Evaluation capacity building in response to the agricultural research impact agenda

Turner, James A.; Guesmi, Bouali; Gil, José M.; Heanue, Kevin; Sierra, Miguel; Percy, Helen; Bortagaray, Isabel; Chams, Nour; Milne, Cath

Published in:
Evaluation and Program Planning

DOI:
10.1016/j.evalprogplan.2022.102127

Print publication: 01/10/2022

Document Version
Peer reviewed version

Evaluation capacity building in response to the agricultural research impact agenda: Emerging insights from Ireland, Catalonia (Spain), New Zealand, and Uruguay

James A Turner[a,1], Bouali Guesmi[b], José M. Gil[b], Kevin Heanue[c], Miguel Sierra[d], Helen Percy[a], Isabel Bortagaray[d,e], Nour Chams[b], Cath Milne[f],

[a] Farms Systems and Environment, AgResearch, 10 Bisley Rd, Hamilton 3214, New Zealand. james.turner@agresearch.co.nz

[a] Adoption and Practice Change, AgResearch, 10 Bisley Rd, Hamilton 3214, New Zealand. helen.percy@agresearch.co.nz

[b] Center for Agro-Food Economics and Development (CREDA-UPC-IRTA). Parc Mediterrani de la Tecnologia, Edifici ESAB, 08860 Castelldefels (Barcelona), Spain.

bouali.guesmi@upc.edu ; chema.gil@upc.edu; nour.chams@upc.edu

[c] Evaluation Unit, Teagasc, Mellows Campus, Athenry, Co. Galway, H65 R718, Ireland, kevin.heanue@teagasc.ie

[d] Instituto Nacional de Investigación Agropecuaria (INIA). Edificio Los Guayabos - Parque Tecnológico LATU, Av. Italia 6201, 11500, Montevideo, Uruguay. Innovation and Communication Management, msierra@inia.org.uy

[e] Institute for Sustainable Development, Innovation and Social Inclusion, University of the Republic, Uruguay, isabelbortagaray@gmail.com

[f] formerly of SRUC, Peter Wilson Building, West Mains Road, Edinburgh, EH9 3JG UK, cemilne@live.com

[1] corresponding author

Abstract

Performance-based funding and calls for public-funded science to demonstrate societal impact are encouraging public research organisations to evaluate impact, the so-called impact agenda. This paper explores evaluation methods of four fully or partially public-funded agricultural research organisations and how they are building evaluative capacity to respond to the impact agenda. Drawing on cross-organisational comparison of the readiness of each organisation to implement evaluation, the implications for improving evaluative capacity building (ECB) are discussed. This study extends the current literature on ECB, as very little has focussed on research organisations in general, and particularly agricultural research.

Driven by the impact agenda, the organisations are beginning to emphasise summative evaluation. Organisational leaders valuing the demonstration of impact and commitment to building evaluation capacity are important precursors to other aspects of organisational readiness to implement evaluation. However, organisational emphasis remains on using evaluation for accountability and to improve efficiency and allocation of funding. The organisations have yet to systematically embed evaluation processes and capabilities for learning at programme and organisation-levels. There is, therefore, an opportunity to develop organisation and programme-level evaluation processes that inform each other and the pathways to impact from science.

To realise this opportunity, organisations could strengthen internal and external networks of evaluation practitioners and academics to bridge the gap between the theory and practice of monitoring and evaluation for learning (MEL) and to begin to reshape organisational culture by using evaluation methods that are grounded in co-production and integrated scientific and societal values.

Keywords: agricultural research impact; monitoring and evaluation, evaluation capacity building

Highlights

- Explores evaluative capacity building by public agricultural research organisations
- Identifies precursors to ECB – commitment to impact, valued by leadership and support networks
- Evaluative capacity is limited, from programme to organisation-levels
- Evaluation for learning to increase impact is limited

Introduction

Over the past two decades there has been an increasing call for public-funded science to deliver and demonstrate impact (Penfield et al., 2014), going beyond the excellence of academic outputs to the benefits these produce for society (Donovan, 2011). This is challenging the social contract with science through which scientific freedom was exchanged for the expectation of socially beneficial impacts (Joly et al., 2016; Owen et al., 2012). Simultaneously, the diffusion of new public management rules (Gaunand et al., 2015) has encouraged performance-based funding for public-funded science (New Zealand Ministry of Business, Innovation and Employment, 2015; Organisation for Economic Co-operation and Development, 2010).

Delivering and demonstrating impact is especially important for agricultural science (Midmore, 2017), with agriculture at the nexus of significant societal challenges, such as food security, rural resilience, and environmental sustainability (Tilman et al., 2002). Given the nature of these challenges agricultural science needs to deliver social, cultural, and environmental benefits (Donovan, 2011; Kelley et al., 2008), as well as economic (Joly et al., 2015). For the purposes of this study, research impact is defined as: “the demonstrable contribution that research makes to the economy, society, culture, national security, public policy or services, health, the environment, or quality of life, beyond contributions to academia” (Australian Research Council cited in Midmore (2017)).

The reasons research organisations may evaluate science impact can be organised into the four "As": advocacy, accountability, allocation, and analysis (Morgan Jones, Manville, and Chataway 2017; Penfield et al., 2014):

1. Advocacy to demonstrate the value of research to government, stakeholders, and the public (Joly et al., 2015; Penfield et al., 2014).
2. Accountability to funders, and more broadly society, by monitoring and managing research organisation performance to contribute to society (Penfield et al., 2014).
3. To inform funding allocation, based on the potential value of research (Penfield et al., 2014).
4. Analysis to understand how agricultural science generated contributes to changes in practices and policies in the agri-food system (Joly et al., 2016; Midmore, 2017; Pollock, 2012).

The "impact agenda" therefore calls for both summative (advocacy and accountability) and formative (allocation and analysis) functions of evaluation (Joly et al., 2015). Formative evaluation takes place during the development of a programme, with the intention of improving the value or effectiveness of the programme, while summative draws lessons from a completed programme (Superu, 2017a).

While agricultural research organisations are in the emergent stages of meeting the impact agenda, other research organisations have made progress (e.g. CSIRO, 2015; Joly et al., 2016; Kelley et al., 2008; Maredia et al., 2014; Menon et al., 2009). The Consultative Group on International Agricultural Research (CGIAR) has been undertaking ex-post Impact Assessments since the early 1970s (Watts et al., 2008), including non-economic impacts (Kelley et al., 2008). Utilising a suite of evaluation methods (Kelley and Gregersen, 2003), CGIAR has moved from an accountability focus to evaluation for learning (like analysis) (Watts et al., 2008) and includes participation of stakeholders and organisations in evaluation (Mackay and Horton, 2003; Watts et al., 2008; Mayne and Johnson, 2015).

Research organisations working to meet the impact agenda face several challenges including: evaluation of multiple possible impact pathways (Joly et al., 2016); multiple impacts (Bozeman and Sarewitz, 2011; Kelley et al., 2008) at different levels from programme to organisation; attributing

97 impacts in multi-actor innovation systems (Douthwaite and Hoffecker, 2017; Horton and Mackay,
98 2003; Joly et al., 2016); and time-lags between research outputs and impacts (Kelley et al., 2008). To
99 address these challenges agricultural research organisations have begun to invest in evaluation
100 capacity building (ECB) (Joly et al., 2016; Maredia et al., 2014; Stone-Jovicich et al., 2019), which is
101 defined as "the intentional work to continuously create and sustain overall organisational processes
102 that make quality evaluation and its use routine" (Hueftle Stockdill et al., 2002, p.14). The goals of ECB
103 include designing, undertaking, and managing evaluation projects; developing and using evaluation
104 knowledge and skills; encouraging organisational learning as an ongoing process; organisational
105 accountability; creating awareness and support for evaluation within the organisation (Cousins et al.,
106 2014; Preskill and Boyle, 2008); and sustaining evaluation practice over time (Wade & Kallemeyn,
107 2020).

108 However, to what extent, and how, are public-funded agricultural research organisations building
109 evaluation capacity to respond to the impact agenda? To address this question our aim is to describe
110 the current evaluative capacity of four partially to fully public-funded agricultural research
111 organisations in terms of:

112 1. The evaluation methods these organisations are using; and
113 2. Their comparative evaluation capacity.

114 This article proposes to extend the current body of literature on ECB, as very little has focussed on
115 research organisations in general, and agricultural research in particular (Joly et al., 2016). As far as
116 we are aware, this is the first time that four publicly funded research organisations have documented
117 their approaches to ECB. Drawing on our cross-organisational comparison of evaluation capacity,
118 implications are discussed to improve ECB within agricultural research organisations.

119

120

Background

This section provides a brief background to each of the four research organisations that participated in this study. The four organisations established a learning network on evaluation capacity building that was operationalised with workshops in 2017 and 2019 that provided data for the analysis in this paper (see Methodology). The network was based on the organisation's commonalities, particularly public research organisations in the agri-food sector focusing on pastoral systems.

Each of the organisations is described in terms of their purpose, agri-food sectors served, key research and extension capabilities, major funding sources (Table 1), overarching evaluation frameworks and core principles, key characteristics of their organisational impact agenda (Table 2), and timeline of activities to build evaluative capacity (Figures 1 to 4). The description of the organisations is for the period 2015-2018. Given the dynamic nature of each organisation the information presented here will differ from the current reality.

