



**Guidance for the design of qualitative case study evaluation**  
**A short report to DG Regio**

**Prof Frank Vanclay**  
**Department of Cultural Geography, University of Groningen**

**February 2012**

**SYNOPSIS**

This short document provides high-level guidance notes on designing and undertaking qualitative case study research and speculates on its potential application in evaluating regional development projects in Europe. It outlines the background of various qualitative evaluation methods used in outcomes assessment (e.g., Most Significant Change Technique, Performance Story Reporting and Collaborative Outcomes Reporting). Some examples of the practical application of story-based approaches and/or narrative evaluation are provided along with a consideration of their effectiveness.

**AUTHOR BIOSKETCH**

Frank Vanclay is professor and Head of the Department of Cultural Geography in the Faculty of Spatial Sciences at the University of Groningen, The Netherlands, where he has been since July 2010. Australian by birth and Dutch by descent, he has a PhD in environmental sociology from Wageningen University, The Netherlands (1994). With formal training in statistics and social research methods from the University of Queensland, Australia, Frank's early career was in quantitative social research and social statistics. Over time and as a result of the type of social research he was undertaking (social impact assessment, social understandings of place, and social aspects of agriculture, farming and natural resource management), he developed a keen interest in qualitative methods and qualitative evaluation. He taught social research methods for many years, and from 1997 to 2000 he was President of the Australian Association for Social Research. He has edited several books and has authored over 65 journals articles and 35 book chapters. His journal articles have received over 1100 citations in Scopus.

**CONTACT DETAILS**

Prof Frank Vanclay  
Department of Cultural Geography,  
Faculty of Spatial Sciences  
University of Groningen  
PO Box 800  
9700AV Groningen  
The Netherlands  
email: [frank.vanclay@rug.nl](mailto:frank.vanclay@rug.nl)

website: <http://www.rug.nl/staff/frank.vanclay/index>

# Guidance for the design of qualitative case study evaluation

## INTRODUCTION: THE NEED FOR QUALITATIVE EVALUATION

The attempt to identify what works and why are perennial questions for evaluators, program and project managers, funding agencies and policy makers. Policies, programs, plans and projects (hereafter all 'programs' for convenience) all start with good intent, often with long term and (over)optimistic goals. An important issue is how to assess the likelihood of success of these programs during their life, often before their goals have been fully achieved. Thus some sense of interim performance is needed, to provide feedback to finetune the program, to determine whether subsequent tranche payments should be made, and also to assist in decision making about whether similar programs should be funded.

Evaluation in such circumstances is always complex. How can the achievement of goals be assessed if they are long term? Evaluation can not wait years to determine whether a program has been successful. Thus evaluation needs to carefully consider the program logic, whether interim steps have been achieved and whether there are signs that longer-term objectives and goals are likely to be achieved. But even this is not straightforward. All programs, especially long-term ones, should incorporate a degree of adaptive management or reflexivity into them allowing them to respond to feedback along the way. Final success therefore is not just whether the original plan was correct, but the extent to which a program has effective monitoring and is capable of adapting to feedback along the way. Depending on the context, it may be too that external factors have changed and the original goals and/or program logic may need to change to accommodate changed circumstances. Any program seeking to contribute to high-level goals like enhanced community wellbeing, social sustainability, regional development potential, innovativeness, etc, is likely to be affected by a changing context. Therefore a key factor for success (and thus for evaluation) is the ability of the program to be responsive to change.

Another issue is that many programs often create a broad range of social benefits that were not necessarily the core purpose of the program, and frequently there can be many other unanticipated spin-off benefits as well. Collectively these may contribute significantly to the perceived success of the program, especially by program beneficiaries. Should evaluation consider the success of a program on the basis of unintended consequences? At face value, many key decision makers might say no. But on the other hand, if a program caused a lot of unanticipated harm in addition to still achieving its narrow goals, would it be regarded as successful? I doubt it! Evaluation must take an holistic approach considering the potential for harm as well as the potential for good, and it needs to consider the unanticipated consequences as much as the intended goals.

A further issue is that an evaluation can not simply measure whether goals (i.e. desired results) were achieved. If so, how would the evaluation establish causality? Could the observed change have been the result of other things occurring at the same time? What if there were underlying trends in a community anyway? In the field of social impact assessment, the concept of baseline is extended beyond being a single datapoint fixed in time. Instead, it is argued that the meaningful comparison is not time  $x+1$  against time  $x$ ; but rather a comparison at a point in time against what would have happened without the program. The baseline is thus the line (not point) of expected trending without the program. In European policy circles, this is called the 'counterfactual'. Thus programs can still be regarded as 'successful' if an indicator at a future time is worse than it was at commencement, providing that there is a reasonable analysis that there were other changes taking place such that the program made the community better off than it would have been without the program.

In considering a wide range of outcomes<sup>1</sup>, and with the realisation that many of the broader social benefits of programs are subjective, the old adage normally attributed to Albert Einstein that "not everything that counts can be counted" becomes important. Particularly in cases of the enhanced wellbeing type of programs referred to above, the additional benefits may be in terms of an improvement in how people feel about where they live and their lives in general, about how they feel about the future of their community, and about how different groups in a community cooperate or at least get on with each other. While not necessarily impossible to measure, these high level goals are difficult to measure, and are not normally included in routinely-collected data collection processes.

The issue of high level broader social benefits raises the question of attribution. How can the evaluator know whether an observed effect was due to the program? A short and simple answer might be that they can't. A more complex answer questions whether simplistic assumptions of direct cause-and-effect are appropriate. Big programs with high level outcomes don't have simple cause-effect relationships, they have complex interconnecting multi-causal linkages. A deeper understanding of the nexus of these relationships is needed. Such systems are dynamic, are mediated by iterative feedback processes, are confounded by inhibiting and enabling mechanisms, and are potentially affected by catalytic relationships (including nonlinear and exponential) between system elements.

It is important to realise that these debates have existed in the field of evaluation for decades (see Greene 2000 for a discussion on the purpose and history of evaluation). While some evaluators have attempted to persist with ever-improving and ever more sophisticated empirical quantitative techniques (Leeuw & Vaessen 2009), many other evaluation experts fundamentally disagree that such methods can address the complexity of the programs being considered (Guijt et al. 2011). Instead, they advocate the use of robust qualitative measures arguing that qualitative methods are more valid, give better information, are more efficient, include the potential for unanticipated factors to be included, and address causality.

---

<sup>1</sup> While the proposed regulations for the reformed cohesion policy published in October 2011 uses the term 'results indicators', the sense in which the word 'results' is used is generally consistent with the use of 'outcomes' in this document. I use the term 'outcome' here because it is the standard terminology in the international field of evaluation. I also note that it was the terminology originally used in the documents tabled at the High Level meetings on the role of indicators and evaluation which took place in early 2011, refer to "Outcome Indicators and Targets: Towards a Performance Oriented EU Cohesion Policy" by Fabrizio Barca and Philip McCann (together with the complementary Notes), DGREGIO Website: [http://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/performance\\_en.htm](http://ec.europa.eu/regional_policy/sources/docgener/evaluation/performance_en.htm)

A final argument in favour of qualitative methods (especially story-based approaches) is that they can yield powerful stories which are not only useful for media reports, but are often frequently preferred by politicians. It is an illusion of scientists that hard data is the only convincing evidence. As Benjamin Disraeli (or at least Mark Twain) implied many years ago with the now famous “lies, damned lies and statistics” aphorism, a statistic (data, evidence) is only as accurate as the reliability of the processes used to collect it and the extent to which it faithfully represents reality (its validity). But reliability and validity (the once all-important cry of quantitative social researchers) have now been replaced with other criteria. With so much data, evidence, information etc everywhere, the key concerns of the users of information are no longer the old ideal concerns of purist statisticians, but the pragmatic considerations of salience, credibility and legitimacy (Cash et al. 2003). Users of information want to know: “*is it relevant information?*”; “*is it useful information?*”; and “*do I believe it?*” – which is partly based on its credibility to them as individuals (in other words, is consistent with their worldview) and partly on the extent to which they trust the source of that information. Very often, a story conveys this information much more effectively (i.e. convincingly) than other forms of evidence (Denning 2007; Fisher 1989; Kurtz 2009; Sandelowski 1991; Shaw et al. 1998).

## A NOTE ON TERMINOLOGY

A project, program, plan or policy has goals and objectives, the achievement of which occurs through an implementation plan and/or one or more activities which were intended to produce outcomes or results. Historically, because of the difficulties in measuring long-term outcomes, much of evaluation effort measured either outputs or even activities and inputs. Bennett (1975) in a now famous paper, ‘Up the hierarchy’, argued for conceiving of a hierarchy of steps in the project sequence and that evaluation should measure change at all levels.

7. End results
6. Practice change
5. KASA change (knowledge, attitudes, skills and aspirations)
4. Reactions
3. People involvement
2. Activities
1. Inputs

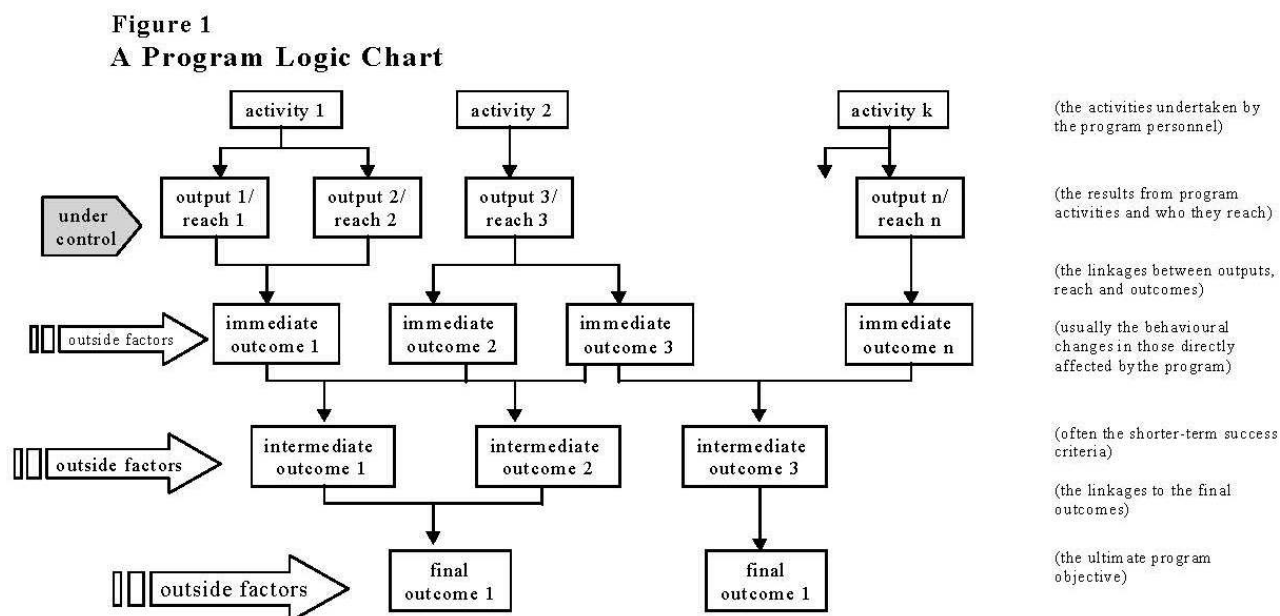
In the decades since Bennett (1975), there has been a considerable move up the hierarchy to such an extent that there is now a much greater focus on outcomes. It is now typical to conceive of projects (and to a lesser extent programs) in the following terms:

inputs → activities → outputs → immediate outcomes → intermediate outcomes → ultimate outcomes

Activities and outputs are often categorised as being ‘primary’ and ‘secondary’ in terms of their centrality to program objectives. The suggestion of Bennett to record at all levels remains relevant in that it is still difficult to evaluate the outcomes of projects and programs that have long lag times (the

time taken for the results to appear). Bennett's conceptualisation of different dimensions of outputs and outcomes also remains relevant.

