

Managing a complex evaluation – the case of “Our Rural Landscape”

Julie McGeary, Victorian Department of Primary Industries (DPI)

Martin Andrew, URS Australia

Simon Smith, DPI

Don Burnside, URS Australia

Lili Pechey, URS Australia

Todd Ritchie, URS Australia

Introduction

Australia’s agricultural landscapes and communities are changing in response to market pressures, consumer concerns about animal welfare and ecological sustainability, increased competition for land and water, and greater recognition of the natural environment’s limits in agricultural production. These pressures require governments to support the development of innovative technologies that will ensure the sustainable development of the food and agriculture sector.

Our Rural Landscape (ORL) was a key Victorian State Government strategic initiative, developed and managed by the Department of Primary Industries (DPI) specifically to tackle these critical issues. With a total budget of \$107 million, including co-investment, ORL ran over four years from July 2003 to June 2007, was delivered through 15 major projects and involved about 120 DPI staff and multiple collaborators.

ORL itself was part of the Victorian Government’s \$310m Science, Technology & Innovation Initiative Program, Phase 2 (STI-2) aimed at improving science, technology and innovation skills, and benefits of commercial, scientific research, collaboration and science awareness natures. The program is linked to the Victorian Government’s *Growing Victoria Together* vision statement.

A final evaluation of the ORL Program was commissioned to provide accountability, organisational learning, and to enable evidence-based decision making. Following a public tender process, URS Australia was engaged in December 2006 to undertake the review.

The ORL Program evaluation aimed to determine the current and potential contribution of ORL funded projects to planned outcomes, and to assess the effectiveness and efficacy of the ORL management systems. The evaluation considered each project’s review and performance data and engaged a mixture of participants, collaborators and stakeholders. Importantly, the external review sought to demonstrate the aggregate impacts achieved by the overall investment.

This paper focuses on the complexity of the ORL Program, and how the independent program evaluation was managed in response to those complexities.

Management of the Evaluation

The URS team comprised four core people; two experienced agricultural program evaluators, and two resource economists. The evaluation was project managed for DPI by the ORL Evaluation Officer, and all was overseen by an ORL Evaluation Steering Group comprising senior DPI Managers and representatives from the program’s Victorian Government investors.

Complexity and evaluation challenges

ORL was a complex rather than a complicated program (Glouberman and Zimmerman 2004). It was a policy driven initiative operating at multiple levels which aimed to make a difference at the system level (public sector science investment policy), on institutions within that system (various government departments, collaborating science institutions, various agricultural production sectors and agri food industries), and providers within institutions (scientists, program managers, land managers, planners) (Owen 2006).

ORL was a multi component program, comprising 15 major projects and almost 50 sub-projects, all expected to contribute to at least some of the overarching policy objectives. It was a multi-site program, with activities conducted at 23 sites across Victoria, with half the sites and half the staff located in regional areas. It involved multiple organisations working together with unpredictable collaboration outcomes, and could be viewed as an open system, influenced by unpredictable factors such as drought and public sector restructure. Causality was therefore likely to be recursive rather than

linear, and program outcomes were likely to be emergent in response to opportunities rather than specific and pre-defined (Rogers 2007).

ORL existed within the Science, Technology and Innovation (STI¹) domain, with its own inherent complexities. Molas-Gallart and Davies (2006) describe the multiple goals of STI policy, the indirect and complex linkages between policy outputs and outcomes, and the serendipitous and long-term nature of policy impacts, acknowledging that these make the evaluation of STI policies challenging.

A recent OECD report (cited in Productivity Commission 2007 p122) highlighted four basic problems affecting the relationship between research outputs and outcome effects, all of which are relevant to the evaluation of ORL. Collectively, these comprise the first challenge faced in the evaluation.