Table 1. Major funding sources for each case study organisation – average for 2015-2018

Income Source	Contribution to the total (%)			
	AgResearch	INIA	IRTA	Teagasc
Structural[1] Funds[1]	27.0	84.0	32.0	69.0
Competitive[2]	17.6	5.0	12.0	17.0
Contracts for services and product sales[3]	44.1	10.0	39.0	14.0
Subsidies[4]	0.0	0.0	7.0	0.0
Other funds	11.2	1.0	10.0	0.0

[1]Funding allocated directly to the organisation to then allocate, [2] Funds that are bid to a funder for, [3] One off contracted projects, e.g. for an agricultural company, and [4] Funds that are received from the provincial councils and for funding researchers from national and regional programs. The availability

of structural research funding from the state depends on the organisational and funding characteristics of each country's science system.

Teagasc, Ireland

Teagasc – the Agriculture and Food Development Authority – is the national body providing integrated research, advisory and education services to the agriculture and food industry and rural communities in Ireland. It is a key node in Ireland's Agricultural Knowledge & Innovation System (AKIS) and has long-established and strong collaborations with many other public and private sector AKIS actors. Teagasc has 1,260 staff based in seven research centres, 52 advisory offices and four colleges of further education around the country.

An Evaluation Unit was established in Teagasc in 2003 in response to recommendations made by the Comptroller and Auditor General (C&AG) in a 1999 report on Performance Measurement in Teagasc (see Figure 4). In 2002, consultants (Capita) outlined a strategy and plan for establishing an evaluation function across Teagasc, not just research evaluation. Although no formal Evaluation Unit existed in the organisation before 2003, research and advisory programmes were reviewed, and other organisational functions evaluated, over the years since Teagasc was established in 1988.

After 2003, a regular cycle of peer assessments of research centres, then research programmes (established in 2008) and advisory regions (established in 2013) began. Organisational functions were evaluated on a non-cyclical basis. In 2014, Teagasc approached the Department of Education & Skills to develop an evaluation framework for its colleges. Externally, in 2016, the Code of Practice for the Governance of State Bodies issued by the Department of Public Expenditure and Reform led to the Oversight and Performance Delivery Agreement (2017 – 2020) with the Department of Agriculture, Food and the Marine (DAFM). This Agreement formalises the arrangements between DAFM and Teagasc for oversight and reporting on performance. Internally, a series of revisions of evaluation

protocols, framework, methodologies and networks has shaped the evaluation function from 2017 to 2021.

Teagasc uses agricultural innovation systems as an overarching framework for understanding the organisation's contribution to the transformation of Ireland's AKIS. To evaluate its research programmes and advisory service Teagasc uses ex-post outcome assessment and peer assessments (Table 2, Figure 1). Key principles that guide evaluation activities in Teagasc are:

1. Using M&E for accountability and learning.
2. Incorporating stakeholder participation.
3. Undertaking M&E of economic, social, and environmental impacts.
4. A focus on improvement of service design and delivery.
5. Strengthening organisational governance.

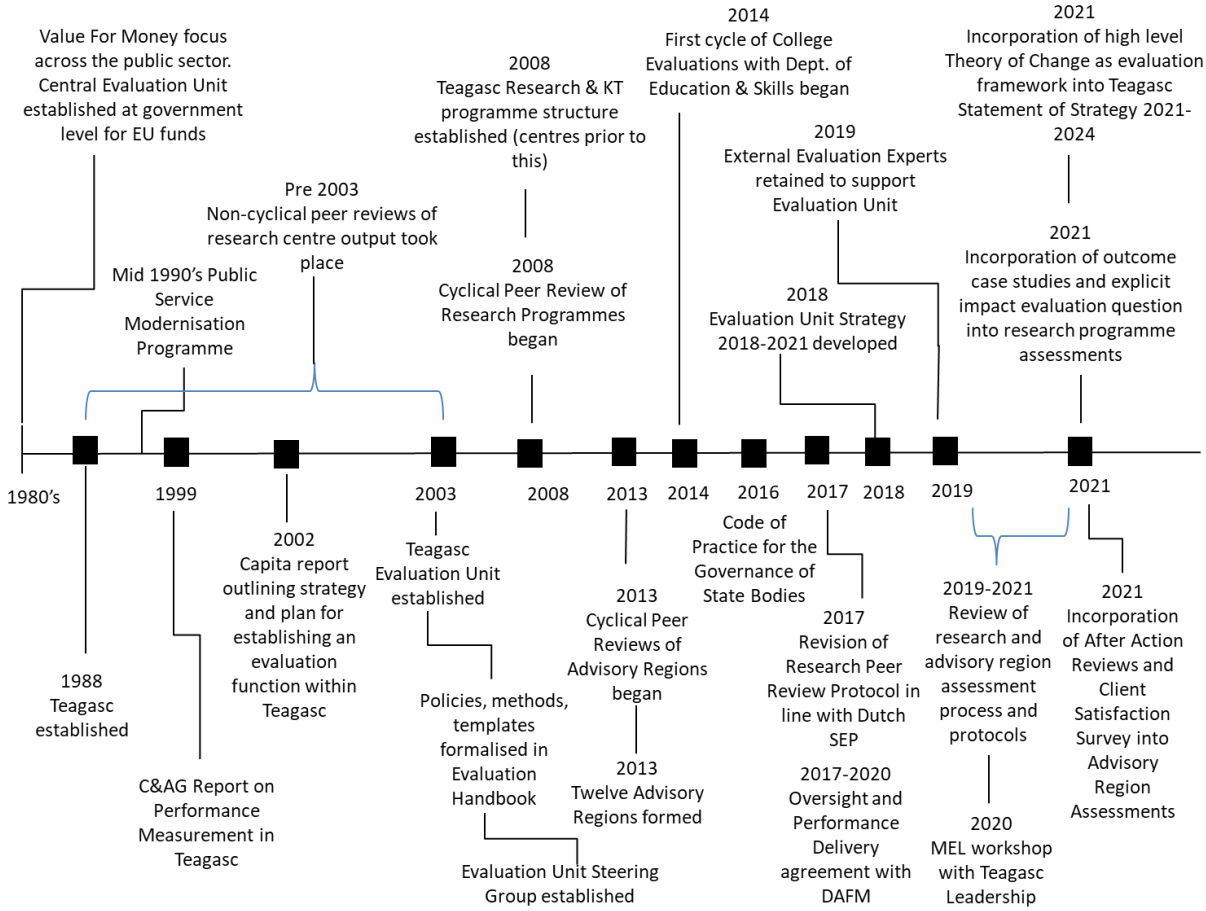


Figure 1: Timeline of key external and organisational events in building evaluation capacity in Teagasc.

AgResearch Ltd., New Zealand

AgResearch is one of seven Crown Research Institutes (CRI), Government-owned organisations tasked with providing research and technologies to deliver outcomes for New Zealand. AgResearch focuses on the pastoral and agri-food sectors, "*to enhance the value, productivity and profitability of New Zealand's pastoral, agri-food and agri-technology sector value chains to contribute to economic growth and beneficial environmental and social outcomes for New Zealand*." (AgResearch, 2018).

AgResearch achieves this by providing research and the transfer of knowledge and technologies in partnership with key Māori (the indigenous peoples of New Zealand), industry and Government partners (AgResearch, 2020). AgResearch has 286 scientists and 178 technicians (2020), but does not have staff dedicated to technology transfer. Instead, AgResearch partners with stakeholders with these capabilities. .

In 2010 the CRI Taskforce Review (Crown Research Institute Taskforce, 2010) (Figure 2), recommended changes to encourage realising impact from science. Recommendations included, improve partnerships between research organisations and industry, increase internal funding managed by each CRI to align with sector needs, and increase each CRI's accountability for impact (Turner et al., 2013). To encourage the last each CRI reports on the percentage of relevant end-users who have adopted their knowledge and technology and provides annual impact case-studies (AgResearch, 2020) (Table 2).