Program Logic refers to the understanding of how the different components of a program work together to produce outcomes. "It captures the rationale behind a program, probing and outlining the anticipated cause-and-effect relationships between program activities, outputs, intermediate outcomes and longer-term desired outcomes. A program logic is usually represented as a diagram or matrix that shows a series of expected consequences, not just a sequence of events" (Roughley 2009, p.7). Mayne (1999) presents a good model illustrating the complexity of program logic (see Figure 1).



Source: Mayne (1999, p.9)

## QUICK OVERVIEW OF QUALITATIVE METHODS USED IN EVALUATION

Qualitative evaluation uses a wide range of standard social research methods (i.e. ways of collecting data) and an ever-increasing number of innovative qualitative social research methods including:

- Open-ended questions in structured questionnaires
- Semi-structured and in-depth interviews with key informants, undertaken in person, by phone and now increasingly by Skype
- Group interviews (essentially interviews with several people at the same time)
- Focus groups (groups discussions that are actively facilitated to focus on specific topics and where the discussion in the group is an important part of the process)
- Workshops, often with table-based or group exercises, including variants such as World Café (Hartz-Karp & Pope 2011)
- Role plays and games

- Expert panels, citizen's juries (Smith & Wales 1999) and other deliberative methods (Hartz-Karp & Pope 2011)
- Document analysis (i.e. analysing all available documents, photos, letters, emails, and other outputs of the activities associated with a project or program)
- Go-along interviews (Carpiano 2009), in situ interviewing, shadowing (where the researcher/interviewer follows the participant as they go about their normal daily business)
- Story-telling with stories either written down, audio-recorded, or video-recorded
- Photo-elicitation (photovoice) and other visual 'triggering' techniques to stimulate participants to recall and articulate their thoughts and/or stories about certain topics (see Wang & Burris 1997)
- Mental models, mind maps, and mud maps (i.e. a representational diagram showing the interconnections between related concept) – sometimes called spidergrams
- Participant diaries, logbooks and audio or visual recordings of reflections/comments as soon after they happen as practical
- Observation (sometimes aided by video recording)
- Participant observation and other experiential techniques
- Researcher diarising (systematically recording notes in a field notebook).

There is a wide range of methodologies (ways of organising data) and frameworks for thinking about information. There are also numerous theoretical frameworks (ways of interpreting data). In general, however, irrespective of the methodology or theoretical framework applied, the above techniques for collecting data tend to be utilised. In all social research, but especially in qualitative methods, partly as a quality-control mechanism, the use of multiple methods is common (also called 'mixed methods' especially when in conjunction with quantitative methods), and is methodologically preferred on the basis of a concept called 'triangulation' – that different methods should be used, with different sources of data, and from different perspectives.

It should be noted that there are many approaches that combine theoretical understandings, methodologies and specific techniques into holistic and coherent frameworks that are suitable for particular situations. Many of the tools/methods used in these approaches can be very creative, which makes the construction of a comprehensive list of methods quite difficult. Some examples of these approaches that have an evaluative dimension include Rapid Rural Appraisal and Participatory Action Research (Chambers 1994), Appreciative Inquiry (Cooperrider et al. 2008) and Outcomes Mapping (Earl et al. 2001).

Sample size is a quantitative concept. In qualitative research there is concern about the veracity of statements made, and there is a much discussion about the robustness of analysis and interpretation. However the justification of the robustness of the research is not drawn from statistical notions of significance drawn from probability calculations. Typically, in qualitative research small numbers of people are included with interviews being conducted until 'saturation' is reached – that is, when no new information arises. People are typically selected purposively (deliberately) for their particular characteristics rather than randomly.

The approach taken in qualitative research is markedly different than in quantitative research. In general, qualitative research (dot points below drawn from multiple sources):

- Focuses on meaning, and on ‘the why’
- Tends to focus on new issues where understanding is required, rather than on confirming prior hypotheses
- Seeks to discover the ‘insider’ rather than ‘outsider’ perspective
- Is person-centered rather than variable-centered; often taking a humanistic or phenomenological perspective
- Often seeks to promote joint learning by all participants, rather than just the learning of the researcher
- Is holistic rather than particularistic
- Is contextual (situated, embedded) rather than decontextual (distant, removed or detached)
- Seeks to understand depth rather than breadth.

In evaluation, qualitative research is used specifically to consider the why questions that quantitative methods typically can not answer:

- Why does the program work (or not work)?
- How does the program achieve its goals?
- Why does it work for some clients (or in some situations) and not others?
- What are/were the needs of the clients that were not anticipated by program developers?
- What were the additional unintended and/or unexpected positive or negative consequences?

For further information about qualitative social research methods, please refer to the key references provided following the list of references to this report.

## **BACKGROUND: A SHORT HISTORY OF QUALITATIVE EVALUATION**

Guba and Lincoln (1989) describe four generations in the history of evaluation: (1) measurement, (2) description, (3) judgement, and (4) constructivist evaluation. Like all such summaries, the primary criticism of Guba and Lincoln is that they over-exaggerate the extent to which the phases actually existed as discrete time periods – as can be amply demonstrated in that the earlier generations are still very much in existence. Nevertheless, the dominant culture of evaluation is changing. While the division between summative evaluation (asking whether a project or program was a success) and formative evaluation (how can the project or program be improved) is widely accepted, evaluation is developing a greater focus as being creative (Patton 1981), qualitative (Patton 1990), participatory (Jackson & Kassam 1998; Whitmore 1998), utilization-focussed (Patton 1997), constructivist or fourth generation (Guba and Lincoln, 1989), empowering or fifth generation (Fetterman et al. 1995; Laughlin & Broadbent 1996; Fetterman 2000), and as a form of action research (Whyte 1990). Evaluation is now viewed as research for informing decision-making (Vanclay et al. 2006). Instead of just being ex-post assessment or audit, evaluation is now understood to contribute to all stages of project or

program development. Rather than being solely the domain of independent experts, evaluation is now widely seen as a participatory approach that empowers and builds capacity within institutions and amongst all program and project partners. Evaluation is now seen as a form of action research that informs project and program design (Vanclay et al. 2006). Ongoing evaluation and adaptive management is an essential part of being innovative and a learning organisation.

Monitoring and evaluating the performance of rural development programs in developing countries has been an area where much innovation has taken place in evaluation methods, largely because of the inappropriateness of many traditional quantitative means of evaluation and because of the strong interest by funders in knowing whether their funding was being used effectively. Many external evaluators and program managers have grappled with how to design fair methods that adequately captured the changes brought about by development interventions, especially when empirical indicators were not available, were too broad or not sufficiently sensitive to change.

One evaluation approach to emerge from the rural development field is the Most Significant Change Technique (MSC) originally developed by Rick Davies (Dart & Davies 2003; Davies & Dart 2005). There were several story-based approaches to evaluation in the 1990s, but two of the main proponents are Dr Rick Davies (an Australian based in the UK) and Dr Jessica Dart (a UK person based in Australia). Each of these developed their thinking both as a practitioner and through doing PhD research on evaluation – Davies completing in 1998 at the Centre for Development Studies at the University of Wales, Swansea; and Dart in 2000 at the University of Melbourne. Although it had previously been called a variety of names, in 2000 Davies and Dart settled on 'Most Significant Change Technique' as the term for the emergent method. MSC has now been widely used in over 10 countries. In 2005, to assist people in utilising MSC, Davies and Dart produced a 100-odd page MSC User Guide, which is readily available for free from their respective websites and from some donor websites as well. Some evaluations using this approach include Wilder & Walpole (2008) and Waters et al. (2011).

Performance Story Reporting (PSR) is similar to the Most Significant Change Technique. Dart attributes the actual name 'performance story' to John Mayne of the Canadian Auditor General's Office (Mayne 2004; Dart & Mayne 2005). Arguably the name 'Performance Story Reporting' overcomes many of the criticisms of 'Most Significant Change Technique' (Willets & Crawford 2007) and in any case its different procedures means that it provides a more sophisticated, yet still workable tool. While not yet widely represented in the scholarly literature, it is well known in the evaluation field (see Dart & Mayne 2005).

Collaborative Outcomes Reporting (COR) is the latest incarnation of these story-based approaches to evaluation. Similar to PSR (in fact Dart provides her PSR work as examples of COR), the technique was renamed because of resistance to the word 'story' by some evaluation clients (Dart pers. com. 20 April 2011) and because it emphasises integrating empirical and qualitative data. There are a range of refinements, but these are minor, and therefore information about MSC and PSR is essentially still relevant to COR. An outline of COR is presented in Appendix 1.



## DESIGNING AND CONDUCTING A STORY-BASED APPROACH TO QUALITATIVE EVALUATION

The intention with story-based evaluation approaches is to be a rigorous qualitative method of reporting the impacts of projects and programs through stories. Intended to be used in conjunction with the reporting of specific empirical indicators where they are available, story-based approaches are especially useful to capture the broader social benefits of programs, particularly in situations where empirical attribution may be difficult. Mayne (2004, pp.49-50) considers that:

There are a variety of ways to present a performance story. All involve a mix of quantitative evidence that certain outputs and outcomes have occurred as well as narrative discussion and further evidence of the contributions made at various points along the results chain, all described within some context. A performance story sets out to convince a skeptical reader that the activities undertaken by the program have indeed made a difference — that the expectations chain has, at least to some extent, been realized, along with any significant unintended results.

Mayne (2004, p.50) outlines the elements of a performance story as follows:

### **What is the context?**

- the overall setting of the program (description, objectives, resources)
- the results chain (program theory)
- the risks faced

### **What was expected to be accomplished at what cost?**

- statement of the (clear and concrete) outputs and outcomes expected
- planned spending

### **What was accomplished in light of these expectations?**

- the relevant outputs delivered at what cost
- the outcomes realized related to the expectations
- a discussion of the evidence available demonstrating the contribution made by the program to those outcomes

### **What was learned and what will be done next?**

- a discussion of what will be done differently as a result of what was achieved

### **What was done to assure quality data?**

- a description of what the organization does to ensure the quality of the data and information reported

The main story line of a performance story is how well the program has performed in relation to what was expected and what will now be done differently to better ensure future performance.

The big difference between the approach by John Mayne and that taken by Jess Dart and Rick Davies is where the stories come from and how they are constructed. For Mayne (at the time in the Canadian Auditor General's Office), a performance story was an effective way by which the evaluator provided information to a client, or how the Auditor-General might report to Parliament. The story technique was an effective way of expressing complex information in a more credible and meaningful way to the target audience.

In contrast, MSC, PSR and COR all utilise the stories of participants themselves (i.e. the program or project beneficiaries) in the report. These techniques have a procedure for collecting stories and for selecting stories that provide a good example of the success of the project or program.

