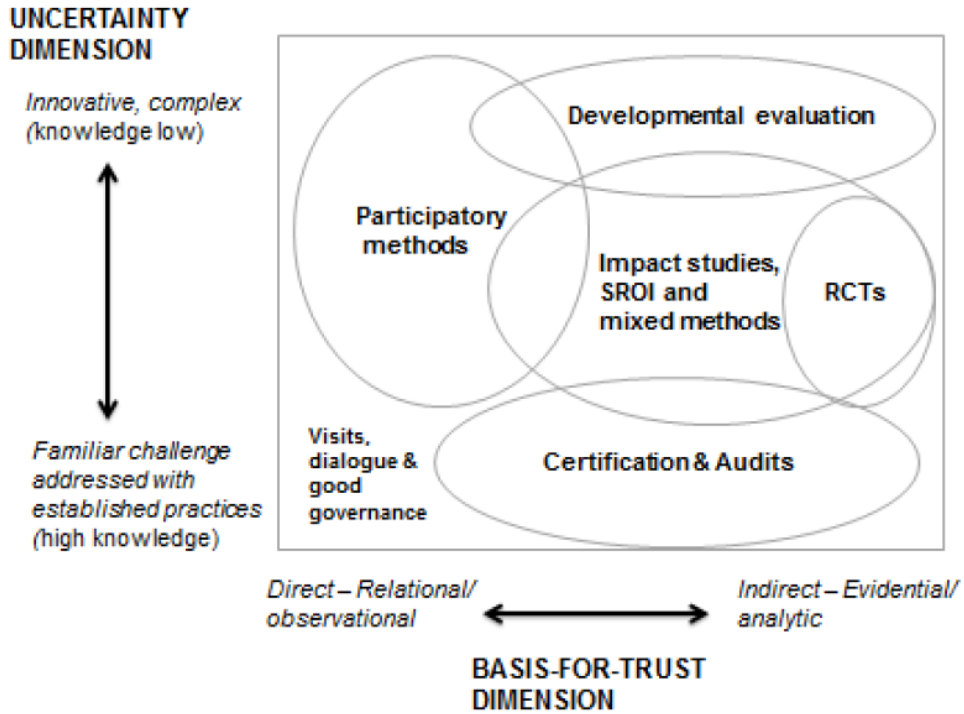
Accepting that formal methods for SIO have severe limitations in terms of do-ability and usefulness, Nicholls (in a report for CRESSI) develops a contingency approach. The starting point for the contingency approach is the question *how can organizations chose an approach that is appropriate to their concerns and context*? This question brings into focus the goals for evaluation.

[12] Thus for example, the GECES report states that: "In the case of all stakeholders, a key need for social impact measurement can be seen in decision making. The investor needs to evaluate the advantages of the impact achieved against the risks of investing. The fund manager needs to consider whether a given investment delivers both acceptable social and financial returns, as well as whether it meets policy and fund focus objectives. The service-user needs to understand the nature of the intervention, and the gains to be enjoyed by engaging with the service. The funder of the service, be it a public body, a service-user, or another party, needs to understand the value it gains and for which it is paying. The needs of all such stakeholders should be recognised and should be balanced" (GECES 2014, p.2).

[13] This extends to the use of impact measures to enable service providers and commissioners to improve effectiveness in delivery, where concerns expressed in the GECES report are for both technical efficiency (of the intervention) and financial efficiency (of the investment).

According to Nicholls, establishing a basis for trust is a goal that is important in its own right. A second important goal is offering an information basis for decision making for the social organisations and those interested in supporting it or evaluating it. These considerations led him to distinguish what he calls *The Basis-for-Trust Dimension.* For stakeholders it is important to know what the social organisation is doing and mission statements and descriptions of activities are valuable for this. For social organisations working on a pay-per-performance basis for public authorities and those funded by social impact investors this will not be enough. External funders may also have an interest to 'look under the bonnet' to "form judgements about the quality of the management team and the challenges and prospects facing the business" (Nicholls, 2015, p. 20). In general, direct contact, dialogue and observation are ways to "allow confidence to build up (or not) and reduce the need for the costly generation and processing of abstracted evidence and reports" (ibid, p. 15). The need for quantitative analysis also will be smaller when the initiative is embedded in a community, a social movement or field of professional activity with well-described practices and ethos as to what is permitted and what not (ibid, p. 20). The second dimension of the contingency framework is the degree and amount of uncertainty. Uncertainty is high for novel and complex initiatives and low for well-established ones. Both trust and uncertainty will affect the confidence of external stakeholders in the occurrence of positive outcomes. The value of the Contingency Framework is that it offers guidance in terms of the most appropriate impact measurement approach for different contexts (shown in Figure 2)

Figure 2: A Contingency Model for Social Impact Measurement



Source: Nicholls, 2015 (p. 21)

According to Nicholls, the central area of the graphs is the most difficult because uncertainty is considerable and the basis for trust is unclear. Here, certification and participative approaches will not be sufficient or appropriate for some stakeholders, but the initiative may not be able to undertake certified method because of costs. Nicholls feels that especially here the SROI has a great to offer because of its flexibility in terms of combining (quantitative) evidence with participatory relationship building.

8. Developmental evaluation for innovation

After having discussed positivistic models of evaluation, especially, SROI, we now offer a deeper discussion of developmental evaluation *as a model of evaluation for innovation*. In developmental evaluation, the focus is not on impact evaluation but on possibilities for *increasing* positive impacts through innovation. Developmental evaluation fits with internal needs for monitoring and external ones for summative evaluation and constitutes an important complement to positivistic forms of evaluation that are grounded in measurement rather than in exploration of possibilities.