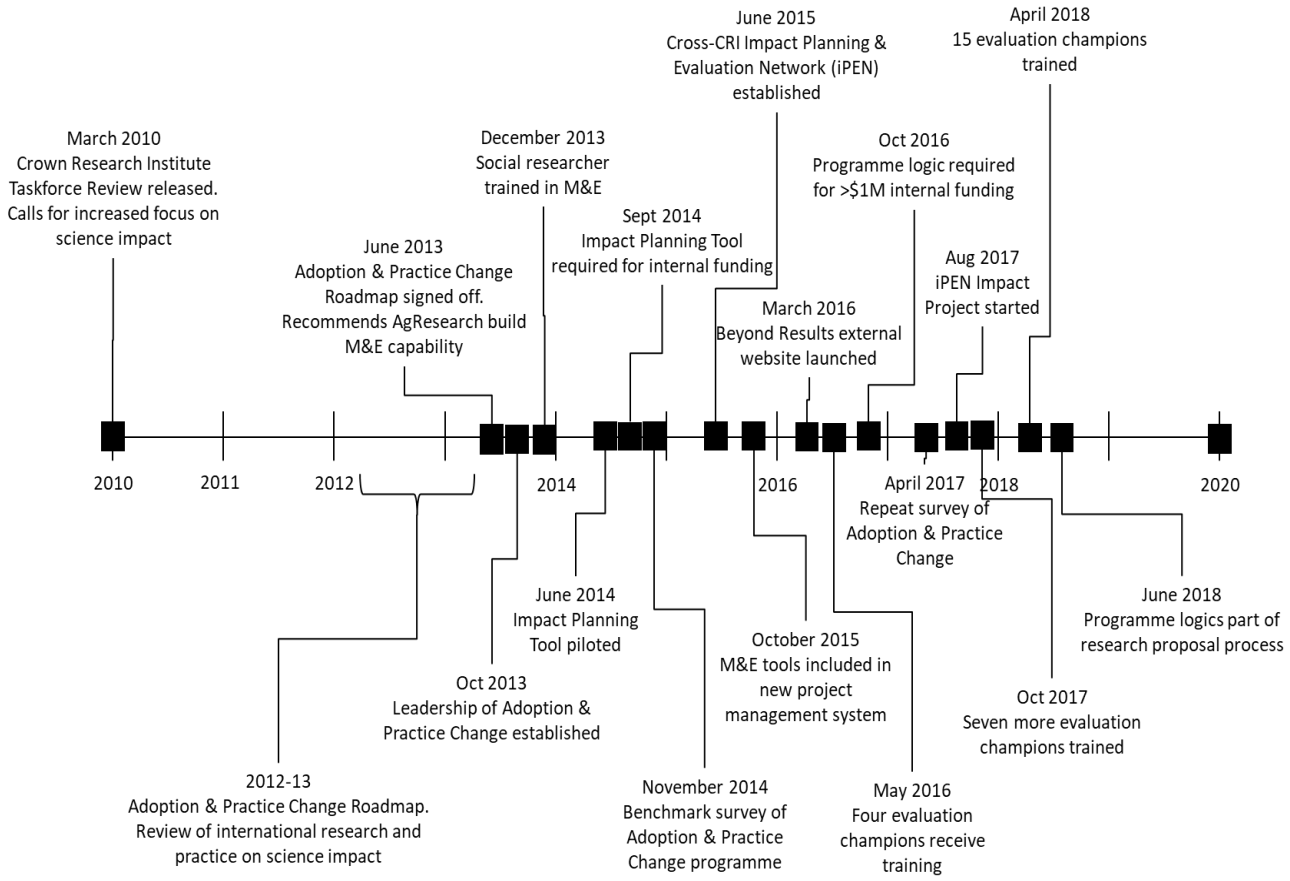


Figure 2: Timeline of key external and organisational events in building evaluation capacity in AgResearch.

In response to the CRI Taskforce Review recommendations the AgResearch Adoption and Practice Change Roadmap (2012-13) developed organisational recommendations to increase science impact, including:

1. Partnering with stakeholders to deliver research outcomes.
3. Planning science programmes with stakeholders to identify issues, outcomes and get a clear understanding of who the stakeholders will be and their roles in achieving impact (Impact Planning Tool).
4. Monitoring and evaluating progress towards outcomes and impact within research programmes (Percy et al., 2015) for accountability to funders and to support learning and

adaptation of research activities (Botha et al., 2017), by requiring Impact Planning Tool and programme logics for internal and Government funded programmes.

5. Embedding evaluation within the organisation, rather than through external evaluators or a separate evaluation unit, by training internal Evaluation Champions, and making monitoring and evaluation tools available on Beyond Results website[1] and included in a new organisational project management system.

Instituto Nacional de Investigación Agropecuaria (INIA), Uruguay

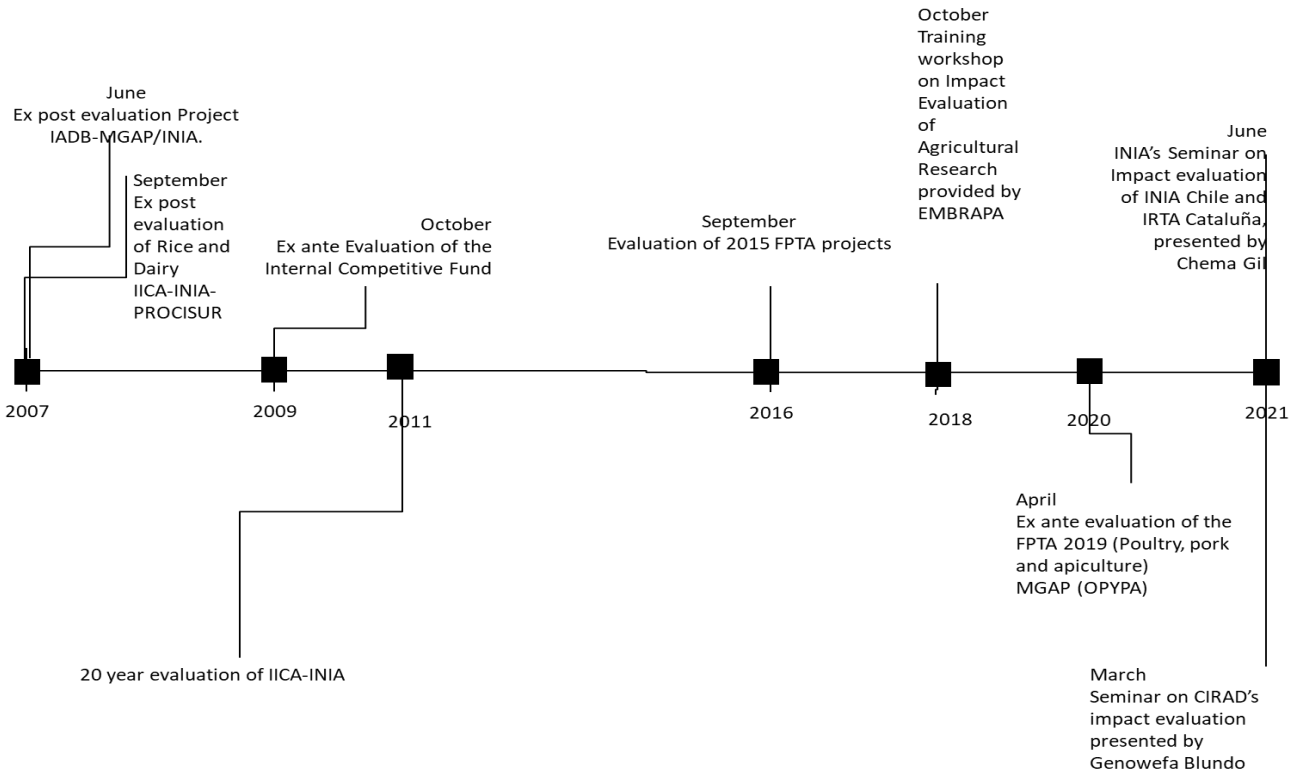
The purpose of INIA is to "produce and adapt knowledge and technologies to contribute to the sustainable development of Uruguay and the agricultural sector, considering State policies, social inclusion, and market and consumer demands". To achieve this, INIA is committed to research, science and technology, and collaborates in the design of public policies for the development of the agricultural sector. Furthermore, through publications and technological developments, it aims at diffusing and transferring scientific knowledge to the agricultural sector and society. INIA's research activities are particularly focused on livestock, agriculture, dairy, forestry, and plant production from extensive family farming through to intensive production, as well as integrated crop-livestock production systems. INIA has 565 permanent staff, 174 university staff, of which 121 are effective researchers, and 391 support staff. It has five experimental stations distributed throughout the country and a mixed unit with the Pasteur Institute in Montevideo.

INIA has created a system of evaluation certification of technologies with participation of external actors in order to improve the adoption and the impact of the technologies (Vasen et al., 2021).

[1] https://www.beyondresults.co.nz/

The private sector provides the main source of structural funding through a tax on the sale of agricultural and livestock assets and Government agricultural departments. This amount represents approximately 82% of the funding managed by INIA (Table 1).

Research evaluation, and impact evaluation in particular, have been an area of concern, and, as seen in the timeline (Figure 3)[2], a series of activities have been implemented to advance this. An institutional assessment was carried out in 2011, and it has recently strengthened its Planning, Monitoring and Evaluation capacity by having a staff member dedicated to M&E. Although different innovation perspectives coexist internally, co-innovation is becoming increasingly important. At INIA co-innovation involves the participation of relevant stakeholders in all phases (design, implementation, and evaluation) with the goal of achieving greater adoption and adaptation of technologies, as well as a greater impact.



[2] The timeline is based on Costa (2022).

Figure 3: Timeline of key external and organisational events in building evaluation capacity in INIA-Uruguay[3].

Institute for Food and Agricultural Research and Technology (IRTA), Catalonia (Spain)

IRTA is a research institute of the Government of Catalonia (Spain) whose purpose is to contribute to the modernisation, improvement and promotion of competitiveness and sustainable development in the agriculture, food, and aquatic sectors, providing safe and quality foods to the final consumer, and contributing to the global improvement of human welfare. IRTA is the leading organisation for agri-food applied research in Catalonia. Aligned with the Catalan/State/EU strategy, IRTA must address the main challenges of the agri-food sector thus generating new scientific knowledge and transferring it to producers and industry to generate innovation.