## A REAL APPLICATION OF PERFORMANCE STORY REPORTING AT THE PROGRAM LEVEL

The Australian Government funds a range of natural resource management (NRM) programs and projects collectively known as the 'Caring for Our Country' Program (see <http://www.nrm.gov.au/>). For the five year period from 2008 to 2013, the total value of the Government's investment in this program will be about AUD\$2.25 billion (Commonwealth of Australia 2008). Investments in the two previous five year periods totaled \$1.5 billion and \$1.3 billion respectively (Auditor General 2008). The program supports a range of disparate projects often in conjunction with local community groups and may include small grant programs, co-funding programs, and support for project staff and project costs. While the objectives are now very clear (Commonwealth of Australia 2008) and there is a clearly elaborated program logic (Roughley 2009), earlier versions of the program did not have clearly identified intended outcomes. While anecdotal evidence suggested that there was much benefit from the program (Curtis & Lockwood 2000; Prager & Vanclay 2010), various official reports were dubious about the benefits. The Auditor General's (2008, p.16) report, for example, concluded that there were "significant areas of noncompliance by State agencies", and that "the quality and measurability of the targets in the regional plans is an issue for attention and ... should be considered nationally – especially as the absence of sufficient scientific data has limited the ability of regional bodies to link the targets in their plans to program outcomes". The report went on:

14. There is evidence that activities are occurring 'on the ground'. For example, Environment's 2006–07 *Annual Report* commented that the programs have 'helped to protect over eight million hectares of wetlands, have treated over 600 000 hectares of land to reduce salinity and erosion, and have involved some 800 000 volunteers in on ground conservation work'. However at the present time it is not possible to report meaningfully on the extent to which these outputs contribute to the outcomes sought by government. There are long lead times for national outcomes and delays in signing bilateral agreements did not help this process. The absence of consistently validated data, the lack of agreement on performance indicators and any intermediate outcomes has significantly limited the quality of the reporting process.

15. Overall, the ANAO [Australian National Audit Office] considers the information reported in the DAFF [Department of Agriculture, Forestry and Fisheries] and NHT [Natural Heritage Trust] Annual Reports has been insufficient to make an informed judgement as to the progress of the programs towards either outcomes or intermediate outcomes. There is little evidence as yet that the programs are adequately achieving the anticipated national outcomes or giving sufficient attention to the 'radically altered and degraded Australian landscape' highlighted in the 1996 *Australia: State of the Environment* Report. Performance measurement has been an ongoing issue covered by three previous ANAO audits since 1996–97 and should be a priority for attention in the lead up to NHT3. (Auditor General 2008, p.16)

This context of official concern about the alleged benefits of the program but strong public and political support for the program led to a real need to prove that the program was being successful, especially in the knowledge that empirical indicators were unlikely to reveal results in the short term.

Jessica Dart had been experimenting with the most significant change technique and performance story reporting for some time, first in the early to mid 1990s in developing country contexts and later in an agricultural extension context in Australia. After working with the Department of Primary Industries in Victoria (Australia) for three years (and completing a PhD at Melbourne University in 2000), she established a consulting company, Clear Horizon. Jess's work was known by various people in government, especially in natural resource management circles, and it became evident that performance story reporting might be a good way to assist the Australian Government in its need to capture the impacts of its investments in natural resource management (NRM).

The Australian Government's Bureau of Rural Sciences (now part of the Australian Bureau of Agricultural and Resource Economics and Sciences, <http://www.abares.gov.au>) conducted a feasibility of the use of story-based approaches by conducting four pilots (Carr & Woodhams 2008). That "independent review" considered three questions:

- 1) Are qualitative approaches such as MSC a) useful and b) appropriate as evidence of outcomes, including intermediate and other outcomes?
- 2) What are the strengths and weaknesses of using PSR to report by outcomes?
- 3) Could MSC and PSR be used by NRM regions in Australia as a form of participatory evaluation for producing program performance reports by outcomes (Carr & Woodhams 2008, p.3)

The report concluded that:

Qualitative approaches to participatory evaluation such as MSC are both useful and appropriate as evidence of outcomes at multiple levels in NRM program logic hierarchies, including intermediate outcomes. Not only are qualitative approaches a valuable source of evidence of the changing human dimension of NRM, they are frequently a profound source of insight and sometimes the only kind of evidence available of the type of practice and attitudinal changes taking place.

...

Four key strengths of PSR were mentioned in reflective interviews: engagement, capacity building, problem-sharing and adaptive learning. These strengths were primarily associated with the MSC phase of the PSR process. Identifying and engaging evaluation stakeholders was seen as a major strength of PSR and was the strength most frequently mentioned by interviewees. Many of the regional staff who took part in the MSC process appreciated the chance to build relationships with resource managers and develop their personal interviewing skills. The MSC process also increased communication about shared experiences and approaches to NRM problems that, in turn, led to an adaptive approach to natural resource management.

Arguably, there are two other key strengths of PSR. First, it integrates qualitative and quantitative evidence. Second, performance story reports rely upon participatory processes using program logic, which allows progressive collection and testing of evidence throughout the life of the investment program.

There are three key challenges for PSR: time and resources; data, results and interpretation; and complexity and preparedness. The biggest challenge across all stages of PSR was a perceived lack of time and resources to conduct the PSR process. Such comments came from all interviewees, consultants, regions and Australian Government representatives. At the regional level, interviewees were concerned that the goodwill and involvement from resource managers and regional staff would start to wane if the process was repeated each year without sufficient resources or local incentive. (Carr & Woodhams 2008, p.61)

As a result of the positive feasibility assessment, the performance story reporting approach was rolled out across Australia with the Australian Government publishing a 'User Guide', *Developing a Performance Story Report* (Roughley & Dart 2009). It is likely that several project-level performance story reports have been prepared, however curiously few of them have been made available on the internet. One example that is available is the performance story report for a biodiversity program in the Mount Lofty Ranges and Fleurieu Peninsula region of South Australia (Dart & O'Connor 2008). Figure 2 is a copy of a page from that report showing how the stories are presented.

The project-level performance story reports can be aggregated into higher level evaluations. One example is the assessment of natural resource management outcomes in the State of South Australia for the period 2001 to 2008, which drew on several performance story reports that were conducted in that state (see DWLBC 2009). The performance story reports can be presented in a variety of formats, as conventional consultancy reports, as websites, as DVDs, as posters or brochures (see Figure 3), or as mixed media. Figure 4 shows a magazine/newsletter representation of a report with links to sound recordings of the participants' stories.

## Section 4: Instances of significant change

The following vignettes were chosen by the participants of the summit workshop as representing the most significant changes occurring as a result of investment in the Recovery Program. These vignettes were chosen out of three sets of vignettes (22 in total).

### What is a vignette?

Vignettes are used to elicit responses, interpretations and judgments about a particular set of circumstances or context within a research setting. Typically used in the qualitative social sciences, vignettes offer a method for simulating complex events, outcomes and/or problems and use these to explore people's perceptions, opinions, beliefs and attitudes. In this case, the vignettes were extracted directly from interview transcripts, which were captured with digital audio recording. While edits were kept to a minimum, some text was removed, this is indicated by three dots between sentences.

### Vignette #1: Changing mind set

*When we came here 10 years ago and found out how valuable the swamps are, and someone said you could be milking cows off it. ...Part of the reason that this property was bought was so that we could use the swamp for the cattle to be in, and now it has completely changed around to where you are not allowed to let, where we won't be letting any cattle in at all. ...I remember a long time ago, someone saying that this may be coming, and at the time I was thinking, oh well, we won't be doing that because we need our swamps for our cows, we can't shut them off. But that was our mind set...*

*And I don't think we contemplated shutting off the swamps when we were fully milking cows because that area was just too important. ... I think the fact that we have stopped milking cows makes it a whole lot easier too, because the milking cows needed to have every amount of grass...*

*Now I am ready to fence off swamp, but I have to do it slowly because Dad doesn't see all this land getting fenced off. ... [the project officer] came out and we got a bit of a line where we were going to run the fence. When I told Dad about it, he wanted to shift the line by quite a few metres further into the swamp so that he didn't lose so much land, but I think we have come to an agreement out of that, of how to do it, and what we should be doing. And usually with my father, he grizzles about things at the start, but once you have done it he has friends coming around and he is out showing them how great it is.*

*I think it is changing, I think there's more people that are into swamps... Maybe the more people do it the more acceptable it will be for other people to do it...It is a bit of an issue but it is more acceptable for us, I guess, and for others, I guess the more you hear about it the more you learn, the easier it is.*

*I guess I know now by seeing the map, the Emu-wrens are probably only a few kilometres away, and if we can certainly fence off our swamp and revegetate it, there is a chance that the Emu-wrens would move onto our property. And think the visual effect of seeing the swamps revegetated will certainly make the property look better, probably not quite as bare as it is. Everything that you hear around the place all the time, is how much everything is changing in our environment. Maybe this is just our way of trying to do our bit to help. It could be nicer for the kids, if the farm stayed in the family, they might see the rewards out of what we have done, maybe the Emu-wren will still be around then.*

*This vignette was considered significant by the participants at the summit workshop for the following reasons:*

- *It provides a real farmer perspective.*
- *It demonstrates how increased environmental knowledge and awareness can lead to attitudinal change and then to behavioral change.*
- *Promotion of change of practice needs to be continuous but also opportunistic. In this case the change in land use created an opportunity to encourage habitat protection (fencing).*

24

Figure 2: Example of the presentation of a performance story

(Source: Dart & O'Connor 2008, p.24)

### WHY ARE THE OUTCOMES THAT RIBBONS OF BLUE CONTRIBUTES TO IMPORTANT?


"I think that these kinds of outcomes are important if we are to produce environmentally aware citizens of the future who can push for and support environmentally friendly policies both in private enterprise and government institutions." Regional primary teacher

"A lot of the problems they hear about, like global warming and stuff, they can't really make a difference, and I try and impress on them that they don't have to solve all the world's problems, they can look after their little patch of it, that's a good start, so I think that's quite important too, that small changes can make a difference." Metropolitan secondary teacher

"It formulates who they are and where they are going and what they want their world to be like." Metropolitan primary teacher

"My classes feel ownership and have accepted responsibility for the Meekup Regional Park and have an ongoing love of the natural environment." Regional primary teacher

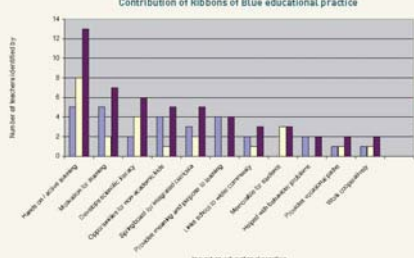
"If they don't take care of it then no-one else is going to." Metropolitan primary teacher



### AN EVALUATION OF Ribbons of Blue: 2008/2009

How Ribbons of Blue has contributed to 'pro-environmental' values, understanding and behaviour in school students; an evaluation of outcomes.

#### Contribution of Ribbons of Blue educational practice



Impact on educational practice

"I believe the hands on learning, especially out in the field, is a far more effective way of learning. Seeing the excitement of kids looking at 'bugs' with binocular microscopes, and catching their own macros, is very rewarding for me as a teacher." Regional primary teacher

"Developing understanding and care for the world doesn't just come from a book - you need sight, smell, touch." Metropolitan primary teacher

"They need to know that it (science) is a real activity, its not just a subject you study at school, it has an impact on all sorts of aspects of their life. And if they don't get that understanding of science then we've failed fairly spectacularly as teachers." Metropolitan secondary teacher

#### Ribbons of Blue

- is an environmental education program which aims to motivate and enable current and future generations to respect and take action for healthy WA waterways
- has been going for 20 years and during 2008 and 2009 was delivered in nine regional catchment areas by 9.2 FTE's
- received \$1 009,630 in funding in 2008/2009 from South West Catchments Council, the Department of Environment and Conservation, Education, Water, Swan River Trust, Avon Catchments Council, WA Plantation Resources and Shire of Manjimup.

This evaluation:

- was conducted by program staff, using the Performance Story Reporting process, with assistance from a consultant
- collected data from a survey of teachers (63 responses); three focus groups with students; 22 semi-structured interviews with teachers and coordinators' records, on 14 Key Evaluation Questions
- had an 'Expert Panel' review evidence and make outcome statements
- ran a 'Summit Workshop' of 22 stakeholders to discuss findings; agree on key outcomes and make recommendations.

Some limitations of this evaluation:

- Data was collected mainly from teachers (other key stakeholders not included).
- Teachers may want to 'protect' the program.
- Inexperience of program staff in conducting evaluation and interviews.