Antadze and Westley (2012) argue that Milbergs and Vonortas (2004) made a major contribution in their analysis of innovation metrics by recognising that innovation is a process that entails change within complex systems, that it is multidimensional, that it is uncertain and that these aspects are not captured through conventional approaches. This line of argument is taken up also by Morris (2011) who says that the pursuit of innovation necessarily involves a venture into the unknown and concludes from this that "opportunities will be lost if we try to pin these unknowns down too early" in our (innovation and evaluation) processes. These aspects "open the possibility that the evaluation of the impacts of innovation is itself an experiment" (Antadze and Westley, 2012, p. 143).

This reflection on innovation as a complex process operating and interacting with complex systems is at odds with the established paradigm. The established paradigm of social impact measurement derives from a positivist tradition that defines social innovation narrowly as a product or service, rather than as a process. The range of evaluation questions asked, the scope of impacts and values considered, and the types of assessment methods used within the established evaluation paradigm are therefore all much narrower than is needed to capture the full range of impacts when social innovation is defined as a process that entails change within complex systems. Yet, crucially, it is in this latter understanding of social innovation that it holds potential to transform society; i.e. by contributing to broad and lasting changes in social relations, institutions, constructs and behaviours.

Elaborating further on conventional measuring approaches and the difference with approaches that might be needed to evaluate social innovation, Antadze and Westley (2012) group currently-used approaches according to whether they are focused on single or on multiple outcomes and according to whether approaches are designed deliberately to capture particular outcomes (which they refer to as approaches of 'deliberate design') or whether they are open to reconfiguration by emergent qualities of the transformation they are measuring ('emergent design').

They find that most currently-used approaches are characterised by deliberate design, with either a single outcome (economic) focus or a multiple-outcome focus. They also show that most

approaches are applied to measure concrete phenomena, such as products, services, and behaviours, rather than something more abstract, like a process or an idea. This, they conclude, is unsurprising, precisely because conventional notions of evaluation have been based generally on traditions of positivist science, leading to an approach that is causal and linear and based on input-output measurements. As a result, they say, "what can be counted tends to be what is evaluated" (p. 144). Concrete phenomena are relatively easy to evaluate. Social innovation, by contrast, demands a link among complex and abstract phenomena, social processes, and multiple outcomes.[14]

Antadze and Westley (2012) draw attention to an alternative approach to evaluation, referred to by Patton (2011) as 'developmental evaluation', which arises from this perspective on and definition of social innovation as" a complex process" determined by "its impact on the broad system", rather than by the conventional definition of social innovation as a specific outcome in the form of a product, service or behaviour. In the perspective of development evaluation the evaluation focus shifts from measuring social innovation as a product or service to evaluating it as a process that has impacts.

By introducing 'development evaluation' Patton draws a distinction with more usual 'formative' and 'summative' modes of evaluation. Effectively, Patton is introducing a third and new mode of evaluation that is designed to serve a hitherto overlooked purpose. Both formative and summative evaluations are tests of a model. Formative evaluation is used to improve a model and bring it to some final stage of refinement. Summative evaluation is used to determine the success and effectiveness of the final model and, especially, to help decide whether it should be continued, extended or disseminated. It implies the existence of well-defined goals, an optimal solution, a targeted intervention and a fairly stable environment. Developmental evaluation, by contrast, "supports innovation development to guide adaptation to emergent and dynamic realities in complex environments". It suggests constant movement back and forth between problem and solution in support of an ongoing and continuous development process. Here Patton is describing a constructivist role for evaluation and a role in the adaptive management of upscaling processes.

Developmental evaluation (see Table 2) is "purpose-and-relationship driven and not method-driven"; making method decisions is part of the process (Patton, 2011, p. 288). "It's all about persistently asking questions and pursuing credible answers in time to be used" (Patton, 2011, 288). **Typical questions** to be used are (pp. 46-47):

- *What is the baseline understanding of the situation?*
- *What are the vision and values that will guide innovation?*

[14] This raises the issue of attribution and causality in social innovation assessment. These are difficult issues to handle generally because social sector impacts can be caused by multiple factors and actors. However, there are additional difficulties of assessing impact and ascribing causality in a complex process when social impacts may result from actions by more than one organisation and lines of influence may be too long, complex and indirect (Earl et al. 2001).

- *What are the initial conditions and the nature of the environment within which action will occurs?*
- *What is meant by innovation?*
- *What do rapid feedback and initial results reveal about progress in desired directions?*
- *What's considered "working" and "not working" as exploration unfolds and innovation is undertaken?*
- *What criteria emerge to tell the difference between working and not working?*
- *What processes and outcomes generate enthusiasm? Why?*
- *How is the programme as an intervention system connected to and affected by larger systems in its environment?*
- *What are the trends in those larger systems?*
- *What can be controlled and not controlled, predicted and not predicted, measured and not measured?*
- *How to distinguish signal from noise to determine what to attend to?*
- *What innovations emerge that merit more formal implementation as pilot programmes ready for formative evaluation?*

It is based on a number of **key conditions**, which are that:

- Social innovators have a strong vision and commitment to making a difference.
- There is a willingness and capacity to act and innovate under conditions of uncertainty and turbulence.
- There is a commitment to use data and rapid feedback to make sense of what emerges during exploration and innovation and to use those emergent understandings to guide next steps.
- Funders are willing to try out and trust the innovation process and developmental evaluation as a way of monitoring what emerges.
- Evaluators are capable of operating without predetermined clear, special and measurable outcome or a well-defined logic model.