IRTA's activities are to promote research and technological development in agri-food, facilitate the transfer of scientific findings, and increase its own technological advances while maximising public and private sector coordination. In the spirit of collaboration as the most efficient way to strengthen capacities and improve societal impact, IRTA established and consolidated a network of partnerships with public and private institutions. IRTA is economically supported by its own funds from royalties and other sources, as well as by regional, state, and international funding (Table 1). Research and development activities are conducted in ten research centres and eight experimental stations

[3] IADB-MGAP/INIA - Banco Inter-Americano de Desarrollo. Ministerio de Ganadería, Agricultura y Pesca. Instituto Nacional de Investigación Agropecuaria. IDB-MGAP/INIA - Inter-American Development Bank. Ministry of Livestock, Agriculture and Fisheries. National Institute of Agricultural Research. IICA-INIA-PROCISUR - Instituto Interamericano de Cooperación para la Agricultura. Instituto Nacional de Investigación Agropecuaria. Programa Cooperativo para el Desarrollo Tecnológico Agroalimentario y Agroindustrial del Cono Sur. Inter-American Institute for Cooperation on Agriculture. National Institute of Agricultural Research. Cooperative Program for the Technological Development of Agrofood and Agroindustrial of the Southern Cone. FPTA projects - Proyectos financiados con el Fondo de Promoción de Tecnología Agropecuaria. Projects financed with the Fund for the Promotion of Agricultural Technology. EMBRAPA - Brazilian Agricultural Research Corporation. Empresa brasileña de investigación agropecuaria. MGAP (OPYPA) - Ministerio de Ganadería, Agricultura y Pesca (Oficina de Programación y Política Agropecuaria). Ministry of Livestock, Agriculture and Fisheries (Agricultural Programming and Policy Office).

throughout Catalonia, as well as two other centres in association with universities and other organisations. IRTA employs approximately 700 staff.

IRTA's evaluation of research is based on ex-post economic impact assessment and social returns approaches to consider multiple impacts of the organisation's research (Table 2, Figure 4). The impact on Catalan agricultural productivity of the research conducted by IRTA has recently been analysed for the period 1985-2015. The main conclusions of the study (Guesmi and Gil, 2021) were:

1. IRTA plays a relevant role in strengthening the system of agricultural technology in Catalonia.
2. Elasticity of Total Factor Productivity[4] with respect to the public knowledge stock was approximately 0.15, indicating that a one per cent increase in the public R&D knowledge is likely to lead to a 0.15 per cent increase in agricultural productivity.
3. The social rate of return on investment[5] in IRTA ranges from 15% to 28%, depending on assumed lag times and discount rate.

[4] Total factor productivity represents the part of output growth that cannot be explained by the growth of inputs used in the production. It is measured as the ratio of total output to total inputs.
[5] The social rate of return could be defined as a percent return on each currency unit spent on R&D investments. It provides an estimate of the benefits from a one-off increase in R&D expenditures, which could be useful as an ex-post measure of the returns achieved and an ex-ante tool to assist in resource allocation (Sheng et al., 2011).

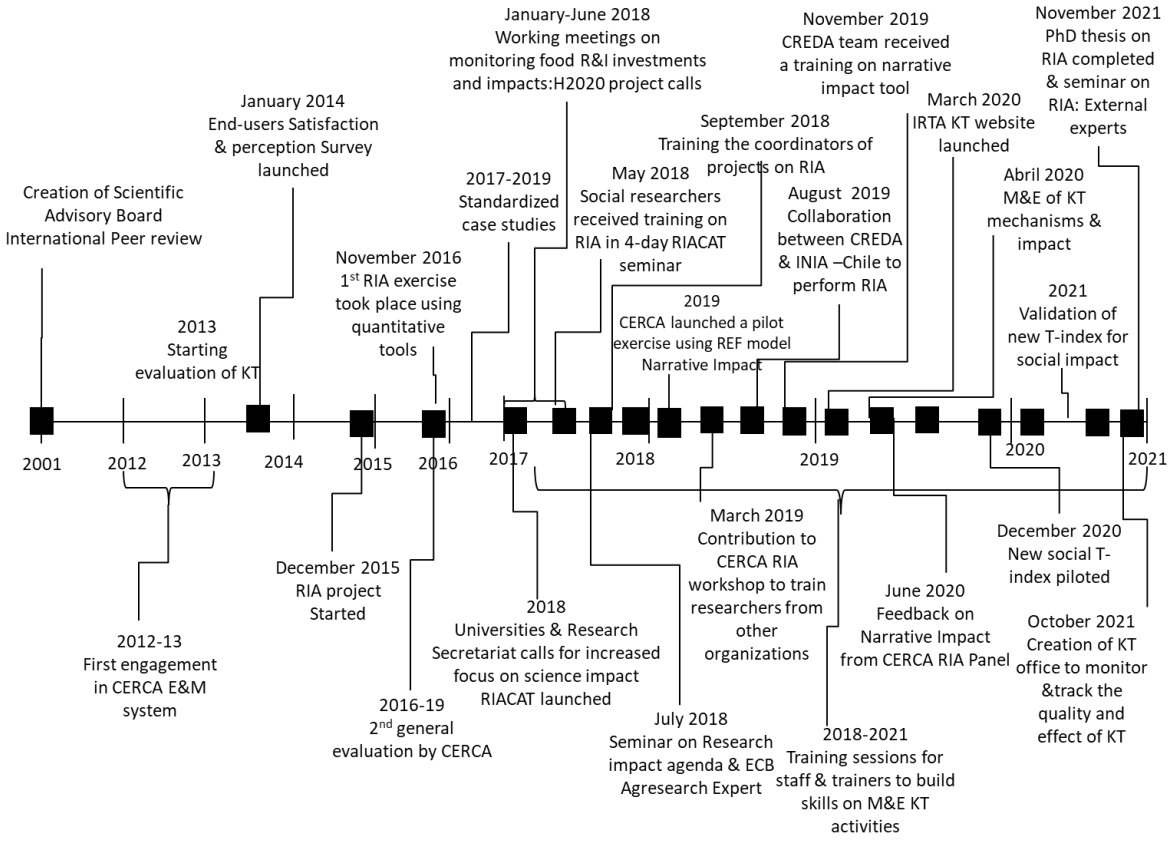


Figure 4: Timeline of key external and organisational events in building evaluation capacity in IRTA.

284 **Table 2**: Characteristics of the agricultural research impact agenda in each organisation – definition, weighting and incentives for an impact focus

Characteristics	AgResearch	INIA	IRTA	Teagasc
Definition of impact	Improved productivity, profitability and value from pasture-based production while reducing environmental footprint.	Contribute to the sustainable development of Uruguay and the agricultural sector considering State policies, social inclusion, and market and consumer demands, strengthening the linkages between research and the productive sector.	Modernisation, improvement, and promotion of competitiveness and sustainable development in the agriculture, food and aquatic sectors.	Underpin profitability, competitiveness and sustainability of agri-food sector and bioeconomy.
Weighting of impact in research assessment	50% in ex-ante assessment of public research funding to show potential benefits to New Zealand (economic, environmental and social).	Almost all efforts are oriented to research monitoring. Ex-ante evaluation and impact are not yet established practices at INIA.	Depends on the type of project. High weights attributed to research related to business activities.	50% in ex-post outcome assessment.
Funder mechanisms for encouraging an impact focus	Ministry of Business, Innovation and Employment.(MBIE) competitive funds require ex-ante estimate of impact and credible impact pathway in funding proposals. Funding contracts with MBIE competitive funds include an outcome statement against which projects are annually assessed. Impact case studies (ex-post) are required as part of annual reporting on all AgResearch science to MBIE. AgResearch provides a selection of annual case-studies as examples of economic, social, and environmental impact.	INIA's last institutional impact evaluation carried out in 2010, through IICA and a multidisciplinary team. In 2015 INIA's Fund for the Promotion of Agricultural Technology (FPTA) had an impact evaluation process with institutional funding.	Evaluation against outcomes (indicators): adopted 10 years ago, though it is no longer used.	Annual reporting to Department of Agriculture, Food and the Marine on Teagasc's performance in delivering on its goals/objectives as set out in its multi-annual Statement of Strategy and the associated annual High Level Business Plan. Focus on outcome evaluation. Ex-ante impact pathway.

Sources	(AgResearch, 2020; Ministry of Business, 2015, 2017)	www.inia.org.uy/online/site/967882I1.php www.iica.int/es/prensa/noticias/iica-evalua-20-anos-de-investigacion-de-inia-uruguay	(IRTA, 2017)	Teagasc (2017)

285

The next section presents the study methodology, defining evaluative capacity building and the analytical framework for comparing the types and uses of evaluation, and evaluative capacities. The results of the comparative analysis are then presented and discussed. The paper finishes with key conclusions regarding how public-funded agricultural research organisations can build evaluation capacity.