1. *Timing*, whereby the effects of research are often manifested long after the research has been completed and the connections obscured. Adding to the complexity is the fact that the time period from investment to delivery of outcomes from agricultural R&D can be particularly lengthy (Chudleigh, Simpson & Schofield 2006).
2. *Attribution*, whereby a given innovation may draw upon multiple research projects and a given research project may impact upon multiple innovations. An innovation also depends upon many inputs other than research before market or social effects are realised (Productivity Commission 2007 p122).
3. *Appropriability*, whereby the beneficiaries of research may not be the same people or organisations who performed it so it may not be obvious where to look for effects. Indeed, ORL's state funded research is addressing national scale issues, and the benefits are likely to be realised nationally, or in some cases, internationally.
4. *Inequality*, whereby the distribution of impacts in a given project portfolio is typically highly skewed. A small number of projects ('blockbusters') may account for the majority of effects, while around half often do no more than advance knowledge in a general way (Productivity Commission 2007 p122). This has implications for sampling strategies, and for this reason, the Productivity Commission (Productivity Commission 2007) recommends that where undertaken, selective case studies of impacts should be placed in a supplementary rather than central evaluation role.

A second evaluation challenge was handling the variation in the type of research being conducted within ORL, and consequent variation in the 'distance' to adoption of the outcomes of that research. ORL projects included cutting-edge plant, animal, fish and soil microbial genomics research, nanotechnologies for food sciences and genomics for food bio-security issues. Research was conducted which enabled better linking of public policy to the changing socio-economic profile of rural Victoria, and yet other work aimed to foster a culture of innovation within DPI and community stakeholders. While most of the research tended to be at the fundamental end of the R&D spectrum, some work was further along the 'route to market' than others. The evaluation was thus an *ex ante* utilisation study (Productivity Commission 2007 p 662), undertaken after most of the research had been successfully completed, but before most of the research outcomes had been commercialised or otherwise utilised.

ORL investment occurred across a range of agricultural industry sectors, and this presented a third evaluation challenge. The consultants required a detailed knowledge of each sector to enable potential benefits of the program to be reliably predicted.

A final evaluation challenge required the reporting of benefits of ORL investment at the aggregated program level. This challenge is compounded where occasional 'failure' at a project level can be expected and where separate but concurrent projects often deliver their outputs and outcomes at different periods of time. Approaches range from simple counts of successful outputs falling within range of pre-specified targets, to more sophisticated summation and weighting measures. Frequently, however, informed professional judgment will be required (Productivity Commission 2007 p 326).

Our approach

Evaluation framework

We adopted a multi-method approach for the evaluation of ORL. URS began with a comprehensive review of contextual documents, as well as program and project documentation and data. This achieved an understanding of the policy context in which ORL operated, the science and technology base upon

¹ Note – this is a generic classification of research type, not to be confused with the Victorian Governments STI-2 Program of which ORL was a part

which the research was founded, and the rationale and expectations of individual research components of the program. The documents also served as a means of 'truthing' subsequent claims of achievement made by program participants.

Figure 1 shows an Outcomes Hierarchy for ORL, developed early in the program's planning stage. While useful in guiding development of individual projects to align with program theory, these documented outcomes and impacts also became program expectations, presenting the evaluation of ORL with a pre-defined set of criteria against which success would be measured.

To accommodate reporting against these measures of success, URS developed a program logic framework, identifying performance indicators for each of the practice change, outcome and impact measures appearing in the Hierarchy. Both lagging indicators – measures of actual achievement, and leading indicators – measures showing the program is on track to achieve, were developed.

Stakeholder consultation

Comprehensive stakeholder consultation was a requirement of the evaluation contract.

A workshop was conducted with senior DPI Agriculture Managers to obtain their views on ORL and its contribution to broader policy objectives. Another four workshops were conducted with ORL project participants at regional locations across Victoria. Information was gained concerning participants' awareness and understanding of projects other than their own. The workshops provided an opportunity for people to reflect collectively on their separate and shared ORL experiences.

Seventy two interviews were conducted with ORL project leaders, project team members, next users, end users and other external stakeholders. Half of these interviews were face-to-face and half by telephone. These conversations, and those at the workshops, were guided by sets of questions derived from the evaluation framework.

An on-line questionnaire was developed based on the program log frame. It contained questions about the respondent's familiarity with ORL, separate sections for those directly or indirectly involved with ORL, which used Likert scales for scoring of achievements at the project level, and questions for all respondents regarding the ORL Program as a whole, using Goal Attainment Scaling (Kiresuk, Smith & Cardillo 1994). All workshop and interview participants were encouraged to complete the questionnaire, and after several reminders, 60 responses were received, mostly from those people involved in program activities.