Ontologically, it is based on the view that "conclusions include reasoning, critical thinking, judgment, and argument—and cannot be reduced to methods and data".[15]

Patton (2011) lists five main uses for developmental evaluation: for ongoing development in adapting a project, program, strategy, policy or other initiative to new conditions in complex dynamic systems; for adapting general principles to a new context as ideas and innovations are taken from elsewhere and developed within a new setting; for developing a rapid response (in real time) in the face of sudden major change; for developing the performance of a potentially scalable innovation; and for evaluating major systems changes and cross-scale developments to provide feedback on unfolding systems changes, evidence of emerging tipping points and/or on how an innovation may need to be adapted as it is taken to scale in the effort to have broader impact.

[15] Reconstructing potentially motivating concepts, such as fairness, autonomy, resilience and sustainability, (in terms of one's own understanding and that of others) may be part of the process, giving the evaluation an element of "phronesis" (practical judgment) (cf. Loeber, 2007 and Avelino and Grin, 2017).

Table 2: Features of Developmental Evaluation

Purpose	Key conditions	Priority questions	Common evaluation approaches	Key factors affecting use
Developmental evaluation				
▪ Help social innovators explore possibilities for addressing major problems and needs, and identify innovative approaches and solutions. ▪ Develop promising innovations. ▪ Support adaptation in complex, uncertain, and dynamic conditions. ▪ Document what actions innovators engage in, the short-term results and consequences of those actions, and their connections to the larger vision of the innovators. ▪ Identify *emergent* processes and outcomes that accompany innovation, and support making sense of their implications. ▪ Support ongoing development and adaption to changing conditions.	▪ Social innovators with a strong vision and commitment to making a difference. ▪ Willingness and capacity to act and innovate under conditions of uncertainty and turbulence. ▪ Commitment to use data and rapid feedback to make sense of what emerges during exploration and innovation, and use those emergent understandings to guide next steps. ▪ Funders willing to try out and trust the innovative process and developmental evaluation as a way of monitoring what emerges. ▪ Evaluators capable of operating without predetermined clear, specific, and measurable outcomes or a well-defined logic model.	▪ What is the baseline understanding of the situation? ▪ What are the vision and values that will guide innovation? ▪ What are the initial conditions and the nature of the environment within which action will occur? ▪ What is meant by innovation? ▪ What do rapid feedback and initial results reveal about progress in desired directions? ▪ What's considered "working" and "not working" as exploration unfolds and innovation is undertaken? ▪ What criteria emerge to tell the difference between "working" and "not working"? ▪ What processes and outcomes generate enthusiasm? Why?	▪ Developmental evaluation. ▪ Systems-and-complexity-based interactive design. ▪ Rapid assessment, rapid feedback. ▪ Emergent evaluation. ▪ Real-time evaluation. ▪ Ongoing environmental scanning and outcomes monitoring. ▪ Reflective practice. ▪ Participatory action research. ▪ Network analysis. ▪ Systems change mapping.	▪ Innovators and developmental evaluator(s) able to work together in partnership: mutual respect and trust. ▪ Getting busy and action-driven innovators to value and spend time in sense making, reflection, and data interpretation to inform ongoing innovation. ▪ Openness to what emerges. ▪ Adaptive capacity. ▪ Tolerance for ambiguity and uncertainty ("getting to maybe"). ▪ Balancing quality and speed of feedback. ▪ Nimble, agile. ▪ Integrate and synthesize multiple and conflicting data sources.

All of these different uses for development evaluation stress the importance of context. Context sensitivity includes paying attention to the primary intended users of an evaluation, priority uses, the political environment within which the evaluation occurs, the stage of development of the innovation and other factors relevant to the use of the evaluation. A key conclusion drawn by Patton (2011) and reinforced by Antadze and Westley (2012) is that standardized metrics are not appropriate for developmental evaluation given the diversity of innovation contexts, so development of metrics and their continuous review and revision (as emergent metrics) needs to be built-into the social innovation process as a central aspect of development evaluation.

> "*Evaluation isn't something to incorporate only after an innovation is underway. The very possibility articulated in the idea of making a major difference in the world ought to incorporate a commitment to not only bringing about significant social change, but also thinking deeply about, evaluating, and learning from social innovation as the idea and process develops*." (Westley et al., 2006)

Antadze and Westley further argue that formative, summative and developmental evaluations are not exclusive; rather they can play complementary roles at different stages in the development of social innovation. Using the adaptive cycle of Gunderson and Holling (2002), Antadze and Westley (2012) suggest that formative evaluation can support the exploitation stage by fine-tuning a model, summative evaluation can support the more stable and grounded conservation phase by judging overall effectiveness of a model, and developmental evaluation can support the reorganisation and release phases "where social innovators need to make sense of the emergent opportunities, understand the ongoing dynamics, and try out new ideas and approaches" (Antadze and Westley 2012, p. 146). As social innovation processes progress this will imply change in requirements not only in terms of evaluations, but also in terms of suitable business, financial, learning and governance models.