Ribbons of Blue - Our catchment, our future

## Ribbons of Blue PERFORMANCE STORY REPORT

Ribbons of Blue's delivery in each region is tailored to local issues, conditions and needs. In 2008/2009 the program delivered:

- over 1000 'occurrences' with schools
- 67 professional learning workshops attended by 466 teachers
- 700 Teacher Resource Kits
- on-ground works: 15.25 hectares native vegetation enhanced in South West




Photo competition: Hand touching water

### KEY OUTCOMES - 12 outcome statements were made by the Expert Panel based on the evidence presented; the Summit Workshop selected those they believed most important, they were:

- Ribbons of Blue has definitely been successful in enabling students to develop values related to environmental responsibility.
- Ribbons of Blue has definitely been successful in enabling students to develop their knowledge of local catchments and waterways.
- There is evidence that Ribbons of Blue is successful in enabling students to develop their knowledge of how people impact on their local catchment.
- There is some evidence to suggest that Ribbons of Blue is successful in encouraging students to adopt behaviour/practices to improve the environment.
- The majority of teachers who responded to the survey are using Ribbons of Blue as an important part of their teaching and learning program.

### PARTICIPATION:

11,250 students in 2008, and 10,000 in 2009, directly participated in Ribbons of Blue activities. The breakdown by region is:

ROB region	Number schools each year	Number classes each year	Number students involved in Ribbons of Blue as a percentage of the total number of students in each region of Blue (2008)
Avon	14	32	95%
Blackwood	14	32	95%
Classy Deer	9	18	100%
Donkeyhide	14	32	95%
Eastonia	27	52	64%
Leathemup	20	37	95%
Mitri Well	16	32	95%
Swan	100	190	98%
Thames	9	18	95%
Water	10	21	9%
Western	210	347	60%
Yalumba			

Note - Coordinators ceased operation in September 2009 due to loss of funding

### TEACHERS' COMMENTS:

excerpts from the three 'stories of significant outcomes' chosen by Summit Workshop as best representing how Ribbons of Blue contributed to significant outcomes for students. Twenty two 'stories' were considered by the workshop.

"I think the most important outcome for the students is the result, the outcomes of the 'macro' snapshot and looking at that data and analysing it and tying it in to what we do on the river... So it's using that data and the process to reinforce that what we're doing, when we get our hands dirty (vegetation), is the right thing and is helping to make a difference."

Metropolitan primary teacher

"Ribbons of Blue has contributed to the development of our students as it offers meaningful, authentic learning experiences; it engages boys; enables community involvement; encourages the use of common language; and is very positive in forming attitudes (e.g. appreciation of the natural environment) in families and communities. Many of our families have not been into the regional park!"

Regional primary teacher

"Over the years the biggest change I've seen is in attitude. Attitude, attitude, attitude. It's changed everything - their caring, understanding, appreciating the importance of the environment. It's not just academic; they truly 'get' why we don't litter, that we need to care for our patch, that it's our responsibility. ... For example, we had a student who heard about the planned waste treatment plant next to The Spectacles, and because she knew and cared for the site she complained because of the possible effect on water quality!"

Regional primary teacher

### STUDENT VOICES

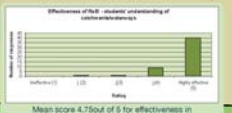
Do you feel differently about the river?

"Yes - way different, in a good way because you never really pay attention to the river until you learn about all this stuff."

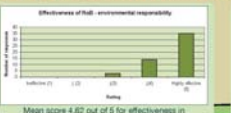
"Like, you can look at a river and think it is just water and trees and stuff but now we look at the water and know that all these things are happening." Regional primary focus group 2

"We learnt about the water quality and how the animals react to it. We can study it later on if the creatures are dying or something and we can find out what's wrong and help fix it." Secondary focus group

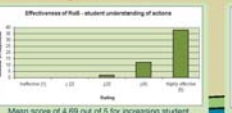
### DATA FROM TEACHER SURVEY: perception of the effectiveness of Ribbons of Blue in contributing to outcomes



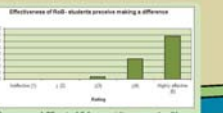
Mean score 4.75 out of 5 for effectiveness in enabling students to develop an understanding of local catchments.



Mean score 4.82 out of 5 for effectiveness in enabling students to develop values related to environmental responsibility.



Mean score of 4.69 out of 5 for increasing student understanding of how their actions impact on the environment.



Mean score 4.82 out of 5 for providing opportunities for students to feel they can make a difference.

Some things students said they'd do differently:

- More careful when we go to the beach-always stick to tracks.
- Don't allow the dog to run up in to the sand dunes anymore.
- Pick up litter instead of leaving it there, seeing litter spoils the enjoyment of the beach now, felt yucky and annoyed about the litter and surprised how much and what sort of litter.
- Water conservation at home-shorter showers.
- Picking up dog poo because of the phosphates and potential algal blooms.

Regional primary focus group 1

Figure 3: Example of the creative presentation of a performance story as a poster/brochure (Source: [http://www.ribbonsofblue.wa.gov.au/component/option,com\\_docman/task,doc\\_download/gid,145/](http://www.ribbonsofblue.wa.gov.au/component/option,com_docman/task,doc_download/gid,145/))

Appendix IV Innovative Communication  
NAILSMA Newsletter, October 2008

Please use this link to access this Newsletter online: <http://www.nailsma.org.au/newsletters>

NAILSMA Dugong and Marine Turtle Project News

Having trouble reading this newsletter? Click here to see it in your browser.



- October 2008 In This Issue...
Dugong and Marine Turtle Management Summit: Huge Success!
Performative Story: Bears Completion

Dugong and Marine Turtle Management Summit: Huge Success!



"Thumbs up!" say the participants of a recent Dugong and Marine Management Summit held by NAILSMA and hosted by the Kimberley Land Council and Barji Jaw Rangers at Mudrum on the Dampier Peninsula...

The Summit comes at the tail-end of the first phase of the Dugong and Marine Turtle Project and was an opportunity for Rangers and Communities involved in the Project to look back and share their successes over the past three years...

Most significantly, all Project participants gave their overwhelming support to continue the Dugong and Marine Turtle Project with a focus on further increasing the capacity of Ranger Programs so that they can take their regional activity plans to the next level.

Also agreed to by the participants was an expansion of the

The Contribution of Investment in the Dugong & Marine Turtle Project Towards Community Capacity, Biodiversity and Other Outcomes Performance Story Report, Appendix IV Page 1

- Terrence Taylor
Terry Mountain
Zoea Muzzas
Vanessa Walsh
Veron Varner

NAILSMA Dugong and Marine Turtle Project

Email: [manager@nailsma.org.au](mailto:manager@nailsma.org.au)  
Web: [www.nailsma.org.au](http://www.nailsma.org.au)



Project Partners and Participants

- Kimberley Land Council
Barji Jaw Rangers
Northern Land Council
Purnululu Rangers
Karrakatta Rangers
Walesley Islands Rangers
Carpentaria Land Council
Aboriginal Corporation
Walesley Islands Rangers
Bathurst Regional Council
Murrumbidgee Regional Council
James Strait Regional Authority
Mar (Murray) Islandery and Creek
Gannawarr Rangers
Radium Camp Park
Barku Rangers
Mabo Rangers
Mabo Rangers

project to include more Indigenous Ranger units from across the North into the Project. The participants showed enthusiasm for the network to develop and see the opportunities for collaboration and sharing ideas across the regions as an important aspect of the Project.

A strong show of support for the Dugong and Marine Turtle Project by state and federal government was evident by the big mob of officials who participate in the four-day Summit. Heading the Department of the Environment, Water, Heritage and the Arts representation was Working On Country Assistant Secretary Kathleen Mackie who was accompanied by colleagues Donna Kwan, Ike Kessling, Jennifer Rahmy and Les Russell.

Listen to what the participants had to say about the Summit by clicking the pictures below.

Listen to what the participants had to say...

Click the picture to be taken to the NAILSMA website where you can listen to audio from the participants recorded at the Summit.

The Summit ran from October 6 - 10



The Contribution of Investment in the Dugong & Marine Turtle Project Towards Community Capacity, Biodiversity and Other Outcomes Performance Story Report, Appendix IV Page 2

- Islanders from Guda
Murrumbidgee
Karrakatta
More from the Project
Project Website
Progress Newsletter
Contact Us
More from NAILSMA
About NAILSMA
NAILSMA News
NAILSMA Events
NAILSMA Publications
NAILSMA Website
Contact NAILSMA
Send to a Friend
Subscribe
Unsubscribe
Vote1
NAILSMA Dugong and Marine Turtle Project
Voting Closes Midnight 22 October!



The Contribution of Investment in the Dugong & Marine Turtle Project Towards Community Capacity, Biodiversity and Other Outcomes Performance Story Report, Appendix IV Page 1

- Coastcare Award\* and a "People's Choice Award"
The project is currently in the top 5 and with only days remaining until voting closes (midnight 22 October), this is your last chance to show your support for the invaluable work of Indigenous Rangers across the north.
To vote, visit the Landcare Heroes Website.
When you place your vote, you can also leave a message of support for the Indigenous Rangers and communities involved in the project.
Please pass this message on to your networks and friends.
Call for Local Heroes!
This could be a great opportunity for ranger groups to give recognition to people involved in their program that have inspired and excited. It would also be an opportunity to profile your ranger program in a national forum.
Applying is easy and NAILSMA will support you to complete the application.
For more information or to apply, contact Josh at NAILSMA.



The Contribution of Investment in the Dugong & Marine Turtle Project Towards Community Capacity, Biodiversity and Other Outcomes Performance Story Report, Appendix IV Page 4

Figure 4: Example of the creative presentation of a performance story as mixed media Note: When originally created and viewed with an internet connection, each speech bubble contained a link to an audio file which played the story of the person in the picture. (Source: <http://www.savanna.org.au/nailsma/publications/downloads/Dugong-and-Marine-Turtle-Project-PSR.pdf>)

## HOW TO UNDERTAKE A PERFORMANCE STORY REPORT EVALUATION

The Australian Government's User Guide (Roughley & Dart 2009) and various other instruction manuals (e.g. Dart et al. 2000; Davies & Dart 2005; Silver et al. 2009) provide ample advice on how to actually conduct evaluations using performance story reporting.

A performance story report provides (Roughley & Dart 2009, p.7):

- a view of progress towards outcomes at a glance
- insight into what's working and what's not and why
- a succinct account of program achievements
- an understanding of the links between investment and intended results, at intervals throughout a program
- a way for organisations to answer some of their more strategic evaluation questions on an as-needs basis or to use as a 'kick-start' process to revise the monitoring system
- information for future investment strategies
- a structure for an organisation's annual report.

Roughley & Dart (2009, p.12) suggest that a typical report is between 10 and 30 pages long and comprises five parts:

**Program context** — background information about the program and the context in which it operates (how the program began, its location, objectives and key strategies, funding sources, structure and expected achievements), as well as an outline of the objectives and boundaries of the performance story evaluation and a summary of key outcomes and what has been learned.

**Evaluation methodology** — a brief overview of the process used in undertaking the evaluation.

**Results** — a listing of the most relevant and rigorous sources of evidence against the outcomes from the program logic hierarchy. This includes data as well as stories of change which are excerpts from interviews that best illustrate change that has occurred as a result of the program.

**Findings and implications** — a discussion framed by the evaluation questions that covers how the study has illustrated the program's impact (intended and unintended outcomes), the progress it has made towards its expected outcomes and how it has contributed to the long-term outcomes of NRM or a large NRM initiative. This part also includes recommendations for applying the findings to future phases of the program.

**Index** — a list of all the sources of evidence considered in the evaluation, including any additional references and the categories of interviewees and study participants.

Roughley & Dart (2009, p.15, slightly modified) describe the seven steps to produce a report:

**Scoping** — inception/planning meetings are held to determine what will be evaluated, develop the program logic (if not already existing), set evaluation questions, and identify existing evidence and people to be interviewed.