Methodology

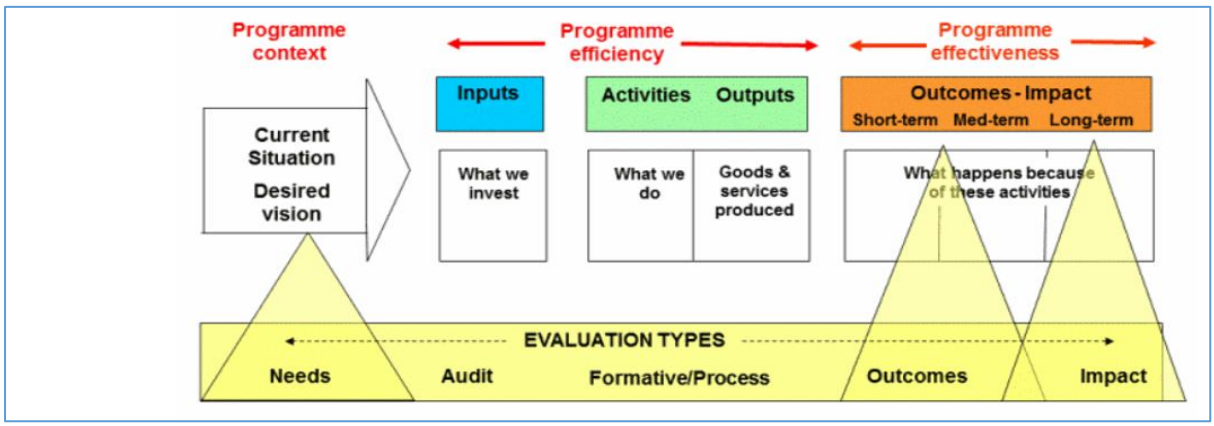
The Evaluation Capacity Building (ECB) literature points to the need to increase the capacity within the individual, organisation, and wider science system to do and use evaluation (Bourgeois & Cousins, 2013; Chaplowe & Cousins, 2016). However, as every organisation works and operates within different contexts there is no single approach to building evaluative capacity (Bourgeois et al., 2015). White, Percy & Small (2018) suggest using a hybrid model of ECB based on the multi-disciplinary model of Preskill & Boyle (2008) and Cousins, Goh, Elliott & Bourgeois (2014) as a useful foundation for science organisations building an ECB culture. This aligns an evaluative culture of "*doing"* that includes such things as organisational support systems, structure, and leadership; capacity to do evaluation within the organisation; and capacity to do evaluation within programmes, with a culture of "*using and valuing"* that involves organisational learning capacity; programme and individual learning capacity; and capacity to use the evaluation by the organisation and within programmes. This hybrid model formed the basis of the analytical framework described below.

Analytical framework

For the purposes of this study, evaluation is defined as the structured interpretation of predicted or actual impacts of proposals, organisations, programmes, projects, or individuals (Taylor-Powell & Henert 2008). It looks at original objectives, and at what is either predicted, or what was accomplished and how it was accomplished. To understand the types of evaluation in each organisation, we estimated the allocation of funding to internal and external evaluations, and the proportion of evaluation undertaken as (Taylor-Powell & Henert 2008; Superu, 2017b):

1. Needs – Characterisation of the needs and priorities of the population targeted by a programme, as well as the potential barriers and enablers to the programme meeting these.
2. Audit – Evaluation of a programme for the purpose of verifying adherence to a set of pre-defined processes.
3. Formative (Process) – Method of assessing how a programme is being implemented. Process evaluation focuses on the programme's operations, implementation, and service delivery.
4. Outcome – Focuses on the effectiveness of the programme and its outcomes. Outcomes can be produced in the short or long term.
5. Impact – Assesses programme effectiveness in achieving its ultimate goals. This is done several years after the programme is implemented.

Figure 5: Types of evaluation.



Source: Taylor-Powell & Henert (2008)

To understand why each organisation undertakes evaluation, the frequency of evaluation for each of the four "As" was estimated using a four-point Likert scale (never/occasionally/frequently/always):

- Allocation – Evaluation is used to instigate changes within a project or programme.
- Advocacy – Evaluation is used to support a position or to justify action or inaction.
- Analysis – Evaluation is used to learn about the programme, its staff, its operation, or outcome.

- Accountability – The process of evaluation is used to support programme management and engagement (e.g. use of programme logics, Theory of Change, Evaluation Plans etc.).

To consider the readiness of each organisation to implement evaluation, questions adapted from the Checklist for Organisational Readiness for Evaluation (Superu, 2017b) were used (Table 3). For each of the questions in Table 3 readiness was assessed as:

- Emerging: There is very little or no evidence of this in the organisation.
- Developing: there is some evidence of this in the organisation – it happens sometimes but is not embedded in routines.
- Consolidating: This is seen occurring regularly in the organisation, however, not everyone is on board.
- Highly developed: This is evident at all levels of the organisation and appears strongly embedded in the organisation's governance, leadership, system, structures, and practices.

Table 3: Questions used to consider organisational readiness for evaluation.

Our organisation uses evaluation to be accountable to our funders and/or internal and external stakeholders
Our organisation uses evaluation for learning and reviewing to improve the ways our organisation delivers impact from science and/or investments in science
Our organisation understands the value of demonstrating the difference we have made from science for our stakeholders
Leaders and/or managers are committed to developing evaluation capacity
Systematic M&E occurs at all levels of our organisation (if not at all levels, specify which, e.g. project-level, organisation-level)
Our organisation has the systems and processes in place to collect data for evaluation
Our organisation has the skills and knowledge to collect data for evaluation
In our organisation, we have the skills and knowledge we need to understand and use data required for evaluation
In our organisation, staff have time, funds and/or resources to collect and use data required for evaluation
In our organisation we (decision-makers and staff) have opportunities to involve a wide range of people (internal and external to our organisation) in evaluation
In our organisation, we participate in external networks of learning to increase our knowledge of evaluation

Source: Superu (2017b)

Data collection and analysis

Two workshops were held on November 2017 and February 2019 at Teagasc (Carlow, Ireland) and IRTA-CREDA (Barcelona, Spain), respectively. At the first workshop there were three participants from Teagasc, two from CREDA-UPC (who provided information on IRTA), one from INIA, and two from AgResearch. At the second workshop there was one participant from Teagasc, three from CREDA-UPC (who again provided information on IRTA), one from AgResearch, one from Scotland, and one from Udelar (University of the Republic) (who participated virtually and provided information on INIA). In the second workshop participants reviewed and updated data collected in the first workshop. All the workshop participants are tasked with undertaking evaluation within their organisations, as well as one participant each from Teagasc, INIA and AgResearch being tasked with building evaluation capacity in their organisations.

At the 2017 workshop the participants mapped the reasons evaluation is undertaken, the types of evaluation, approaches and activities, and actions to build evaluation capacity (see Analytical Framework). The selection of questions (Table 3) were circulated prior to the workshop so that workshop participants could consult within their organisation to provide an assessment of organisational readiness.

At the workshop each organisation used a colour-coded dot to show where they were positioned (Table 6) and then a facilitated discussion was held around the gaps and trends emerging. After the workshop the participants undertook verification with key staff in their respective organisations.

Having identified the types of evaluation and evaluative capacities, workshop participants then identified the initiatives, strategies, actions, and key resources each organisation engaged in to support evaluation at the stages of planning, monitoring and communication.

Findings and Discussion

Purposes for evaluating research

The four organisations undertook evaluation for all four purposes, though with different frequency (Table 4). Evaluation for allocation purposes ranged from always in Teagasc, to frequently in INIA, to occasionally in the other organisations. Teagasc has more initiatives in place to assist with being able to make changes within a project or programme, including a formal cyclical peer review system of research programme objectives and evaluation of knowledge transfer services. AgResearch uses its annual programme reporting and funding processes to reflect on programme contributions to the Science Plan and identify changes to allocation of funding to different research areas. IRTA also use evaluation to allocate resources to different research programmes based on ex-ante estimation of the potential impact of research. INIA are looking to increase the use of evaluation for informing funding allocation at the programme-level.

Accountability is traditionally a strong driver for undertaking evaluation (Midmore 2017; Penfield et al., 2014), and agricultural research organisations are no different. INIA, Teagasc and Agresearch had strong accountability drivers. In contrast, the remaining organisation identified that while evaluation for accountability was occurring, it was not well embedded throughout the organisation. For example, through its Annual Reports to the New Zealand Government, AgResearch reports on organisational performance against Key Performance Indicators (KPIs), such as commercial revenue, science publications per scientist, and understanding of stakeholder strategies. Another example is INIA which intends to hold a workshop with producer associations and the government, at the end of its five-year strategic plan, to evaluate performance in seven production systems against the objectives and goals proposed at the beginning of the plan. In the case of Teagasc, through its oversight agreement with DAFM, it provides an evaluation of its performance in delivering on its goals/objectives as set out in its multi-annual Statement of Strategy and the associated annual High Level Business Plan for the period under review.

Another significant driver for evaluation is learning (or analysis) (Midmore 2017; Pollock 2012; Wade & Kallemeyn, 2020). While this driver was evident within Teagasc, AgResearch, and INIA, IRTA is at the

very early stages of realising organisationally the potential for evaluation to improve their science programmes through evaluation for analysis (Table 4). To support this use of evaluation AgResearch is building evaluation into all major research programmes; though this is at an early stage (Table 8).

Table 4: Frequency with which evaluation is used in each research organisation for the four purposes.

Purposes for evaluating research	Never	Occasionally	Frequently	Always
Allocation		AgResearch IRTA	INIA	Teagasc
Advocacy		Teagasc IRTA	AgResearch INIA	
Analysis		IRTA INIA	AgResearch Teagasc	
Accountability		IRTA	AgResearch Teagasc INIA	

Since this research was undertaken the four "As" have been extended by Parks et al. (2019) to include acclaim (comparing and recognising the value of higher education institutions and the research conducted within them) and adaptation (to steer change in organisational structures, behaviours and cultures, and research activities and priorities). Future research should consider the evaluative capacities needed to support evaluation for acclaim and adaptation.

Internal and external evaluation

The four organisations have different balances of evaluation undertaken by internal and external parties (Table 5). For IRTA and AgResearch, evaluation is predominantly undertaken internally. In contrast, evaluation is equally split between external and internal in Teagasc, and in the case of INIA is predominantly undertaken externally.