Next to the stage of development, the characteristics of the context space matter. Kirkland (2013) distinguishes three evaluation 'spaces': simple, complicated and complex.[16] Complex space call for open exploration and put a premium on flexibility. Complicated space involves a better, but still not fully, understood space where there is still room for debate and decision about the best way forward. Here the focus is on measuring what is thought will happen, so the focus is on measuring sought outcomes. There is a need also to consider the efficacy of different approaches, so that different variants of an intervention might be tested and evaluated to see which gives best results. There is still a need to be mindful of unintended consequences, so evaluations systems are needed to capture those things that do not show up with established measures. Simple space by contrast is where a good deal is known already about the context, the intervention and the mechanisms and nature of its effects. Here the focus of evaluation is on measuring what is thought or is expected to happen. Approaches, like randomised controlled trials can be appropriate here because the effects sought are known and the need is for the most robust method to detect these.

For social innovation, useful alternatives to developmental evaluation exist. One proposed approach is the model of dynamic evaluation, which is based on innovators coming together to share intervention suggestions and experiences via stories (Kieboom and Vahl, 2014).[17] Like

[16] http://www.nominettrust.org.uk/knowledge-centre/blogs/evaluating-social-innovation-what-and-when

[17] This section is based on https://www.kl.nl/wp-content/uploads/2013/01/MD-Deliverable-3.2-Case-Study-Education-Pioneers.pdf, https://www.kl.nl/wp-content/uploads/2014/04/pres_mk_de2.pdf

developmental evaluation, dynamic evaluation seeks formative improvements. It exploits people's natural interest in stories and their proneness to understand causal elements of "why and how" if these are presented in the form of a narrative.[18] The stories themselves may need further articulation, systematisation and scrutiny, which is done by bringing people in direct contact with each other (allowing them to ask questions) and by breaking down the stories into 4 parts: the problem that was being addressed; the idea; the actions; and the results obtained. It helps social innovators with the critical issue of finding partnerships of collaboration and subjects them to external feedback, which its proponents claim is something to which developmental evaluation is less suited. [19]

9. Evaluation anxiety

Imposed forms of evaluation create anxiety. Funders (or governments) often want social impact to be demonstrated in return for funding. However, initiatives often do not have the necessary expertise or resources and would rather focus the monitoring on different aspects, especially if their intervention model is an asset-based approach, since evaluations using baseline approaches tend to focus on deficits and needs rather than on assets and potentials. Social innovators also often prefer to spend their scarce resources on *making* impact rather than on measuring it.

Action research projects can provide useful support in such cases, as shown by the monitoring and evaluation of eco-localisation projects in the UK (described in Hobson et al., 2016). In the action research, the interest in monitoring and evaluation (M&E) was explored amongst low carbon community groups and partnerships (LCCGPs). The interest in M&E was examined and exploited in a step-wise process, which started with a one-day M&E workshops in Oxford, London and Manchester. The aim of the research project was not so much to collect and analyse M&E data, but rather "to explore what happens when groups are given the space, resources and tools to do so themselves, in-keeping with calls for 'a more holistic evaluative frame' when examining the impacts of community groups" Hobson et al., 2016, p. 1398).

In a second step, experience was gained with M&E methods in the course of several months. Each group was asked to trial at least two M&E tools and was allocated one research team member as support over the trial period. When the trials had ended, feedback was being sought from each participating group on the process and tools. Overall, the feedback was "positive and constructive". Participants said that the evaluation led them into discussions on their mission and 'theories of change'. It enabled some groups "to plan future projects in line with the desire to create specific impacts" rather than using M&E for capturing "outcomes after-the-fact". In the words of a Transition Network staff member, the M&E exercise 'stimulated me to do some of the work that I've been wanting to do, like around theories of change because that was the first thing that confronted me when I started doing this project. It was ''so what's your theory of change and how are you going to fit the impacts and what you want to measure into that logic model''? (p. 1401)

The exercise revealed anxiety towards the element of judgment and the possibility that it might "highlight shortcomings within the group, either in terms of tensions between members or

[18] A narrative is "an account of the unfolding of events, along with an effort to explain how and why these processes and events came to be". http://understandingsociety.blogspot.nl/2014/02/a-causal-narrative.html

[19] Personal communication of Marlieke Kieboom in an email.

perceived missing skills and constituents". The exercise also showed that not all of them are interested in expansion. One group in particular, chose "not to take up the mantle of becoming a well-funded and expansive group, able to deliver on national level policy goals and prove their impact to attract competitive external funding" (Hobson et al., 2016, p. 1406). People in this group were content with their current size. For this group, "doing projects that were of interest to them – not ones that delivered the most quantifiable impacts – was their intended pathway"

Another important conclusion found by the researchers is that action research can help people to find their own way of dealing with M&E, in ways that help them gain traction with the public and specific stakeholders, through expressions of their ethos and particular representations of impact (Hobson et al., 2016 p. 1505).

10. Monitoring of context to find suitable partners and strategies

The importance of context sensitivity has been mentioned in section 8. Here, we examine this element further. At the moment, there is a favourable opportunity context for social innovation (Weaver and Marks: 2017a, 2017b; Marks et al, 2017) because there are shifts underway that favour social innovation initiatives as service providers. We see, for example, the shift from local authorities acting as direct service providers to becoming service commissioners in areas of adult social care, urban poverty relief, etc. Another relevant development is the discussion about the current welfare systems and experiments with new forms of social welfare provision and rules, such as incentives for welfare claimants to join social innovation initiatives, experiments with basic income, possibilities for health service beneficiaries to engage in activities that will help them to improve their own health or engage them in helping improve the health of others (Weaver, Boyle and Marks, 2017).