**Evidence gathering** — an evidence trawl is conducted to identify existing data that will provide best evidence for expected outcomes. This is followed by the social inquiry process, where interviews are conducted with people who can provide additional information about program outcomes. Specific questions are asked and recorded to provide stories of significant changes that have occurred as a result of the program.

**Integrated data analysis** — quantitative and qualitative data is analysed to identify evidence corresponding to the outcomes in the program logic and integrated within the results chart.

**Expert panel** — people with relevant expertise assess the evidence of outcomes that has been gathered. They judge and make statements about the extent to which the evidence is adequate to assess the progress the program is making towards its stated outcomes. The panel may also identify further evidence that may be needed to make a conclusive statement about the achievement of

program outcomes. Following the panel meeting, the evaluator integrates all of the analysed evidence and assesses the amount and quality of evidence available for each outcome in the program logic to inform a draft set of recommendations.

**Summit meeting** — evaluation participants come together to consider and discuss the findings, nominate the stories that best illustrate the impact of the program and make recommendations for the program in future.

**Integration, report and communications** — the evaluator prepares the performance story report, which is a synthesis of all the above steps including recommendations from summit meeting participants. A plan is established to communicate the findings of the evaluation.

**Revising the program logic** — program managers, staff and other stakeholders meet to consider the report and revise the program logic as needed to plan for the next phase of the program. The next phase can incorporate the lessons and recommendations from the previous phase.

Figure 5 shows how the seven steps link to the different parts of the report. The User Guide (Roughley & Dart 2009) outlines each of the seven steps in some detail.

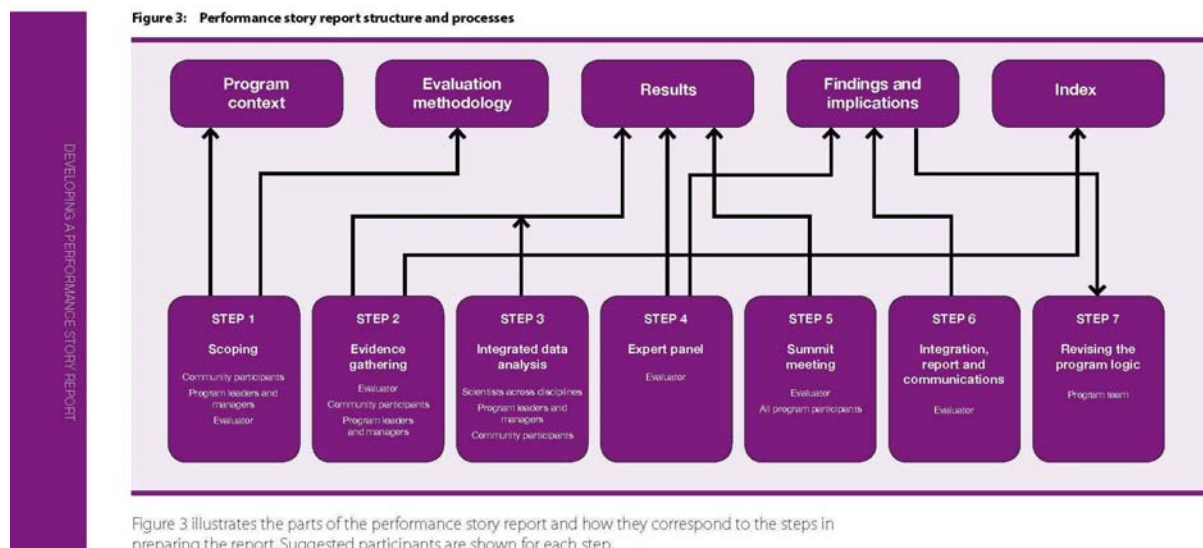


Figure 5: the steps in producing a performance story report  
(Source: Roughley & Dart 2009, p.13)

There are two critical steps in the process. The first is the process of recording stories (Step 2) and preparing them for presentation; and the second is the process of selecting stories to be included in the report (Step 4). A wide range of people should be interviewed. In addition to project participants and people involved at the project at all levels, strategic informants who can comment at a strategic level about how the program has been experienced or how it has addressed policy goals, such as funding agency representatives and policy officers from local, state or federal government can also be interviewed. The User Guide provides guidelines for how the stories should be prepared.

The expert panel comprises a group of people charged with having oversight for ensuring that the body of evidence collected, including the stories of change, are credible and realistic examples of the changes that have been made.



## A PERSONAL ASSESSMENT OF THE EFFECTIVENESS OF STORY-BASED EVALUATION

From January 2002 to June 2010 I was professor of rural sociology and Leader of the Rural Social Research Group in the Tasmanian Institute for Agricultural Research at the University of Tasmania, Australia. In June 2008, I was involved in the development of one of the performance story reporting case studies in the Australian Government's assessment of the Caring for our Country Program (see earlier in this report). My participation in that process means that I have personal experience by which to make informed comment (albeit from a limited personal perspective).

I had been familiar with Jessica Dart's work while she was in the Victorian Department of Primary Industries. While I previously had misgivings about the 'Most Significant Change Technique' she was advocating at that time, primarily because its naming gives the wrong impression, I was much more satisfied with the naming of 'Performance Story Reporting' and with the process it uses. My overall impression, as is evident in this report, is that the method (in generality) has real potential to be useful, particularly in complex situations with multi-causal processes, long lag times, and a wide range of potential outcomes.

Taking a more critical perspective and being more specific, my personal assessment is that there is potential for considerable variation in the way the stories are compiled especially when multiple interviewers are used. It is necessary to ensure good training, supervision and monitoring of the interviewers so that they are relatively consistent in the way stories are collected. There is potential confusion and/or overlap between the role of the expert panel and the summit meeting. In the particular case in which I was involved, the independent consultant seemed to have made some departures from the stated protocol. The expert panel and the summit need to know what their respective roles are, and what the criteria and/or instructions are for what they are supposed to be doing. If there is to be a panel of experts, they must not be fundamentally opposed to the use of story-based approaches. Finally, selecting a set of stories is not in itself 'analysis'. The analysis comes from distilling common themes and topics. Arguably, in the public representation of the process there has been an over-emphasis on the story component. As is evident in the preceding section, "How to undertake a performance story report evaluation", the overall methodology strongly emphasises analysis and the utilisation of other data (including empirical indicators) to have a comprehensive evidence base to support the evaluation report's conclusions.

From a social research perspective, the analysis should be done on all available data, specifically using all interview transcripts that are available, not just the stories selected for inclusion in the report. Using software tools like NVivo<sup>2</sup> to code the original interview transcripts (rather than the enhanced stories) would be appropriate. The selected stories are intended only to exemplify the information in the evaluation report and do not constitute the data or the analysis. However, the performance story reporting process provides a method for selecting and ratifying the stories used as vignettes.

---

<sup>2</sup> See <http://www.qsrinternational.com>

There are other ways of constructing and/or selecting vignettes (indicative stories). In fact, there is no reason why the researcher-evaluators themselves can't construct them drawing on a cross-section of interviews (e.g. as has been done by Dare et al. 2011). It is not clear that the method described in the preceding pages is necessarily the most robust, depending on the audience and the role of the vignette in the report. Conceivably the expert panel process of selecting the stories in performance story reporting is about ensuring the external legitimacy of the process, rather than necessarily being about the integrity of the data and the analysis.

Potentially there might be stakeholder concern about the performance story reporting method because of a lack of awareness of the underlying analysis procedures used and because of a concern that the stories have been somehow manipulated to exaggerate the success of the program and/or its outcomes. When it is clearly explained that the process of enhancing the stories is only to ensure that they fairly represent the actual events and that any redrafting is not intended to change the substance of the story but only to ensure that it communicates the story more effectively, then there might be greater acceptance, especially when it is clear that there is a wider evidence basis used in the analysis. Performance story reporting is meant to be a rigorous analytical process, it is not just about the stories – this is partly why Jessica Dart has now chosen to rename the method to 'Collaborative Outcomes Reporting'.

Like all forms of social research, the method relies on the integrity of all participants to be truthful – participants in what they say, facilitators in what they write down and how the stories are prepared, the evaluators in the way the analysis is undertaken, and the selection panel (expert panel) in how stories are selected. In the Performance Story Reporting process, the expert panel is meant to provide credibility and reputation, and the members of the panel are meant to be in a position of knowledge to collectively know what is reasonable. There are two problems with this. First, I'm not sure that a group of external experts who are beyond reproach actually necessarily can know what is happening at the ground level and therefore be an authority on the integrity of the stated outcomes and the accuracy of the stories. Secondly, as with many group processes, there is potential for 'groupthink' to occur, and while the experts potentially influence the process in various ways, arguably their primary role is to provide credibility and legitimacy to the process. It is unlikely that an expert who was somewhat uneasy about something would speak up unless it was a serious matter and/or they felt very strongly about something. This does not mean that the stories (and/or resultant evaluation) are necessarily exaggerated as this would depend on the quality of the stories. In my personal experience, the stories we were considering were not particularly well constructed or told, and my hunch is that they under-represented the outcomes of the program.

A final point is that many commentators might argue that the expert panel should comprise the participants who contributed stories themselves, and not external authority figures. Apart from being more consistent with principles of participatory process and deliberativeness, many would argue that the participants themselves are the best ones to judge what is realistic. Personally I am in two minds about this, but I do think it fair to say that there would be much more positive feedback to the

participants if they were involved in the selection panel, and therefore to the extent that this was intended to be formative evaluative, it would be more effective.

To summarise my reflections: performance story reporting has real potential, but I don't think it is yet fully mature as an evaluation method. I think it is confused by its purpose – on the one hand trying to be a rigorous procedure for integrating multiple methods, on the other hand being a genuine process of narrative collection and analysis. From my perspective, as an integrating process it is either under-developed or it offers little more than what is available with other mechanisms for integration, including good social research generally (which is inherently multi-method and integrating). As a mechanism for qualitative, story-based evaluation, performance story reporting is not fully convincing and not fully thought through. In the Australian context, Dart and collaborators conceivably have had to make many concessions to sceptical key stakeholders who were not fully committed to a narrative approach. Therefore the resultant method as elaborated in the Australian Government's User Guide is a compromise approach.

The key point, however, is that story-based evaluation is an effective way of collecting evidence of change, especially in contexts where there are not adequate empirical indicators, where causality is hard to establish, and where there may be external factors that influence the outcomes. In these contexts, the stories of participants provide a means of determining success. Just like with all qualitative social research, the robustness of the research is established through the professionalism of the researchers, and the consistency of stories from multiple sources (triangulation). I personally feel that story-based evaluation could be an appropriate approach in the context of EU cohesion funds, especially as an augmentation to other forms of evaluation.

## **SPECULATION ON THE FEASIBILITY OF STORY-BASED EVALUATION IN THE CONTEXT OF EU COHESION POLICY**

The application of a story-based approach to evaluation in the Australian natural resource management context establishes the potential of the approach. There is nothing particularly unusual about the Australian situation that would mean that it would apply there but not elsewhere such as in Europe. Indeed, it has been used in many other contexts, including widely in developing countries, and in Canada where it was used to evaluate the effectiveness of the various cultural programs undertaken by the City of Surrey as part of its Cultural Capital of Canada Award (see Withers 2010).

Although an official evaluation of the utilisation of performance story reporting in Australia has not been publicly released (other than the feasibility assessment, Carr & Woodhams 2008), information received through informal channels suggests that in general it worked well, although there were some teething problems with its application. As an evaluation technique, it was unfamiliar and some people thought it a little strange and were cautious about accepting it. There was concern about the reliability and validity of the approach. Because it was done on a large scale, there was inconsistency between the consultants used, especially in exactly how the stories were collected and shaped. The various

manuals that were prepared were meant to ensure relative consistency, but there was still the possibility of either ambiguity in the directions, or in various people deciding to make their own adaptations. There was also concern about the time taken to do it and the cost. These comments are typical of something being tried for the first time, and do not represent a substantive concern with the method. With more experience in its use, it is likely that a more standardised approach and increasing familiarity and acceptance would be achieved.