Economic analysis

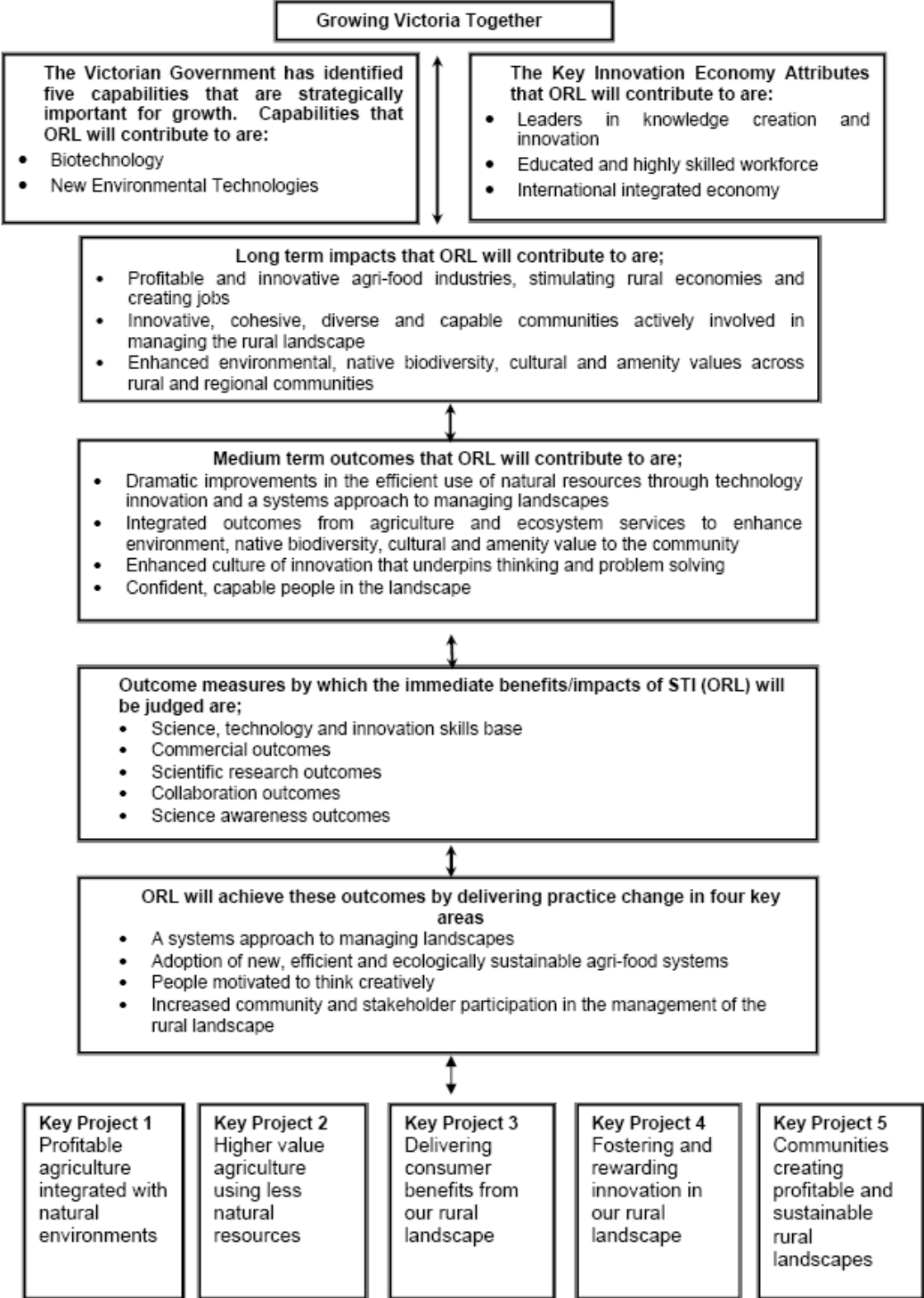
As per the contract, URS undertook an economic analysis to assess benefits of the research for five case studies. Benefits and costs associated with adoption of new technologies developed were discussed with project staff, as well as potential adoption rates. The DPI Economics and Policy Research Branch used this information to model the potential gross margin benefits, calculated as \$/ha/annum. Break-even analyses were done to illustrate the area of land over which new technologies would need to be adopted to cover the research costs for each case study. Because few of the case study projects collected actual cost, benefit and adoption rate data, there was uncertainty associated with the estimated parameters, and they could only be considered reflective of the order of magnitude of the potential gross benefits. While it may be impossible to give accurate estimates of the benefits for the economy of R&D stimulated by public finding support, (Productivity Commission 2007 p 139), the level of accuracy achieved in the ORL evaluation was disappointing.