But such evaluations come with additional challenges, often in the form of dilemmas. For example, when establishment actors set the agenda and expect the SIs to play along, they are creating problems of co-opting. An example is the UK government seeking to signpost benefits claimants to time-banks without due consideration of how easily this could overwhelm the absorptive capacities and resources of time banks. SI's can easily collapse when agendas are imposed top-down by (single-topic) agencies. The development goals of funders and SI leaders may also conflict with the wishes of volunteers not to be judged and their desires for keeping to old ways of doing. SI leaders must also take care not to lose the grassroots element.

To go to scale, social innovation initiatives may use different routes (Table 3). They may campaign for recognition and support by government and incumbent actors, they may build a delivery network, form strategic partnerships and grow an organisation to deliver.

The choice of stakeholder-collaborators is a critical issue for going to scale (NESTA, 2014). Social innovators are advised to differentiate people into those who'll pay, those who'll take part, use and benefit, and those who'll devote their time to the innovation and make it happen (NESTA, 2014, p. 6.). This helps them to identify allies and work with them in mutually beneficial ways.

Table 3: Scaling routes for social innovators

Scaling route	Models and approaches	Activities
Influence and advise	Campaigning and advocacy Consultancy Training	Public speaking Publishing Engaging with policymakers Communicating via traditional and social media Advising or training others
Build a delivery network	Federations and membership models Communities of practice Kitemarks and quality marks Licencing Franchising Delivery contracts Collaborations	Representation Advocacy and awareness raising Transferring knowledge, codifying processes, sharing good practices, providing tools Training, support and quality assurance Community and movement building
Form strategic partnerships	Strategic alliances Mainstreaming into the public sector Piggybacking on another organisation's infrastructure Joint ventures Mergers and acquisitions	Brokering and managing partnerships with other organisations that allow a step change in scale Transferring knowledge Creating a sense of common values and mission
Grow an organisation to deliver	Setting up new branches Growing delivery capacity of a central team	Building staff and team capabilities Raising funds/investment Developing organisational capacity and systems

Source: NESTA (2014, p. 5)

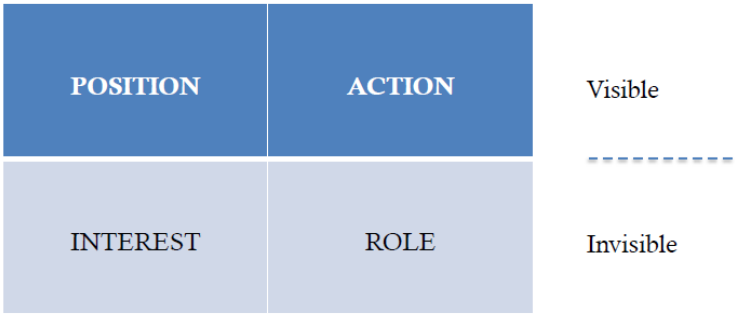
A possible way of investigating motivations of actors (especially those who hold the key to system change and transformative impact) is the PAIR matrix of Henk Diepenmaat of Actor Management (Figure 3). The PAIR-analysis is a method that helps to improve on understanding of the players in a certain field of players, where *P* stands for *Position*, *A* for *actions*, *I* for *Interests* and *R* for *Role*.

- **'Position':** A position is the espoused viewpoint, which is usually declared to others in public, in the press, in a position paper, or at a private negotiation table. An example statement is that the 'empowerment' of citizens and communities is "*absolutely essential to our economic, social and political future. If our local economies are vibrant and strong we are far less vulnerable to global shocks or the failures of a few dominant industries. If people know that their actions can make a real difference to their local communities, they're far more motivated to get involved – and civic pride is revived.*" (Cameron, 2009: n.p.). Positions are often about the concrete characteristics of the *present* or the *future* situation of an

underlying change perspective, as seen by the actor who is taking up a position. They say something about the intentional logics of an actor.

- **'Role'** is a name that covers a set of acting- and other perspectives that go together and are typical for a certain actor (for example a certain *business company, a research organisation*, *intermediary*, *advisory body*, *mayor*, *teacher*).
- **'Interest'** is something that an actor wants for himself (to be re-elected, to achieve particular goals, to have a good income or to maintain the one he or she has). If an interest is under threat, the actor will take a curative or preventive action.
- **'Actions'** are the actual conducts/activities carried out by a party. Actions are the entire concrete manifestation of interpreting one's own intentional logics in a world shared with other actors. (Diepenmaat, 2011).

Figure 3: The pair matrix



The PAIR-matrix is made up of six "halves": a bottom half, a top half and 2 diagonals *(from bottom left to top right and from top left to bottom right)*

Source: Diepenmaat

The PAIR matrix helps to think about partners for collaboration in a more structured way, and for offering guidance for what to do. The PAIR-matrix is made up of six "halves": a bottom half, a top half and 2 diagonals (from bottom left to top right and from top left to bottom right).

- **Top half: positions and actions.** This is what can be observed directly: *positions* are expressed and *actions* are carried out. Interests and roles, by contrast, are to a large extent invisible and are hidden in the actors' inner selves. The advocacy for Big Society by the Cameron government in the UK was suspected to be a cover for spending cuts (Small Government). Positions are only credible if they are combined with action, which can be asked for.
- **Bottom half: interests and roles.** The bottom half is the invisible half of the PAIR-matrix. It is also the half that allows us to overcome the deadlock in the multi-actor process. After all, interests can be served by more than one position and roles can be fulfilled by more than one action. That is why looking for cooperation is more likely to be successful if one is not focussed on the positions that have been taken up and the actions that have been carried out (top half), but on interests and roles. When looking for cooperation, one should look down to the bottom half. It may be less tangible, more difficult to grasp, but *that* is where the 'wiggle room' is, if there is any.