Table 5: Estimated proportion of internal and external evaluation undertaken within each organisation

Organisation	% undertaken internally	% undertaken externally
AgResearch	~ 85	~ 15
INIA	40	60
IRTA	90	10
Teagasc	50	50

Teagasc and INIA evaluations always involve an external provider in collaboration with internal staff. These organisations also use external peer review panels chosen to represent the different areas of the programme being evaluated. Using external providers and review panels enables evaluations to be completed more rapidly and the evaluation is considered more impartial. This emphasis on external evaluation may be related to Teagasc and INIA both emphasising the use of evaluation for accountability purposes (Table 4).

Types of evaluation

For most of the organisations, evaluation is concentrated on internal research performance: needs, auditing, and formative evaluation (Table 6 and Figure 5). INIA, IRTA, and AgResearch are all undertaking similar proportions of evaluation in these areas. For example, AgResearch reports on milestones, deliverables, and case studies of research programmes in its Annual Reporting. The emphasis on formative (or process) evaluation, relative to summative (or outcome) evaluation, may be due to impacts occurring post-programme and the difficulty of allowing for the time lag to when impacts occur, and lack of investment in evaluation of impacts post-project.

There appears to be less analysis of changes in the social, economic and production context of stakeholders to analyse research effectiveness and changes beyond the research organisations. Evaluation of impact is, therefore, limited in all organisations. However, the organisations are beginning to increase their emphasis on summative evaluation. Teagasc is strategically moving into

formative and summative evaluation and INIA recently incorporated the evaluation of outcomes and the goal is to strengthen the evaluation of impact in the future.

Table 6: Proportion and examples of types of evaluation being done in each organisation.

Types of evaluation	AgResearch	INIA	IRTA	Teagasc
Needs	**30%** Stakeholder participation in programme logic development; participatory project development with stakeholders.	**20%** 5-year strategic plan defining research areas and technology transfer activities.	**25%** 2020-2023 Strategic Plan.	**1%** Stakeholder consultation on research programmes and advisory services.
Audit	**25%** Reporting on programme milestones and deliverables usually to meet requirements of funders. Organisational performance is annually reported on against KPIs for annual reporting	**45%** Research projects and outputs are annually monitored against organisational priorities and KPIs[1].	**30%** Individual evaluations of the programmes by external panels (2015).	**1%** Annual progress reporting on milestones, deliverables and KPIs.
Formative	**35%** Planning and monitoring of outputs and activities and outcomes within individual projects. Requirement to record in project management system.	**20%** Mainly implemented in relation to the Fund for the promotion of Agricultural Technology (FPTA) projects. Currently an evaluation strategy is being developed with the Policy Unit of the Ministry of Agriculture.	**35%** Annual objectives of programmes (scientific, economic & transfer).	**50%** Part of cyclical review process of research programmes and advisory regions.

[1] KPIs – Key Performance Indicators

Table 6: Proportion and examples of types of evaluation being done in each organisation (continued)

Types of evaluation	AgResearch	INIA	IRTA	Teagasc
Outcomes	**5%** Programme case studies in AgResearch Annual Report. Some outcome reporting for funders for individual projects.	**10%** Research products are evaluated against organisational KPIs, and from there to individuals using a cascade-approach [6].	**5%** Being part of the CERCA system in 2018 (The Agency for the Research Centres of Catalonia).	**43%** Research evaluated against organisational KPIs. Case studies in annual reports and part of cyclical review process.
Impact	**5%** Programme case studies in AgResearch Annual Report.	**5%** INIA is currently defining its evaluation strategy, and the impact evaluation approach. In 2015 the FPTA fund and selected individual programmes were evaluated.	**5%** CREDA's ex-post evaluation of the societal impact of research and innovation (2016-2019).	**5%** Annual Research Impact Highlights Cyclical review process

[6] The cascade approach goes from the macro (organisations) to the meso and then the micro-level of individuals, through stakeholders working at each level, e.g. managers and regional directors, and then through the programme directors.

Evaluative Capacity Building Activities

While not an entirely comprehensive list, Table 7 provides an indication of specific initiatives, strategies, actions, and key resources that the four organisations used to implement evaluation capacities in planning, monitoring, and communication (including reporting). In the planning stages, all organisations had strategy documents, and most utilised an advisory board to support them. AgResearch has targeted moving evaluation to the programme-level and indicate several initiatives to assist with this. Monitoring activities were evident around organisational KPIs and annual reporting. At the programme-level, reviews or evaluations were also a key feature in all four organisations. To communicate impact, all four organisations utilised both widespread communication, as well as more focused means. M&E activities that have been easier to implement provide an obvious and immediate value to staff or management. For example, in INIA-Uruguay and IRTA this has been evidencing the economic return on research investment for accountability purposes. In Teagasc this has been M&E as part of cyclical reviews of programmes. In AgResearch this has been programme logics because they have helped research teams to prepare proposals for funding, which is evidenced by the tools page of AgResearch's Beyond Results website being the site's second most visited page.

A key evaluative capacity building activity in all four organisations has been establishment of groups or teams (internal in INIA-Uruguay, Teagasc and AgResearch and external for IRTA) with dedicated budgets, time, and staff to undertake M&E (Table 7 and Figures 1 to 4). This has demonstrated the organisational commitment to M&E.

Table 7: Organisational activities to support planning, monitoring, and communicating for impact as of 2017 baseline workshop.

Evaluation stage	AgResearch	INIA	IRTA	Teagasc
Plan	Adoption & Practice Change programme funding Benefits/impact assessment for large science programmes Training (internal and external) to develop capability Advisory Board Participatory programme logics for planning research (outcome focus) Evaluation champions initiative to support planning Processes to support incorporation of evaluation in application process Science Plan/Statement of Core Purpose Co-innovation guidelines and success principles Impact planning tool for new programmes to understand stakeholders Internal community of practice around evaluation champions	Strategic Plan 2017-2020 International Advisory Board	Strategic Plan 2017-2020 International Advisory Board	Teagasc Statement of Strategy Evaluation Unit Strategy Annual Business Plans

1 **Table 7**: Organisational activities to support planning, monitoring, and communicating for impact as of 2017 baseline workshop (continued)

Evaluation stage	AgResearch	INIA	IRTA	Teagasc
Monitor	Evaluation plans for large science programmes Organisational KPIs Ad hoc reviews, surveys, etc. for different parts of the organisation	Project level experience for +20 years Organisation level KPIs (in 2017 used an international board)	Annual reports Impact on society document Three indices: Research, Economic, Transfer	Business Planning mid- and end-year reviews Internal audit Cyclical peer reviews Ad hoc reviews
Communicate	AgResearch website and publications Impact case studies for annual report AgResearch Annual Report/KPI reporting In development – upskilling/training/resources of staff to tell impact story Publications regarding process of evaluation Beyond Results resources available (www.beyondresults.co.nz)	Press conference Report on 20 years of impact evaluation Surveys – about image of INIA in society Surveys – about knowledge application and satisfaction with technology	Invited speakers Publications Mass media Activities Newsletter	Board of Directors (Teagasc Authority) Website publications Action Plans Research impact highlights Project teams

2

Good practice examples in the organisations (Table 7) have involved M&E being built into project or programmes as part of cyclical reviews (e.g. Teagasc and IRTA) so that programmes learn from and are guided by M&E. In the case of Teagasc cyclical reviews are an established organisational process. In AgResearch there are examples of projects including M&E at the proposal stage and then revisiting M&E through the life of the project.

Another good practice example of activities for building evaluative capacity is using opportunities for internal and external practitioner and academic feedback on M&E practices as part of continuous improvement (Table 7 and Figures 1 to 4). For example, the CREDA team working with IRTA have improved M&E practice through collaboration with other international research centres (e.g. INIA-Chile), feedback from IRTA staff and management, and academic feedback through conferences and peer-review of publications on M&E methodologies. This builds a network, including social scientists, to systematise activities to bridge the gap between academic research and the practice of M&E, which has been highlighted by Joly et al. (2016) as an opportunity to strengthen evaluation capacities in agricultural research organisations. This was also highlighted in Hall's (2018) review of AgResearch in recommending "The need to create a critical mass of thought and practice leadership dedicated to the programme to anchor and frame it and to act as a bridge between theory and practice. This is an issue of adequate resourcing and appropriate skills".

Lessons Learned

Leadership is important for organisational readiness for evaluation

The degree to which the organisations have systems and capabilities in place to collect and use evaluation data appears to be positively related to the degree to which the organisations value demonstrating impact and in which leaders are committed to building evaluation capacity (Table 8). For example in a review of AgResearch practices for science impact Hall (2018, p. 16 highlighted "The importance of senior management level support and championing of capacity building for impact programs in research organisation". Teagasc has already begun to consolidate the systematic use of

evaluation at all levels of the organisation, including processes for collecting data. Teagasc's recognition of the value of demonstrating impact is also reflected in the organisation being at the consolidating stage of skills, knowledge, and resourcing to collect and use evaluation data. The other organisations are emerging or developing the skills, knowledge, and resources to collect and use evaluation data.