The anecdotal evidence suggested that the participants liked the process of developing stories and story selection. They felt involved and empowered in a way that other evaluation processes can not provide for. The story process gave them immediate feedback and not only validated their experience, but sometimes helped them to conceptualise or consolidate their experiences.

As was done in Australia, it would be wise to have a pilot program to undertake a feasibility assessment and to develop and perfect the approach in a European context. Nevertheless, thinking about the European situation and the types of interventions typically funded suggests the following:

- Local development interventions in an urban or rural setting – this seems to be an ideal situation for the use of story-based evaluation. Because of the disparateness of the range of activities, selecting common indicators could be difficult. Story-based evaluation would enable the collation of evidence of change even where the on-ground activities varied considerably.
- Innovation support and/or enterprise support – companies and other organisations (e.g. universities) could create stories about what they used the support for and what difference it made. In these cases, empirical indicators may be available – at least in terms of the improvement in the financial performance of profit-oriented companies – however the stories are likely to highlight additional added-value dimensions that are not revealed through the figures alone.
- Support for dynamic systems involving numerous different stakeholders such as clusters – measurement in such contexts is always difficult because of differing units of analysis. The dynamic nature implies that the system is in a state of flux such that movement on any one indicator does not necessarily explain what is happening. Stories have a greater potential to explain how the funding support made a difference to different stakeholders in the system.
- Incentive grants and loans intended to stimulate behaviour change – potentially these initiatives are difficult to evaluate with existing indicators because they are often quite specific. Nevertheless, often ad hoc indicators can be developed that measure the changed behaviour, especially where the behaviour relates to consumption. However, where the behaviour change is not immediate it may be harder to identify appropriate indicators. Here story-based evaluation can assist. Qualitative approaches might also provide a greater understanding about why the program led to changed behaviour. Perhaps the grant or loan was not the main cause and the measured relationship is spurious. Qualitative evaluation would also be able to

provide a sense of whether the size of the grant or loan was appropriate to have the necessary stimulus effect, or whether it need to be increased to expand uptake, or whether it could be decreased without sacrificing it effectiveness.

- Training and capacity building – measures of investment for training and capacity building can readily assess the numbers of people attending activities, and where there are assessment (examination) processes, the number of who passed. Satisfaction surveys can also be undertaken. All these however are largely outputs, not outcomes. The outcomes of training are what difference it makes to the lives of those who did the training. Story-based evaluation is likely to provide much more information about the outcomes of training and capacity building programs than any empirical measure can.
- Investment in infrastructure (roads, rail, environmental infrastructure) – it is worth highlighting that story-telling approaches might reveal many more benefits (and potentially problems) about improvements in infrastructure. Improved public transport for example not only reduces journey-to-work time, but making it accessible to mobility-restricted people potentially makes a world of difference to them enabling them to get jobs, have a wider range of entertainment, better access to shopping, and greater autonomy over their life. The crude empirical indicators give no sense of the richness and value that increasing transport options can have.

Thus in all types of EU funding, whether or not empirical indicators are available, story-based approaches will always provide additional information. In a summative context, story-based approaches will provide additional evidence especially of the extra 'social return on investment'. In a formative context, story-based and qualitative evaluation will provide more information about how the program can be improved.

## ANSWERS TO SOME FREQUENTLY ASKED QUESTIONS

### **1. Are there any ethical considerations that are particular to the approach?**

There are no particular ethical considerations applying to story-based approaches to evaluation than would also apply to other evaluation techniques or social research. The standard ethical principles that apply to all social research equally apply. It would be generally expected that:

- participants have given informed consent
- their participation is voluntary and that there has been no coercion or threatened or implied retribution for non-participation
- that participants can withdraw at any time, and have their data removed from the analysis (where this is possible)
- that there is full disclosure of funding sources
- that there will be no harm to participants; that the researchers have given consideration about the potential for inadvertent harm; and where emotional distress etc arises as a result of participation in the research, that the researcher is able to assist the participant to resolve this or to seek appropriate professional care
- that interviewers and other staff demonstrate respect towards participants
- that anonymity can be presumed, or if it is intended to attribute statements that the expressed permission for the use of names be provided, including where a person's identity is evident from the context (for example, the mayor, governor, president of a particular organisation etc)
- that there be confidentiality with respect to all private matters (or when any such an undertaking is given)
- that there be full disclosure of research methods used to enable replication of the research by another researcher, and to enable peer review of the adequacy and ethicality of the methodology, and to encourage critical self-reflections on the limitations of the methodology and the implications of this for the results and conclusions.

### **2. Is it subject to bias or open to distortion or manipulation?**

All forms of social research are affected by 'bias'. Bias is a technical term in social research meaning a systematic tendency to favour one outcome of another as an inherent feature of the methods used. It is different to random error. Bias can occur in terms of response bias to surveys (to what extent are people who answer surveys different to those who don't respond). Bias can occur in situations where social desirability encourages people to understate or sometime overstate their experience (for

example their weekly alcohol consumption). People's beliefs about what the research will be used for will affect their likely answers. Thus all forms of social research are affected by bias, and qualitative methods are also so affected.

It is highly likely however that qualitative methods are less subject to bias than many quantitative methods. Firstly, it is part of the code of practice of qualitative researchers that they be acutely aware of how their techniques may cause bias and that they reflexively take steps to minimise bias. Secondly, in an interview where an interviewer has an inkling that the statements of the respondent are exaggerated, this can (and should) be annotated on the transcript.

In terms of the Performance Story Reporting method used in Australia, bias was minimised by the use of expert panels whose task was to ensure that the stories selected accurately and fairly represented the typical experiences of the participants in the program activities.

### ***3. I can see how it works at the project/activity level, how does it apply at the program level?***

There are two responses to the question. The first response is that stories can be collected from all participants, not only community participants in program activities, but also program managers and coordinators. The stories that are collectively assessed for the evaluation could come, for example, from the people in charge of the program in their own region.

The second response is that program evaluation can be seen as a type of meta-evaluation. Programs ultimately come down to activities on the ground that are coordinated at various levels. The way the evaluation of the program in Australia worked was that there were story-based evaluations done at the regional level, with a report being written for each region. The overarching program evaluation was then a meta-analysis of the various regional level Performance Story Reports.

### ***4. Does it work for summative evaluation as well as for formative evaluation, and can it measure 'impact'?***

Summative and formative are terms that are used to describe the purpose to which evaluation is being put to. They represent a continuum rather than completely separate entities. Formative is focused on contributing to the ongoing development and improvement of the program, whereas summative attempts to be an ex-post, independent, objective assessment of outcomes that assesses whether the extent to which the program was responsible for (i.e. caused) the outcomes, and potentially undertaking a cost-benefit or return on investment calculation. For these reasons, summative evaluation tends to require empirical indicators, while formative evaluation tends to be qualitative.

Story-based approaches to evaluation, however, tend to be for summative evaluation purposes. Formative evaluation uses a wide range of qualitative methods and because of the developmental process of the program, there usually is no issue that questions the validity of the qualitative data, especially in terms of views about how the program could be improved. Formalised story-based

approaches tend to be used for summative evaluation in situations where there are no empirical indicators, no baseline measurements, no previously-identified program logic, and complex multi-factor and/or changing environments, and/or situations with long lag times between program/project activities and likely outcomes. In situations with long lag times, even if empirical indicators are available, they may not show movement in the political timescales necessary to make decisions about funding. In these situations, just as occurred in Australia with the need to evaluate the Caring for our Country program, performance story reporting provided a solution.

To repeat: story-based evaluation (and performance story reporting in particular) are forms of qualitative evaluation specifically adapted to be appropriate in the context of summative evaluation. As qualitative methods, they are not intending to provide empirical measures, however as Dart emphasises, performance story reporting and collaborative outcomes reporting are meant to be integrative approaches. Furthermore, the qualitative research may generate ideas that could lead to the identification of possible empirical indicators.

## **CONCLUSION**

There is ample testament to the power of stories. Tell a person an isolated alleged 'fact', and they wonder about its veracity. Tell a person a proper story and it will likely be accepted. Stories are more engaging, stories are more meaningful. Stories are more real. Stories convey information more effectively and are more likely to be remembered.

An effective story has to be a proper story. It can't be an inchoate amalgam of odd ideas. To be an effective story, it needs to conform to the standard basic elements of all stories. It needs to have a beginning, a middle and an ending. It needs to have a coherent and credible storyline running through the story. It needs to be multi-dimensional, but the different components need to be connected and the causal relations between the components needs to become clear in the course of the story. It needs to be personal and emotional. Kurtz (2009) provides much advice on how to construct good stories.

Telling stories as a means of effecting behaviour change is an ancient art. Biblical parables, children's fables, classic mythology, and good literature all seek to influence their readers. Using stories to understand, analyse and make sense of things is more recent, but has been part of strategic planning in business for some decade. Using stories as an evaluation methodology is more recent, but has much appeal and as has been shown in this report, has been effectively implemented.

It is not intended that story-based approaches replace quantitative indicators where they are available. The intention is that the stories complement the quantitative indicators and indeed that they add value to those indicators by providing meaning and interpretation. Big programs are subject to long lag times, and can be subject to the influence of multiple external influences. It is naïve to think that have a list of simple indicators will reveal the complex processes taking place. Stories are therefore a much more effective way of understanding what is happening.