- **The diagonal from top left to bottom right: positions and roles.** Ideally, positions and roles mutually enhance each other: important positions have competent and recognisable roles; and robust roles are secured by a clearly articulated position. But if a position is created based on a role rather than based on an interest, this is highly vulnerable to change. SII should be aware of that.
- **The diagonal from top right to bottom left: interests and actions.** A similar consideration goes for the diagonal interests-actions. Actions should be rooted in interests and interests should be secured by actions. Actions that fit with their own natural roles and direct interests, but that are not rooted in a deeper interest, can change when deeper interests can be served better

The behaviour implications of the PAIR matrix are as follows: *If you want to form coalitions, focus on the bottom half. If you want to enrol new actors, look at the left half. If you want to know whether actions and roles fit together perfectly, look at the right half. If you want to see through monopolies, look at the diagonal from top left to bottom right. And if you want to identify shortfalls in action (given interests) and actions which are weakly connected to interests (and thus may disappear), you should look at the diagonal from top right to bottom left.*

Researchers of Wageningen University and the VU Amsterdam developed a method called **Reflexive monitoring in action** (RMA) which offers a more systematic approach to the scanning of opportunities for system innovation by undertaking a system analysis, stakeholder analysis, causal loop mapping and other activities. It is a form of developmental evaluation with an important role for research, where "the insights gained from monitoring are tried and experimented with in the projects and activities" (van Mierlo et al., 2010).

11. Conclusions

In reviewing the literature on and tools for social impact evaluation it is apparent that there are different perspectives on evaluation systems and tools and that these reflect different perspectives on social innovation. The perspectives of social sector actors differ from those of for-profit sector actors, so that the adaptation and application of models of performance used in the for-profit sector – such as measures of profit or productivity – are potentially problematic for social sector actors. For social sector actors financial performance is often a means rather than an end of social sector activity. Social sector actors are mission-focussed and therefore they need mission-based measures for performance evaluation. It has nevertheless been observed that, to date, the debate around the measurement of social innovation outcomes and impacts "mainly reflects the perspective of private social finance that is attempting to be more strategic about its capital allocation" (Ebrahim and Rangan, cited by Antadze and Westley 2012, p. 135). Existing tools for social impact assessment are therefore ones mostly borrowed from economics and accounting. Not being developed specifically for this field and for the needs within it, they are not always appropriate.

Even if there is some overlap in interests, approaches, metrics and data needs, stakeholders' different evaluation needs therefore require different focuses and methods, as even what is to be evaluated – the productivity of finance, the effectiveness of activities, the nature of benefits experienced by different beneficiaries – differs. Further, the evaluation needs of a social organisation are likely to evolve as the organisation and its innovations evolve, diffuse and go to scale. Overall, because of the importance now applied to the sector, this is a very fast-moving field where progress is being driven by the imperatives and opportunities (political, financial

and social) to harness social organisations and social innovation both to maximise social impact and to optimise social investment.

The diversity and dynamics of needs are stimulating new work to develop evaluation frameworks based on generalizable principles and protocols that seek to combine consistent guidelines with flexibility over indicators and metrics. These put premium on understanding the processes and pathways through which outcomes and impacts are produced. However, even the most-recently proposed frameworks are still being developed from a largely positivist perspective. Furthermore, the issues of attribution and causality, which are difficult issues to handle generally in social innovation assessment,[20] are especially problematic when lines of influence may be long, complex and indirect and when outcomes are influenced by multiple factors and interactions among these, as is likely to apply when concern is for transformative societal change (Earl et al. 2001).

Policymakers, researchers and social innovators interested in whether and how bottom-up social innovations might contribute to positive societal-level change and in understanding how any such contribution might be maximised have a yet different concern from those that are typically addressed within the established evaluation paradigm that is focussed on what has been achieved. The question of how to enlarge positive impacts requires an approach to evaluation that is very different from the approaches developed under a positivist perspective that seek evidence in relation to pre-specified cause-effect links and chains. Whereas in a traditional impact assessment the focus is on measurement of outcomes and attribution, developmental evaluation uses monitoring to obtain new data and understandings about the complex dynamics behind positive outcomes and reasons for failure, with the help of stories of success and failure in combination with a look at data.

All methods have strengths and uses, but also limitations. For measuring what has been achieved with allocated funding, SROI is a useful method. For fostering innovation and improvement of social innovation, developmental evaluation (Patton, 2011) and dynamic evaluation (Kieboom and Vahl, 2014) are useful methods. There is a need to select evaluation methods that are 'fit for purpose' and there is added value in using combinations of methods. We propose that SROI pays more attention to the stories of people involved (those helped by a SII and the professionals in the case of help services), to explain to interested outsiders what the SI is about and to understand the role and influence of the SI on outcomes and impacts. Action research can be used to find useful ways of monitoring, as shown by the experiences of the eco-localisation project of Kersty Hobson and co-workers (Hobson et al., 2016). In action-based forms of evaluation, such as developmental evaluation, evaluators do not take distance but immerse themselves in contextual specifics, they "co-create interpretations and arguments, examine the evidence and reason together" (Patton, 2011, p. 287). Focussing on only those factors that can be measured (as happens in randomised controlled trials), may keep from view essential factors and processes that link causes to effects.