Perhaps reflecting that leadership commitment to building evaluation capacity is still developing in AgResearch, and INIA, these organisations are at the stage of developing processes for the systematic use of evaluation but are still emerging in having processes to collect evaluation data. For example, all AgResearch competitively funded programmes are required to develop a programme logic. However, in most cases this has yet to be translated into evaluation plans and collection of evaluation data for funded programmes. Leadership and valuing the demonstration of impact, therefore appear to be important precursors to other aspects of organisational readiness to implement evaluation as well as sustaining evaluative capacity, as has previously been observed by Stone-Jovicich et al. (2019), Wade & Kallemeyn (2020), and White et al. (2018). This is not surprising given the importance of these factors in the success of organisational change initiatives, as highlighted by research on successful organisational change (e.g. Kotter (2012) and Brinkerhoff and Morgan (2010)).

Table 8: Degree of organisational readiness to implement evaluation capacities as of 2018.

Evaluation capacities	Emerging	Developing	Consolidating	Highly developed
Organisation values demonstrating impact		IRTA INIA	AgResearch	Teagasc
Leadership committed to building evaluation capacity	IRTA	AgResearch INIA	Teagasc	
Systematic M&E occurs at all levels of organisation	IRTA	AgResearch INIA	Teagasc	
Systems and processes to collect M&E data	AgResearch IRTA		Teagasc INIA (in projects)	
Skills and knowledge to collect M&E data		AgResearch IRTA	Teagasc INIA	
Skills and knowledge to use M&E data	AgResearch IRTA	INIA	Teagasc	
Sufficient resourcing to collect M&E data	AgResearch IRTA	INIA	Teagasc	
Opportunities to involve stakeholders in M&E		AgResearch INIA IRTA	Teagasc	

Internal and external support networks to build evaluative capacity

All four organisations have benefited from having both internal and external support for M&E activities (Table 7 and Figures 1 to 4). Internal support included senior management resourcing and championing the inclusion of M&E in organisational processes. External support included bringing in M&E experts to facilitate evaluation or train social scientists in evaluation best practice, particularly through "train the trainers" and "learning by doing". For example, IRTA has worked closely with a team of social researchers at CREDA who are experts in research impact assessment.

Also beneficial was growing internal and external networks of M&E champions to build legitimacy for M&E and share good practices. The latter has previously been recommended by Joly et al. (2016) and our findings highlight the benefits of this for creating an enabling environment for M&E. Internally, science staff leading large research programmes or with experience in evaluation can be influential champions. For example, Teagasc has supported evaluators to move into other roles in the

organisation to embed evaluative thinking. Three of the organisations (INIA, IRTA and AgResearch) have benefited from joining a network of other research organisations and agencies implementing M&E. This includes bringing in external experts to participate in seminars and workshops on evaluation methodologies to build awareness and evaluation capacity, an approach also used by Teagasc.

Build evaluation practices into existing planning and review processes

Each organisation had unique reasons for the effectiveness of activities to build evaluative capacity. These reflect the specific characteristics of existing evaluation and science impact activities. Teagasc has embedded evaluation activities into the cyclical reviews of research programmes and knowledge transfer activities that it conducts. These reviews provide an opportunity for staff and stakeholders to reflect on and learn from programmes outside of the annual business planning process. IRTA has been effective in embedding evaluation of impact by having a dedicated knowledge transfer unit that is assessed on its knowledge transfer activities through end-user assessment, technology transfer measures and use of this to inform further improvement. This has included increased resourcing and staff to undertake M&E of knowledge transfer activities. In contrast AgResearch did not have staff dedicated to knowledge transfer nor conducts cyclical reviews. In this context embedding M&E into organisational planning and funding processes, including making M&E tools easily available, has been most effective for increasing M&E activities. These differences in programme and project-level evaluation are like the differences observed in the five agricultural research organisations studied by Joly et al. (2016).

A common theme across the organisations though is that M&E activities were built into existing organisational processes (cyclical reviews, knowledge transfer, project planning) and incentivised (KPIs, funding). The value of cyclical reviews by IRTA and Teagasc for incentivising M&E of impact and reflection and learning suggests this could be an opportunity for other research organisations to encourage on-going M&E. Teagasc is seeking to deepen the integration of M&E with regular business

84 planning activities. There is a risk, however, that as was experienced by AgResearch, operationalising
85 M&E means that it can become a 'tick box' activity for compliance. Overall however, evaluative
86 capacity building will benefit from having organisational processes for planning and review, as well as
87 incentives for delivering impact, already in place into which M&E activities can then be embedded.

88 ***Different drivers of organisation- and programme-level evaluation***

89 Given different organisational emphasis on evaluation for accountability, advocacy, allocation, or
90 analysis (Table 4), and how these may drive a focus on organisation or programme-level evaluation,
91 none of the organisations would consider themselves highly developed in systematically embedding
92 evaluation at all levels from programme to whole of organisation (Table 8). Out of the four
93 organisations in this study, INIA, and IRTA appear to emphasise accountability and advocacy, which
94 focuses evaluation on organisation-level processes and reporting. For example, strategic plans inform
95 KPIs used in reporting on organisational performance to funders (Table 6). AgResearch appears to
96 emphasise allocation purposes, and a focus on programme-level evaluation, through cyclical peer
97 reviews of individual programmes or areas of research by advisory boards. Teagasc appears to be the
98 most advanced in undertaking evaluation at organisation and programme-levels for accountability,
99 advocacy, allocation, and analysis.

100 This suggests there is an opportunity to develop organisation and programme-level evaluation
101 processes that inform each other. Organisation-level evaluation sets longer-term outcomes
102 (accountability and advocacy) to which multiple programmes deliver, and then supports reporting on
103 the contribution of a suite of programmes to these outcomes. Programme-level evaluation therefore
104 assists with the allocation of funds to the mix of programmes to realise outcomes.

105 There is less use of evaluation for analysis (or learning) to better understand the processes that could
106 lead to greater impact from research. However, since this research was undertaken, Teagasc has
107 begun to address this gap by incorporating a theory-led approach into its research evaluations. This
108 role of evaluation in supporting increased research impact has recently been emphasised in other

agricultural research organisations such as CIRAD (Blundo-Canto et al., 2019), the International Livestock Research Institute (Kristjanson et al., 2009), Wageningen University, and the Dutch Agricultural Research Institutes (Spiertz & Kopff, 2011). In the case of the four research organisations studied here, the emphasis remained on using evaluation for improving the efficiency and allocation of funding within the organisations and with sporadic participation of stakeholders (Table 8). Evaluation provides the potential for organisations to work differently with stakeholders to realise impact by keeping them more engaged in programmes, their progress, and impact (Blundo-Canto et al., 2019; Percy et al., 2015; Stone-Jovicich et al. 2019). This practice ranges from emerging (INIA) through to highly developed (Teagasc). For example, Teagasc uses stakeholder engagement both in programme development and evaluation. These programme logics and evaluation plans can then be used to assess how a project is progressing toward desired outcomes, provide accountability for public investment in research, and produce evidence of the benefits of this investment.

Evaluation capacities for learning remain limited compared with capacities for accountability

M&E for learning remains under-resourced relative to evaluation for accountability and the organisations lack the necessary structures to embed M&E for learning (see also Joly et al. (2016)). The four organisations have therefore found it challenging to implement M&E to understand and learn from the pathways by which science contributes to wider societal impact. This challenge has previously been highlighted by Joly et al. (2016) in a review of M&E practices in five agricultural research organisations. In the current study, two reasons for this challenge were identified.

Firstly, M&E for learning is more complex as it needs to reflect the multitude of contexts, actors, interactions, impact pathways and outcomes of individual projects. This places greater demands on organisational resources and capacities, for example to undertake case studies, support the information systems to collect and track a wider variety of data at programme, project, and individual levels, and to standardise findings across case studies to report on organisational-level impact (see

also Joly et al. (2016)). It also makes analysis of attribution of individual organisations challenging, if not impossible.

Secondly, it was highlighted that strong cultural obstacles to embedding learning from M&E remain. For example, a senior manager in one of the organisations highlighted that "one of the major difficulties is … still deeply rooted belief that science already fulfils its objectives by generating knowledge and that impact is not a priority or even it is assumed that it will happen if the science is good, a fact that rarely happens." This may reflect a deeply rooted belief in the separation of scientific and societal values and linear model of knowledge production (Smit & Hessels, 2021). That is, that science is focused on knowledge generation and therefore is not immediately about impact or that impact is something that will automatically happen if research meets science excellence criteria. This culture is emphasised by the tendency for the organisations to focus on traditional measures of scientific excellence (e.g. h-index, citations), stakeholder satisfaction surveys, economic impacts evaluated at the end of a project, or include evaluation only at the planning stage. Thus, there has been limited on-going monitoring and use of that data for reflection and learning, with these activities seen as a "nice to do" that compete with science activities and funding. As Smit and Hessels (2021) highlight in a review of ten evaluation methodologies, the culture influences the types of evaluation methods prioritised by research organisations.

There is an opportunity to build capacity amongst in-house M&E teams to conduct M&E for learning by deeper engagement with the academic research and researchers on this type of evaluation. To then begin to reshape organisational culture, evaluation methods that are grounded in co-production and integrated scientific and societal values should be emphasised, e.g. contribution mapping and evaluative inquiry (Smit & Hessels, 2021). Teagasc has begun to address on-going evaluation for learning by introducing three high-level impact pathways to describe the ways Teagasc contributes to change in the agri-food sector. AgResearch also identified the need for clear narratives of the

pathways to impact from its research (Hall, 2018), and has found that the process of project teams developing programme logics raised awareness of the existence of different impact pathways.