## REFERENCES

- Auditor General 2008 *Regional Delivery Model for the Natural Heritage Trust and the National Action Plan for Salinity and Water Quality* (Audit Report No.21 2007-08). Canberra: Australian National Audit Office. Last accessed 27 January 2012, at: [http://www.anao.gov.au/uploads/documents/2007-08\\_Audit\\_Report\\_21.pdf](http://www.anao.gov.au/uploads/documents/2007-08_Audit_Report_21.pdf)
- Bennett, C. 1975 Up the hierarchy, *Journal of Extension* 13(2), 7-12.
- Carpiano, R. 2009 Come take a walk with me: The 'Go-Along' interview as a novel method for studying the implications of place for health and well-being, *Health & Place* 15(1), 263-272.
- Carr, A. & Woodhams, F. 2008 *Monitoring, Evaluating, Reporting and Improving (MERI) Natural Resource Management: Making MERI in Australia*, Canberra Bureau of Rural Sciences.
- Cash, D., Clark, W., Alcock, F., Dickson, N., Eckley, N., Guston, D., Jäger, J. & Mitchell, R. 2003 Knowledge systems for sustainable development, *Proceedings of the National Academy of Sciences USA*, 100(14), 8086-8091.
- Chambers, R. 1994 The origins and practice of participatory rural appraisal, *World Development* 22(7), 953-969.
- Commonwealth of Australia 2008 *Caring for Our Country Outcomes 2008-2013*. Canberra: Commonwealth of Australia. Last accessed 27 January 2012, at: <http://www.nrm.gov.au/resources/index.html>
- Cooperrider, D., Whitney, D. & Stavros, J. 2008 *Appreciative Inquiry Handbook: For Leaders of Change* (2<sup>nd</sup> edn), San Francisco: Berrett-Koehler.
- Curtis, A. & Lockwood, M. 2000 Landcare and catchment management in Australia: Lessons for state-sponsored community participation, *Society & Natural Resources* 13(1), 61-73.
- Dare, M., Vanclay, F. & Schirmer, J. 2011 Understanding community engagement in plantation forest management: Insights from practitioner and community narratives, *Journal of Environmental Planning and Management* 54(9), 1149-1168.
- Dart, J. & Davies, R. 2003 A dialogical, story-based evaluation tool: The Most Significant Change Technique, *American Journal of Evaluation* 24(2), 137-155.
- Dart, J., Drysdale, G., Cole, D. & Saddington, M. 2000 The Most Significant Change approach for monitoring an Australian extension project, *PLA Notes* 38, 47-53 (International Institute for Environment and Development, London).
- Dart, J. & Mayne, J. 2005 "Performance Story". In: Mathison, S. (ed.), *Encyclopedia of Evaluation*. Thousand Oaks: Sage, pp.307-308.
- Dart, J. & O'Connor, P. 2008 Performance Story Report: A Study of the Mount Lofty Ranges Southern Emu-Wren and Fleurieu Peninsula Swamps Recovery Program and how it contributed to biodiversity outcomes in the Adelaide and Mount Lofty Ranges Natural Resources Management Region. Canberra: Commonwealth of Australia. Last accessed 27 January 2012, at: <http://nrmonline.nrm.gov.au/catalog/mql:2167>
- Davies, R. & Dart, J. 2005 *The 'Most Significant Change' (MSC) Technique: A Guide to its Use* (self-published manual for a consortium of development agencies). Last accessed 27 January 2012, at: <http://www.mande.co.uk/docs/MSCGuide.pdf>
- Denning, S. 2007 *The Secret Language of Leadership: How Leaders inspire Action through Narrative*. San Francisco: Jossey-Bass.
- DWLBC, 2009 *Our Changing Environment: Outcomes and Lessons from Investment in South Australia's Natural Resources 2001-2008*, Government of South Australia, through the Department of Water, Land and Biodiversity Conservation, Adelaide. Last accessed 27 January 2012, at: [http://www.nrm.sa.gov.au/Portals/0/NRMinSA/Our\\_changing\\_environment\\_full.pdf](http://www.nrm.sa.gov.au/Portals/0/NRMinSA/Our_changing_environment_full.pdf)
- Earl, S., Carden, F. & Smutylo, T. 2001 *Outcome Mapping: Building Learning and Reflection into Development Programs*, Ottawa: International Development Research Centre.
- Fetterman, D.M. 2000 *Foundations of Empowerment Evaluation*. Thousand Oaks, CA: Sage.
- Fetterman, D.M., Kaftarian, S.J. & Wandersman, A. (eds) 1995 *Empowerment Evaluation*. Thousand Oaks, CA: Sage.

- Fisher, W.R. 1989 *Human Communication as Narration: Toward a Philosophy of Reason, Value, and Action*. Columbia: University of South Carolina Press.
- Greene, J.C. 2000 "Qualitative program evaluation: practice and promise". In: Denzin, N.K. & Lincoln, Y.S. (eds), *Handbook of Qualitative Research*. Thousand Oaks: Sage, pp.530-544.
- Guba, E. & Lincoln, Y. 1989 *Fourth Generation Evaluation*. Newbury Park, CA: Sage.
- Guijt, I., Brouwers, J., Kusters, C., Prins, E. & Zeinalova, B. 2011 *Evaluation Revisited: Improving the Quality of Evaluative Practice by Embracing Complexity (Conference Report)*, Wageningen: Centre for Development Innovation, Wageningen University & Research Centre. Last accessed 27 January 2012, at: [http://capacity.org/capacity/export/sites/capacity/documents/topic-readings/110412-evaluation-revisited-may-2010\\_small-version.pdf](http://capacity.org/capacity/export/sites/capacity/documents/topic-readings/110412-evaluation-revisited-may-2010_small-version.pdf)
- Hartz-Karp, J. & Pope, J. 2011 "Enhancing the effectiveness of SIA through deliberative democracy". In: Vanclay, F. & Esteves, A.M. (eds), *New Directions in Social Impact Assessment: Conceptual and Methodological Advances*. Cheltenham: Edward Elgar, pp.253-272.
- Jackson, E.T. & Kassam, Y. (eds) 1998 *Knowledge Shared: Participatory Evaluation in Development Cooperation*. West Hartford, CT: Kumarian Press.
- Kurtz, C. 2009 *Working with Stories in your Community or Organization*. Last accessed 27 January 2012, at <http://www.workingwithstories.org>
- Laughlin, R. & Broadbent, J. 1996 Redesigning fourth generation evaluation. *Evaluation* 2(4), 431-451.
- Leeuw, F. & Vaessen, J. 2009 *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*. Washington DC: The Network of Networks on Impact Evaluation (c/- Independent Evaluation Group, The World Bank). Last accessed 27 January 2012, at: [http://siteresources.worldbank.org/EXT/OED/Resources/nonie\\_guidance.pdf](http://siteresources.worldbank.org/EXT/OED/Resources/nonie_guidance.pdf)
- Mayne, J. 1999 *Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly* (Discussion Paper) Office of Auditor-General of Canada. Last accessed 27 January 2012, at: <http://dsp-psd.pwgsc.gc.ca/Collection/FA3-31-1999E.pdf>
- Mayne, J. 2004 Reporting on outcomes: Setting performance expectations and telling performance stories, *The Canadian Journal of Program Evaluation* 19(1), 31-60.
- Patton, M.Q. 1981 *Creative Evaluation*. Beverly Hills, CA: Sage.
- Patton, M.Q. 1990 *Qualitative Evaluation and Research Methods*. Newbury Park, CA: Sage.
- Patton, M.Q. 1997 *Utilization-Focused Evaluation*. Thousand Oaks, CA: Sage.
- Prager, K. & Vanclay, F. 2010 Landcare in Australia and Germany: Comparing structures and policies for community engagement in natural resource management, *Ecological Management & Restoration* 11(3), 187-193.
- Roughley, A. 2009 *Developing and Using Program Logic in Natural Resource Management: User Guide*, Canberra: Australian Government. Last accessed 27 January 2012, at: <http://nrmonline.nrm.gov.au/catalog/mql:2164>
- Roughley, A. & Dart, J. 2009 *Developing a Performance Story Report User Guide*, Canberra: Australian Government. Last accessed 27 January 2012, at: <http://nrmonline.nrm.gov.au/catalog/mql:2162>
- Sandelowski, M. 1991 Telling stories: Narrative approaches in qualitative research, *Image: The Journal of Nursing Scholarship* 23(3), 161-166.
- Shaw, G., Brown, R. & Bromiley, P. 1998 Strategic stories: How 3M is rewriting business planning, *Harvard Business Review* 76(3), 41-50.
- Silver, M., Richards, R. & Cowell, S. 2009 *A Guide to Designing and Implementing Performance Reporting to increase the Confidence of Conservation Investors*. Melbourne: Bush Heritage Australia. Last accessed 27 January 2012, at: <http://lwa.gov.au/files/products/reporting/pn30209/pn30209.pdf>
- Smith, G. & Wales, C. 1999 The theory and practice of citizens' juries, *Politics and Policy*, 27(3), 295-308.

- Vanclay, F., Lane, R., Wills, J., Coates, I. & Lucas, D. 2004 'Committing to Place' and evaluating the higher purpose: Increasing engagement in natural resource management through museum outreach and educational activities, *Journal of Environmental Assessment Policy & Management* 6(4), 539-564.
- Wang, C. & Burris, M. 1997 Photovoice: Concept, methodology, and use for Participatory Needs Assessment, *Health Education & Behavior* 24(3), 369-387.
- Waters, D., James, R. & Darby, J. 2011 Health promoting community radio in rural Bali: an impact evaluation, *Rural and Remote Health* 11(1), article #1555. (online journal) Last accessed 27 January 2012, at: <http://www.rrh.org.au/articles/subviewnew.asp?ArticleID=1555>
- Whitmore, E. (ed.) 1998 *Understanding and Practicing Participatory Evaluation: New Directions for Evaluation*. San Francisco: Jossey-Bass.
- Whyte, W.F. (ed.) 1990 *Participatory Action Research*. Newbury Park, CA: Sage.
- Wilder, L. & Walpole, M. 2008 Measuring social impacts in conservation: experience of using the Most Significant Change method, *Oryx* 42(4), 529-538.
- Willets, J. & Crawford, P. 2007 The most significant lessons about the Most Significant Change Technique, *Development in Practice* 17(3), 367-379.
- Withers, D. 2010 *A narrative evaluation of the City of Surrey's Cultural Capital of Canada Award Program*. Last accessed 27 January 2012, at: <https://files.me.com/denisewithers/73mf37>

#### **EXAMPLE KEY REFERENCES FOR QUALITATIVE SOCIAL RESEARCH METHODS**

- Bryman, R. 2001 *Social Research Methods*, Oxford: Oxford University Press.
- Kvale, S. 1996 *InterViews: An Introduction to Qualitative Research Interviewing*, Thousand Oaks, CA: Sage.
- Mason, J. 2002 *Qualitative Researching*, London: Sage.
- Patton, M.Q. 2002 *Qualitative Research and Evaluation Methods (3<sup>rd</sup> edn)*, Thousand Oaks, CA: Sage.
- Ritchie, J. and Lewis, J. 2003 *Qualitative Research Practice*, London: Sage.
- Silverman, D. 2000 *Doing Qualitative Research: A Practical Handbook*, London: Sage.

## EXAMPLES OF PERFORMANCE STORY REPORTS AS PART OF AUSTRALIAN NRM

(all accessed 23 January 2012)

A study of the Mount Lofty Ranges Southern Emu-wren and Fleurieu Peninsula Swamps Recovery Program and how it contributed to biodiversity outcomes in the Adelaide and Mount Lofty Ranges Natural Resources Management region

<http://www.conservation.sa.gov.au/files/emuwren/Performance%20Story%20-%20MLRSEW%20and%20FPS%20RP.pdf>

An evaluation of outcomes of the Ribbons of Blue environmental education program

[http://www.ribbonsofblue.wa.gov.au/component/option,com\\_docman/task,doc\\_download/gid,139/](http://www.ribbonsofblue.wa.gov.au/component/option,com_docman/task,doc_download/gid,139/)

Evaluation of investment in the dugong and marine turtle project

<http://www.savanna.org.au/nailsma/publications/downloads/Dugong-and-Marine-Turtle-Project-PSR.pdf>

Performance Story Report for the contribution of Strategic Tree Farming to Regional NRM Outcomes

<http://nrmonline.nrm.gov.au/catalog/mql:2165>

A study of land use change projects delivered in the Bundella catchment and how these investments have contributed to land use outcomes for the Namoi Catchment Management Authority

[http://www.namoi.cma.nsw.gov.au/namoi\\_cma\\_bundella\\_psr\\_25june09.pdf](http://www.namoi.cma.nsw.gov.au/namoi_cma_bundella_psr_25june09.pdf)

## EXAMPLES OF OTHER REPORTS USING STORY-BASED APPROACHES

(all accessed 23 January 2012)

A narrative evaluation of the City of Surrey's Cultural Capital of Canada Award Program (Denise Withers, 2010)

<https://files.me.com/denisewithers/73mf37>

The "Real Book" for story evaluation methods (Marc Maxson, Irene Guijt, and others, 2010)

<http://www.globalgiving.org/jcr-content/gg/landing-pages/story-tools/files/-story-real-book--2010.pdf>

## CONSULTANTS WHO ARE KNOWN FOR DOING STORY-BASED APPROACHES

***Disclaimer:*** The list below comprises those consultants who emerged from this research as operating in this area of activity and is not necessarily exhaustive of all consultants that might be suitable for such work. They are presented in alphabetical order. While Jessica Dart and Irene Guijt are personally to me, I have no relationship with any of the consultants likely to affect my judgement. The listing of these consultants here does not imply any endorsement by me.