[20] The difficulty arises, inter alia, because social sector impacts can be caused by multiple factors and actors and the precise contribution of a particular activity to an outcome and impact may not be separable.

Acknowledgements

This article is based on research carried out as part of the Transformative Social Innovation Theory ("TRANSIT") project which is funded by the European Union's Seventh Framework Programme (FP7) under grant agreement 613169. The views expressed in this article are the sole responsibility of the authors and do not necessarily reflect the views of the European Union.

References

Antadze, N., and Westley, F.R. (2012) Impact Metrics for Social Innovation: Barriers or Bridges to Radical Change? *Journal of Social Entrepreneurship* 3(2): 133-150.

Avelino, F., Grin, J. (2016). Beyond deconstruction. A reconstructive perspective on sustainability transition governance, *Environmental Innovation and Societal Transitions* 22: 15–25

Bishop, M. and Green, M. (2008) *Philanthrocapitalism: how giving can save the world.* New York: Bloomsbury Press.

Burgess, R. and Pande, R. (2005) "Do rural banks matter? Evidence from the Indian social banking experiment". *The American Economic Review*, 95 (3), 780–795.

Canadian Task Force on Social Finance (2010).*Mobilizing private capital for public good*. Toronto: Social Innovation Generation.

Deekens, G., de Ronde Bresser, S., Duchène, S., Zonneveld, A., den Heijer, I. (2015) De maatschappelijke stand van het land, publicatie van Kracht in Nederland, Oktober 2015.

Diepenmaat, H. (2011), Multi-actor procesmanagement in theorie en praktijk, Serie Samenleving in perspectief Deel 4, Almere: Parthenon.

Eadery, Y., (2006). Ethical developments in finance: implications for charities and social enterprise. *Social Enterprise Journal*, 2 (1), 82–100.

Elder, J., Didham, R. Sano, D. (2016) Process Indicators to Measure Intermediate Progress of Social Impacts of an Individual Organization's Transition-Related Research, paper for International Sustainable Transitions Conference IST 2016, Wuppertal, Germany, September 6-9, 2016

Edwards, M. (2009). Why 'philanthrocapitalism' is not the answer: private initiatives and international development. In: M. Kremer, M. van Lieshout, and R. Went, eds. *Doing good or doing better: development policies in a globalizing world*. Amsterdam: Amsterdam University Press, 237–254.

Fujiwara, D. and Campbell, R. (2011) Valuation Techniques for Social Cost-Benefit Analysis: Stated preferences, revealed preferences and subjective well-being approaches – a discussion of the current issues. HM Treasury and Department for Work and Pensions, July, 2011.

GECES (2014) Proposed Approaches to Social Impact Measurement in the European Commission legislation and practice relating to: EuSEFs and the EaSI, GECES Sub-group on Impact Measurement. http://ec.europa.eu/internal_market/social_business/docs/expert-group/social_impact/140605-sub-group-report_en.pdf

Harding, R., (2007) Social Entrepreneurship Monitor. London: London Business School and Barclays Bank.

Hobson, K., Mayne, R., Hamilton, J. (2016) Monitoring and evaluating eco-localisation: Lessons from UK low carbon community groups, Environment and Planning A, 48(7): 1393–1410.

Hoffmeister, K.L., (2016) Identifying the Challenges of Social Entrepreneurs in Applying the Shared Value Concept in Practice, Master Thesis of Sustainability Science and Policy programme, Maastricht University - ICIS

Jørgensen, M.S., Dorland, J., Pel, B., Wittmayer, J., (2015) Characterisation and comparison of case study findings – Batch 1 cases, Deliverable 4.2 TRANSIT project.

Kieboom, M., and Vahl, M. (2014). Social Innovation and the magic of experience. Emergence by Design - D. 3.2: Case study Education Pioneers. Amsterdam: Kennisland. Licensed under CC-BY.

Loeber, A., (2007). Designing for phronèsis: Experiences with transformative learning on sustainable development. Critical Policy Analysis, 1(4): 389–414.