Reporting

The URS team collated and summarised the evaluation data and brainstormed it as input to the draft report. The Steering Group workshopped the draft with URS, who incorporated the formal feedback which was collated by the ORL Evaluation Officer from Steering Group members and ORL senior managers. Two feedback iterations resulted in a final report which accommodated all requested factual amendments and clarifications.

URS presented a summary of key findings to an audience of ORL project team members during a program wind up meeting in June 2007. The results were robustly discussed, and the value of the evaluation as a project promotion tool was appreciated. URS gave a second presentation to an audience of DPI senior executives in July 2007.

Fig 1 Outcomes hierarchy depicting how ORL contributes to high level State outcomes



Key Findings

In short, the evaluation found that ORL as a program has contributed to STI-2 outcomes, particularly in scientific research, science capability and collaboration areas ORL was on track to achieve its DPI outcomes and the program was addressing leading indicators relating to innovative policy making, generation of intellectual property (25 provisional patents) and investment in R&D, with almost dollar

for dollar leveraged as cash co-investment. URS concluded: “*ORL has been an impressive program. It has invested significant money in a range of strategic R&D that will generate benefits for Victoria and beyond. It has national and international relevance, and demonstrates the commitment of the Victorian Government to the statements about resourcing science and innovation made in Growing Victoria Together*”.

Key issues and learnings from the Evaluation process

Document review – the sheer volume of documentation provided to URS by DPI was problematic. Apart from overall program planning, implementation and early review documents, each project contributed project plans, quarterly reports, annual business plans, annual output monitoring data, as well as technical notes, conference papers and journal articles. No attempt had been made to ‘map’ the documents, so URS were required to spend considerable time sorting the material. Version control by ORL program management was less than ideal, and additional time was incurred locating latest or current versions of some documents.

Outcomes hierarchy – the outcomes hierarchy shown in Figure 1 represents a theory of action at the program level. Each individual project was also required to develop its own theory of action, and these were generally based on a Bennett’s Hierarchy, often modified for a R&D configuration (Bennett & Rockwell 1995). DPI has for many years used program logic for project planning, and Bennett’s has been its preferred model because it was designed for agricultural extension-type projects. However it did not fit well the fundamental research nature of most of the ORL projects, and its use caused angst amongst some ORL researchers.

Program log-frame – much effort was invested in developing the complete program log frame, with Steering Group and ORL Management Team involvement in selection of the performance indicators even though it was realised that only the first steps could be achieved during the four years of ORL, which was indeed the case. The evidence was that ORL was aligned with leading indicators i.e. ORL was on track to achieve outcomes and impacts over time.

Profile of interviewees – The list of people for interviews was determined by ORL Project Leaders and the ORL Management Team. It comprised mainly DPI staff. The few external stakeholders were scientists and consultants delivering services under contractual arrangements to ORL projects, or were employed in collaborating organisations (but not as members of project teams), or were involved as members of Reference or Steering Groups, or were next users from industry. It became evident that although project teams had given thought to next and end users in a generic sense, they had some difficulty identifying specific individuals who were truly external to the program and would know sufficient about it to provide informed advice. Thus the interviewees and questionnaire respondents were skewed towards DPI staff with a high knowledge of, and often a high commitment to, ORL projects. This skewed sample was problematic for interpreting the interview and questionnaire data. URS addressed this problem by focusing questions on how the work in ORL would be carried forward, and by whom, and through a deliberate search for relevant supporting information in the documents.