Jessica Dart, Clear Horizon, Melbourne, Australia

<http://www.clearhorizon.com.au>

email: [info@clearhorizon.com.au](mailto:info@clearhorizon.com.au)

Rick Davies, Cambridge, United Kingdom

<http://richardjdavies.wordpress.com/>

email: [rick@mande.co.uk](mailto:rick@mande.co.uk)

Irene Guijt, Learning by Design, Randwijk, The Netherlands

Email: [iguijt@learningbydesign.org](mailto:iguijt@learningbydesign.org)

Cynthia Kurtz, "upstate New York", USA

<http://www.cfkurtz.com/>

Email: [cfkurtz@cfkurtz.com](mailto:cfkurtz@cfkurtz.com)

## APPENDIX 1 Brochure from Clear Horizon (reproduced with permission)

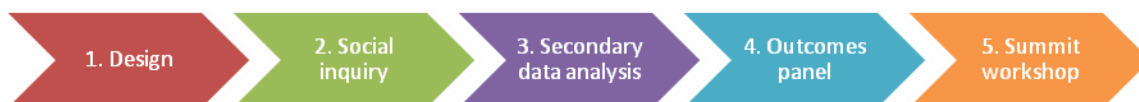
**Collaborative Outcomes Reporting technique (COR)**

The Collaborative Outcomes Reporting technique (COR) was developed by Dr Jess Dart and is characterized by two elements: a participatory process whereby the information is generated and a five-part structure in which the report product is presented.

**Report structure:** the report aims to explore and report the extent to which a program has *contributed* to outcomes. Under the COR, reports are short and generally structured against the following sections:

1. A narrative section explaining the **program context** and rationale.
2. A **'results chart'** summarising the achievements of a program against a theory of change model.
3. A narrative section describing the **implications** of the results e.g. the achievements (expected and unexpected), the issues and the recommendations.
4. A section which provides a number of 'vignettes' that provide **instances of significant change**, usually first person narratives.
5. An index providing more detail on the sources of **evidence**.

**Participatory process:** COR uses a mixed method approach that involves participation of key stakeholders in a number of process steps:



Participation can occur at all stages of this process for example:

1. **In the planning workshop.** In this workshop the theory of change is clarified, existing data is identified and evaluation questions developed. Consultants play the role of facilitation and documentation.
2. **In the data trawl.** Program staff may be enlisted to help with the collation of data, although in our experience consultants usually lead this process as the evaluation managers.
3. **The social inquiry process.** Volunteers can be given a short training session in interviewing and with the help of an interview guide can conduct the interviews. This is a very effective way to involve staff in the data where there is sufficient enthusiasm around the process. Otherwise consultants or the evaluation managers conduct all or a proportion of the interviews.
4. **Outcomes panel.** People with relevant scientific, technical, or sectoral knowledge are brought together and presented with a range of evidence compiled by the consultants. They are then asked to assess the contribution of the intervention towards goals given the available knowledge. We call this an outcomes panel and it is usually facilitated. It is sometimes also referred to as an expert panel. It can be substituted for a citizen's jury.
5. **Summit workshop.** At a large workshop instances of significant change are selected (incorporating aspects of Most Significant Change Technique) and key findings and recommendations are synthesised. The summit should involve broad participation of key stakeholders such as program staff and community members.

**Advantages:** Organisations often place a high value on the reports because they strike a good balance between depth of information and brevity and are easy for staff and stakeholders to understand. They help build a credible case that a contribution has been made. The participatory process by which reports are developed offers many opportunities for staff and stakeholder capacity building. Compared to standard outcomes evaluations approaches they are relatively straightforward. They are a great way to kick off a new monitoring and evaluation system, because they involve synthesising and reflecting on all existing data and data gaps (a great platform to think about what data is really needed!). It has been used in a wide range of sectors from overseas development, community health, and Indigenous education. But the majority of work has occurred in the Natural Resource Management Sector, with the Australian Government funding 20 pilot studies in 2007-9.

**Limitations:** COR's have been criticised for being too appreciative, or for being incapable of telling a bad story. While this is certainly a risk, the technique does attempt to address this in a number of ways. Firstly all informants are asked to describe the strengths and the weaknesses of the program. These weaknesses or issues are documented in the report. Secondly, the outcomes panel is encouraged to report on negative as well as positive trends in terms of the outcomes. So the "negatives" are not avoided in this process. However where COR is used as a case study approach, the choice of topic for a case outcomes report is often purposeful rather than randomly selected. Topics for reports are often selected on the basis that they are likely to show some significant outcomes. In addition COR only address one type of key evaluation question. That is the question concerning the extent to which an investment contributes to outcomes. It is an extremely important question, but it is not the *only* type of key evaluation question that is important. This needs to be understood and acknowledged. For this reason, COR should not be seen as the only reporting tool. The idea is that it should complement other reporting processes or be extended to encompass more.

**Values:** COR is based on the premise that the values of stakeholders, program staff and key stakeholders are of highest importance in an evaluation. The evaluators attempt to "bracket off" their opinions and instead present a series of data summaries to panel and submit participants for them to analyse and interpret. Values are surfaced and debated throughout the process. Participants debate the value and significance of data sources and come to agreement on the key findings of the evaluation.

**Quality:** The focus on quality is largely associated with *process quality*: ethical conduct; culturally appropriate methods; ownership of the evaluation process; ensuring that the evaluation provides credible but useful recommendations to inform the next phase of the program. Interviews are usually taped and transcribed. Data is double analysed by participants at workshops and by the consultants using thematic coding.

**Complexity:** COR is especially useful when a program has emergent or complex outcomes that are not fully defined at the onset of a program. For this reason a theory of change is refreshed at the start of the evaluation process. In addition qualitative inquiry is used to capture unexpected outcomes and deliberative processes are used to make sense of the findings.

**Resources:** Clear Horizon runs a two-day training program on this technique. We have also drafted a comprehensive User Guide that lays out all steps of the process for the NRM sector. See [www.clearhorizon.com.au](http://www.clearhorizon.com.au) Examples report can also be found here.

## APPENDIX 2 Case Study Application of Collaborative Outcomes Reporting

(reproduced with permission from Clear Horizon)

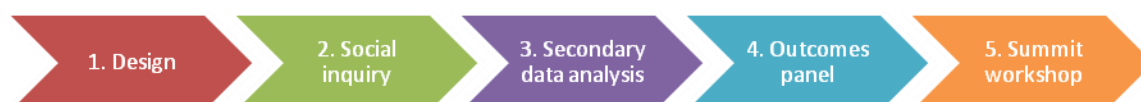
### **Evaluation of the Stronger, Smarter Realities program Jess Dart and the Stronger Smarter Institute**

In Australia there is a big disparity between educational outcomes for Indigenous children compared to non-Indigenous children, and in the last 8 years educational outcomes have been either stable or declining.<sup>3</sup> While Indigenous children have been staying longer at school, too many Indigenous students leave school without a formal Year 12 qualification. Nationally, the proportion of Indigenous students who achieved a Year 12 Certificate (as a proportion of students who were enrolled in Year 11 in the previous year), has decreased from 51% in 2001 to 46% in 2006. During that period the proportion of non-Indigenous students increased from 80% to 86% and the gap between Indigenous and non-Indigenous outcomes has widened.

This case is about a participatory evaluation of the first phase of the Australian “Stronger Smarter Realities Program” (SSR) which ran from 2006 to the end of 2008. This project was about creating systematic and transferable change by arming Australian educators with the belief, skills and capacity to make profound changes to the learning outcomes of Indigenous children. Over 3 years, the project aimed to engage principals, teachers and Indigenous community leaders from 240 schools with high Indigenous student populations, and support them to transform their schools in such a way to deliver dramatically improved educational outcomes for Indigenous students. The program is based on the premise that this can be achieved by providing a supportive educational environment, by providing excellent teachers and by having high expectations.

The program is delivered by the Stronger Smarter Institute (formerly Indigenous Education Leadership Institute (IELI)), Queensland University of Technology, Brisbane, Australia. It was funded jointly by two donors: the Sidney Myer Fund and the Telstra Foundation. The project has two streams: the Principal Leadership Program and Teacher Leadership and Community Leadership Program. This evaluation was primarily concerned with the Principal Leadership Program. The evaluation was completed in 2009 at the end of the first phase of the project by external consultants using a participatory approach. It was guided by two key questions i) to what extent has the SSR project contributed to excellence in Indigenous education in participating schools? And ii) to what extent did the SSR project influence the overall Indigenous education agenda?

The evaluation was both summative and formative in nature and largely focused on outcomes. It was summative in that it was conducted at the end of the first phase of the program and was required by the program funders. It was formative in that it was intended to influence the next phase and scaling up of the program. The evaluation used the “Collaborative Outcomes Reporting Technique (COR)” developed by Jess Dart. This is a mixed method approach that involved key stakeholders in a number of process steps.



Firstly, a design workshop was held where the theory of change was clarified and evaluation questions developed. This was conducted with program team members and key stakeholders in a participatory manner. Social inquiry included over 50 semi-structured interviews incorporating the Most Significant Change technique and 3 case studies from Indigenous communities. The data trawl involved collection and analysis of secondary documents and quantitative data on student outcomes from 10 schools.

<sup>3</sup> National Report on Indigenous Education, 2006 DEEWR

The quantitative data, case studies and qualitative summaries were used as evidence to feed into an 'outcomes panel' with Indigenous educators who examined the data and created statements about: the extent to which the outcomes had been achieved; the plausible contribution of the program to these outcomes and the quality of the data. The panel were selected as they were highly respected, had no vested interest in the program and had an excellent knowledge of Indigenous education policy and practice. The process culminated in an evaluation summit workshop that saw key stakeholders and staff deliberating over qualitative and quantitative data and creating recommendations. The consultants' role was collection and synthesized data and facilitation of the sensemaking process with recommendations created by workshop participants.

While the quantitative data was limited in scope, the evaluation was noteworthy as it managed to capture some of the less tangible outcomes concerning 'breakthrough learning' and raised expectations for Indigenous children. The program itself has been very successful and is being scaled-up and delivered on a national scale. This evaluation has been highly influential as evidenced by all the recommendations been successfully implemented, and one Philanthropic funder stating that the evaluation was well-balanced and gave them confidence to provide further funding for the program.

**Values and Quality.** This evaluation was based on the premise that the values of the Indigenous panel members, program staff and key stakeholders were of highest priority. Funders' views were not considered. The evaluators attempted to "bracket off" their opinions and instead presented a series of data summaries to panel and summit participants for them to analyse and interpret. The evaluators felt they were not the right people to make judgements concerning the effectiveness of a program.

Values were surfaced and debated throughout the evaluation process. The theory of change created a 'normative model' for how program staff viewed the program and this was used as the organising construct for the evaluation. Program staff debated and helped create the evaluation questions. Quotations and data were presented in a non-homogenised manner to panel and summit participants. Vignettes were analysed using the most significant change approach and the reasons for their selection were documented. Participants debated the value and significance of data sources and came to agreement on the key findings of the evaluation. Dot voting was used to determine how different participants rated each issue and outcome. Participant analysis and judgements were used to frame the evaluation report in terms of how findings were grouped, and the recommendations came from the workshops.

The focus on quality was largely associated with *process quality*: ethical conduct; culturally appropriate methods; ownership of the evaluation process; ensuring that the evaluation provided credible but useful recommendations to inform the next phase of the program. All interviews were taped and transcribed. Data was analysed in participatory ways and by the consultants using thematic coding. A weakness of the evaluation was the quantitative data; the data sets were patchy and the sample was limited. It proved extremely difficult to acquire this data from schools. One of the recommendations of the program was to create a more robust quantitative monitoring system, for the second phase of the program. While the program was not overly complex, the program outcomes and process were somewhat emergent, intangible and hard to measure. The "Collaborative Outcomes Reporting approach" has been successfully used with programs with higher degrees of complexity – often with no clear outcomes defined at the onset of the program. It is widely used for evaluating biodiversity conservation programs.