Marks, M. Hogan, L. and Weaver, P. (2017), "Success Case Examples of Sustainable Social Innovation Resourcing Strategies in the United States", TRANSIT working paper #12, July 2017.
McWade, W., (2012) The role for social enterprises and social investors in the development struggle. *Journal of Social Entrepreneurship*, 3 (1): 96–112.
Milbergs, E., & Vonortas, N. (2004). Innovation metrics: Measurement to insight. Center for Accelerating Innovation and George Washington University, National Innovation Initiative 21st Century Working Group, 22.
Moore, M-L., Westley, F. R., and Nicholls, A. (2012) "The Social Finance and Social Innovation Nexus", Special Issue "Funding Experiments: The Role of Social Finance in Supporting Social Innovation". *Journal of Social Entrepreneurship*, 3(2): 115-132.
Morris, L. (2011). The innovation master plan: the CEO's guide to innovation. Innovation Academy.
NESTA (2014) Making it big, report by Madeleine Gabriel for NESTA.
Nicholls, A., (2010) The institutionalization of social investment: the interplay of investment logics and investor rationalities. *Journal of Social Entrepreneurship*, 1 (1), 70–100.
Nicholls, A., (2010) The legitimacy of social entrepreneurship: reflexive isomorphism in a pre-paradigmatic field. *Entrepreneurship Theory and Practice*, 34 (4), 611–633.
Nicholls, A., and Pharaoh, C. (2007) The landscape of social investment: a holistic topology of opportunities, Third Sector Research Centre.http://www.birmingham.ac.uk/generic/tsrc/documents/tsrc/reports/SEIF/SEIFPhaseOneThelandscapeofsocialinvestmentintheUK.pdf
Nicholls, A. and Murdock, A. (2012) Social Innovation: Blurring Boundaries to Reconfigure Markets, Palgrave Macmillan, Hampshire, England
Nicholls, A. (2015) Synthetic Grid: A critical framework to inform the development of social innovation metrics, CRESSI Working Papers No. 14/2015.
Patton, M.Q. (2011) *Developmental Evaluation: Applying complexity concepts to enhance innovation and use*, New York: The Guildford Press
Patton, M.Q. (2016) *The Developmental Evaluation Mindset: Eight Guiding Principles*, in M.Q. Patton, K. McKegg and N. Wehipeihana (eds) Developmental Evaluation Exemplars. Principles in Practice, The Guildford Press, New York, pp 289-312.
Putnam, R. (1995) "Tuning In, Tuning Out: The Strange Disappearance of Social Capital in America", *Political Science & Politics*, 28(4): 664-683.
Van Mierlo et al., (2010) Reflexive Monitoring in Action. A guide for monitoring system innovation.
https://www.researchgate.net/profile/B_Regeer/publication/46383381_Reflexive_Monitoring_in_Action_A_guide_for_monitoring_system_innovation_projects/links/0fcfd505ad7d904336000000/Reflexive-Monitoring-in-Action-A-guide-for-monitoring-system-innovation-projects.pdf
Weaver, P.M., Boyle, D. and Marks, M.B. (2017) Addressing the 'missing' preventive infrastructure: a process toward cross-governmental collaboration in effective public services; paper for the European Prevention Society, Vienna Conference, 2017; Groundswell Research Associates, London, UK.
Weaver, P.M., Marks, M.B., Scropke, C. and Hogan, L. (in press) Social innovations and enterprises as integrative resources, strategies, and models: In: Ramos, B., Miesing, P. and Briar-Lawson, K. [Eds.] *Social and Economic Development in Transitioning and Market-Based Economies: Social Enterprises, Entrepreneurship, and Microenterprise as Integrating Institutions*: State University of New York Press, Albany.
Weaver, P.M. and Marks, MB (2017b) Social Innovation Resourcing Strategies and Transformation Pathways: A First-cut Typology, TRANSIT working paper #11, July 2017.
Weber, H. (2004) The 'new economy' and social risk: banking on the poor. *Review of International Political Economy*, 11 (2), 356–386.
Westley, F., Zimmerman, B., and Patton, M.C. (2006) *Getting to Maybe*, Random House: Canada.

Westley, F., Antadze, N., Riddell, D., Robinson, K. & Geobey, S. (2011-04). Pathways to System Change.

Westley, F.R. and Antadze, N. (2010) Making a difference: strategies for scaling social innovation for greater impact. The Public Sector Innovation Journal, 15 (2), article 2. Available from: http://www.innovation.cc/scholarly-style/westley2antadze2make_difference_final.pdf.

Wood, C., and Leighton, D. (2010), Measuring social value. The gap between policy and practice, a report for DEMOS, http://www.demos.co.uk/publications/measuring-social-value

Appendix: Some examples of M&E activities within the TRANSIT case studies

Network	Social innovation initiative	Monitoring element	Process and impact
DESIS	Creation of Sustainability Report by Study Center of Design & Technology (CedTec/DESIS LAB), for the State University of Minas Gerais in Brazil for the year 2011.	Sustainability report. The report was prepared following the guidelines of the GRI - Global Reporting Initiative.	The report was created despite absence of support from the University. One-off affair.
Impact Hub	Incubation and acceleration programme for enterpreneurs	Social impact analysis is part of the programme	Source of income for the IH Incubation Center in Vienna.
Transition Town	Creation of Impact matrix in TT Tooting which was considered useful for professionalizing and for making funding bids.	Impact matrix	The impact matrix was a result of discussions during the "Visions and Pathways" day
DESIS	educational food activities in schools located in the poorest neighbourhoods of the city of Aveiro In Portugal	A didactic game where students (from the 1st to 4th year) follow a storytelling process and keep record of their meals in a diary.	One-time event involving 5 schools. Discontinued because of lack of (public) funding
iMinds	New structure for the Living Lab Unit in the IMinds platform thanks to a positive impacts evaluation and a merger with imec	Impact evaluation (of the living lab by two external organisations)	The impact evaluation took away negative views on the value of Living Labs and facilitated a merger with imec, a nanotech research center
GovLab	Evaluation of the Co-Bologna project by independent researchers	Survey questionnnaire	The evaluation is not completed, but the agreement to be evaluated signalled openness to innovation and to learning to the Munacipality of Bologna
Living Labs	Application for ISO 9001 quality certificate for the management system of ReGIM Lab	Compliance with ISO 9001 which involves (survey-based) bi-yearly client satisfaction analysis	Acquiring the ISO certification provided the Living Lab unit of ReGIM Lab with national and international recognition and allowed it to become part of the European Network of Living Labs
Budget monitoring	Letting go of the human rights elements in workshops about budget monitoring organized by the Centre for Budget Monitoring and Citizen Participation (CBB)	Human rights as an evaluative element was being removed or downplayed	The effect was a watered down version of budget monitoring fitting with the demands of municipalities
Budget Monitoring	Revoking of policymaking and budget authority of districts by the City of Amsterdam.	More centralized budget information	Districts lost budget authority and a district level budget with it