Questionnaire – The response rate to the questionnaire was modest. Given that the evaluation was required to address program-wide issues, there was a trade-off in the design between length and comprehensiveness. During pilot testing with DPI staff, the questionnaire was recognised as being lengthy, but there was reluctance to remove items, and a slightly lower response rate was anticipated as a result. Consequently, a number of people found the questionnaire difficult and too long. As with the list for interview, very few genuine external stakeholders responded to the questionnaire.

Goal Attainment Scaling – GAS was used in the questionnaire to measure achievement against expectation at the ORL program level. Considerable time and effort was invested in developing the GAS by URS, Steering Group members, and the ORL Management Team. While 60 people responded to the lengthy questionnaire, many fewer completed the final GAS section. Two reasons have been suggested for this. Firstly, the long and repetitive nature of the questionnaire induced fatigue and frustration in respondents, and many ‘gave up’ before the end. Secondly, the GAS section related to the ORL program as a whole, rather than individual projects. Many respondents, while familiar with a particular project or specific sub-project research component, were less familiar with ORL as a program, and were reluctant to attribute a score to program achievements against expectations, or their scores were a ‘guesstimate’.

Economic analysis – The economic assessment was originally intended to consist of Benefit- Cost analysis case studies, but information to inform this analysis was so poor that URS chose instead to limit the analysis to break-even rather than cost-benefit. In addition to providing an indicative measure of value for the case studies, the process brought the DPI researchers and economists together, which has not been standard practice in the past. Following presentation of the evaluation findings, many ORL project leaders expressed an appreciation for the value of economic analysis data to promote the achievements of their projects. It was identified that future program planning needs to incorporate a focus on cost, benefit and adoption information, and processes for collecting such information.

Role of the ORL Evaluation Officer – The ORL Evaluation Officer had management responsibility for the evaluation. She provided a single contact point for the consultants, liaising between them and ORL Management, the Steering Group, and ORL project teams. She provided pre-meeting briefings for URS, anticipating issues and creating response preparation time. This served to minimise delays, and allowed the evaluation to proceed as smoothly as possible. To ensure independence, the role of the ORL Evaluation Officer was restricted to advice only on evaluation theory and practice matters, never direction. This proved extremely difficult for an evaluation practitioner, but the benefits became obvious during Steering Group debriefs, when the level of independence of the evaluation was held in high regard.

Conclusion

The end-of-program evaluation of ORL was a valuable exercise for the DPI. Despite challenges discussed above – the like of which are to be expected for such a complex undertaking – DPI found the findings to be sensible, and many of the insights gained by the URS team to be quite helpful for future program planning.

References

- Bennett, C and Rockwell, K 1995, 'Targeting outcomes of programs (TOP): An integrated approach to planning and evaluation', unpublished manuscript, University of Nebraska, <http://citnews.unl.edu/TOP/english/> (accessed 23 May 2007).
- Chudleigh, P, Simpson, S and Schofield, N 2006, 'Economic evaluation of R&D investment in natural resource management', *Evaluation Journal of Australasia*, vol. 6, no. 2, pp 37-48.
- Glouberman, S and Zimmerman, B 2004, 'Complicated and complex systems: what would successful reform of Medicare look like?' *Changing Health Care in Canada: Romanow Papers*, vol. 2, pp. 21-53.
- Kiresuk, T, Smith, A and Cardillo, L 1994, *Goal Attainment Scaling: Applications, Theory and Measurement*, Lawrence Erlbaum Associates, New Jersey.
- Molas-Gallart, J and Davies, A 2006, 'Toward theory-led evaluation. The experience of European science, technology, and innovation policies', *American Journal of Evaluation*, vol. 27, no. 1, March, pp 64-82.
- OECD (Organisation for Economic Cooperation and Development) 2006, *Evaluation of Publicly Funded Research – Recent Trends and Perspectives*, Committee for Scientific and Technological Policy, OECD, Paris.
- Owen, J 2006, 'Managing the evaluation of multi-level multi-site programs', paper presented to the annual conference of the Australasian Evaluation Society, Darwin, September.
- Productivity Commission 2007, *Public Support for Science and Innovation*, Research Report, Productivity Commission, Canberra.
- Rogers, P 2007, 'Using programme theory for complex and complicated programmes', presentation to the South African Monitoring and Evaluation Society pre-conference workshop, Johannesburg.