Evaluation capacity building increases the frequency of evaluation in agricultural research organisations

For evaluation to thrive within an organisation, it must be valued (Stone-Jovicich et al., 2019) and enabled (Preskill and Boyle, 2008; Wade & Kallemeyn, 2020; White et al., 2018). Figure 6 shows some consistent trends regarding the four organisations' positioning in this regard. At the time of the study, IRTA was in an emerging space whereby the skills, knowledge, resources, and opportunities were yet to be developed; however, it is developing this area. Teagasc had some of the key enablers to collect evaluation data (systems, processes, skills, knowledge) within the organisation, but not yet systemically distributed from the programme to organisation-levels. AgResearch and INIA were positioned in between, considering themselves to be in an emerging to developing position. Organisational evaluation capacity readiness is also reflected in the commitment of leadership. Within AgResearch, INIA and Teagasc, organisational leaders and managers have shown commitment to developing evaluation capacity. Leadership within IRTA has yet to mandate further internal development of systems and processes in place to collect evaluation data and capacity for evaluation.**Figure 6**: Organisations' positioning towards Evaluation Capacity Building.

While the four organisations included M&E for learning in guiding principles, all were predominantly focused on M&E for accountability and advocacy. This was reflected in a disconnect between evaluation at the programme- and organisational-level, with an emphasis on evaluation of organisational KPIs. For example, IRTA has focused on embedding research impact assessment. This was prioritised for accountability and reporting at specified periods. M&E for learning at the programme-level appeared to be left to individual programmes and received less resourcing and capacity within the organisations, and where it did occur was a point in time event, rather than a continuous activity. The experience of Teagasc in embedding M&E as part of cyclical programme reviews could be an opportunity to address the programme- and organisation-level disconnect by evaluating suites of programmes and projects delivering to a common objective (e.g. Sustainability Development Goals) and involving project and programme teams in reflecting on these evaluations.

Conclusion

To the best of our knowledge, the present study is one of the first attempts to provide empirical evidence of how publicly funded agricultural research organisations put research evaluation principles

into practice. Drawing on our analysis, accountability and allocation are strong drivers of evaluation in the four agricultural research organisations studied. For most of the organisations, evaluation is concentrated on internal research performance, with less focus on evaluation of societal, environmental, and economic impacts. However, perhaps reflecting the influence of the impact agenda, the organisations are beginning to increase their emphasis on summative evaluation. This is reflected in organisational leaders valuing the demonstration of impact and commitment to building evaluation capacity. These appear to be important precursors to other aspects of organisational readiness to implement evaluation.

Given different organisational emphasis on evaluation for accountability, advocacy, allocation, or analysis, and how these may drive a focus on organisation or programme-level evaluation, none of the organisations would consider themselves highly developed in systematically embedding evaluation processes and capabilities at all levels of the organisation from programme to whole of organisation. This suggests there is an opportunity to simultaneously develop organisation- and programme-level evaluation processes that inform each other. There is also less use of evaluation by the organisations to better understand the processes that could lead to greater impact from research.

The four organisations have demonstrated their ability to develop evaluation capacities and to establish cycles of periodical impact assessment within their organisations. Given the difference in development across organisations, we believe that more on-going collaboration between the four organisations would provide the opportunity to learn constantly through exchanging information and insights of 'what works' in other research organisations and create a network to improve the evaluation capacity, which in turn helps sustain future science impact. Finally, our empirical findings reveal that the research organisations' challenge is mainly to adapt evaluation practices in terms of skills, knowledge, time, data, resourcing, and supportive structure to build an ECB culture within their own context.

Acknowledgements

Thank you to Teagasc and IRTA for hosting the workshops, and to Penny Payne and Roxanne Henwood for reviewing an early draft of the manuscript.

Funding: This work was supported by the Strategic Partnership; Strategic Science Investment Fund, New Zealand; Scottish Government; Teagasc; Instituto Nacional de Investigación Agropecuaria (INIA) in Uruguay; and the Societal Impact of R&D Investments [IRTA project-61095].

References

AgResearch (2020). *Annual Report 2020*. AgResearch Ltd., Lincoln, New Zealand.

AgResearch (2018). *AgResearch: Statement of Core Purpose*. AgResearch Ltd, p. 2.

Blundo-Canto, G., Triomphe, B., Faure, G., Barret, D., De Romemont, A., & Hainzelin, E. (2019). Building a culture of impact in an international agricultural research organization: Process and reflective learning. *Research Evaluation*, 28(2), 136-144.

Botha, N., Coutts, J., Turner, J.A., White, T., & Williams, T. (2017). Evaluating for learning and accountability in system innovation: Incorporating reflexivity in a logical framework. *Outlook on Agriculture, 46*, 154-160.

Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational evaluation capacity. *American Journal of Evaluation, 35(4)*, 579-593.

Bourgeois, I., Whynot, J., & Thériault, É. (2015). Application of an organizational evaluation capacity self-assessment instrument to different organizations: Similarities and lessons learned. *Evaluation and Program Planning, 50*, 47-55.

Bozeman, B., & Sarewitz, D. (2011). Public value mapping and science policy evaluation. *Minerva 49*, 1-23.

Brinkerhoff, D.W., & Morgan, P.J. (2010). Capacity and capacity development: Coping with complexity. *Public Administration and Development: The International Journal of Management Research and Practice, 30(1)*, pp.2-10.

Chaplowe, S. & Cousins, B. (2016). *Monitoring and evaluation training: A systemic approach*. California: Sage Publications Inc.

Costa, M. (2022). *Experiencia sobre evaluación en INIA Uruguay* INIA, Montevideo.

Cousins, J.B., Goh, S.C., Elliott, C., Aubry, T., & Gilbert, N. (2014). Government and voluntary sector differences in organizational capacity to do and use evaluation. *Evaluation and Program Planning, 44,* 1-13.

Cousins, J., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. *New Directions for Evaluation, 141*, 7-23.

Crown Research Institute Taskforce (2010). How to Enhance the Value of New Zealand's Investment in Crown Research Institutes. *Report of the Crown Research Institute Taskforce*. Wellington: Ministry of Business, Innovation and Employment.

CSIRO (2015). *Impact Evaluation Guide*. Canberra: CSIRO.

Donovan, C. (2011). State of the art in assessing research impact: introduction to a special issue. *Research Evaluation, 20,* 175-179.

Douthwaite, B., & Hoffecker, E. (2017). Towards a complexity-aware theory of change for participatory research programs working within agricultural innovation systems. *Agricultural Systems, 155*, 88-102.

Gaunand, A., Hocdé, A., Lemarié, S., Matt, M., & Turckheim, E.D., 2015. How does public agricultural research impact society? A characterization of various patterns. *Research Policy, 44*, 849-861.

Guesmi, B., & Gil, J.M. (2021). The impact of public R&D investments on agricultural productivity. *Review of Economics and Finance, 19(1), 284-291*.

Hall, A. (2018). Forward Looking Review of AgResearch's Adoption and Practice Change Program. Report to AgResearch Ltd. AgResearch Ltd, Lincoln, New Zealand

Horton, D., & Mackay, R. (2003). Using evaluation to enhance institutional learning and change: Recent experiences with agricultural research and development. *Agricultural Systems, 78*, 127-142.

Hueftle Stockdill, S., Baizerman, M., & Compton, D.W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation, 2002*, 7-26.

Joly, P.-B., Gaunand, A., Colinet, L., Larédo, P., Lemarié, S., & Matt, M. (2015). ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization. *Research Evaluation, 24(4)*, 440-453.

Joly, P.-B., Colinet, L., Gaunand, A., Lemarié, S., & Matt, M. (2016). "Agricultural research impact assessment: Issues, methods and challenges", OECD Food, Agriculture and Fisheries Papers, No. 98, OECD Publishing, Paris. http://dx.doi.org/10.1787/5339e165-en

Kelley, T., & Gregersen, H. (2003). NRM impact assessment in the CGIAR: meeting the challenge and implications for ICRISAT, Methods for assessing the impacts of natural resource management research. *A summary of the proceedings of an International Workshop, 6-7 Dec 2002*, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502 324, India.

Kelley, T., Ryan, J., & Gregersen, H. (2008). Enhancing ex post impact assessment of agricultural research: the CGIAR experience. *Research Evaluation, 17*, 201-212.

Kotter, J.P. (2012). *Leading Change*. Brighton: Harvard Business Press.

Kristjanson, P., Reid, R.S., Dickson, N., Clark, W.C., Romney, D., Puskur, R., MacMillan, S., & Grace, D., 2009. Linking international agricultural research knowledge with action for sustainable development. *Proceedings of the National Academy of Sciences, 9*, 5047-5052.

Mackay, R. & Horton, D. (2003). Expanding the use of impact assessment and evaluation in agricultural research and development. *Agricultural Systems, 78*, 143-165.

Maredia, M.K., Shankar, B., Kelley, T.G., & Stevenson, J.R. (2014). Impact assessment of agricultural research, institutional innovation, and technology adoption: Introduction to the special section. *Food Policy, 44*, 214-217.

Mayne, J., & Johnson, N. (2015). Using theories of change in the Agriculture for Nutrition and Health CGIAR research program. *Evaluation, 21(4)*, 407–428.

Menon, S., Karl, J., & Wignaraja, K. (2009). *Handbook on planning, monitoring and evaluating for development results*. UNDP Evaluation Office, New York, NY.

Midmore, P. (2017). The science of impact and the impact of agricultural science. *Journal of Agricultural Economics, 68*, 611-631.

Ministry of Business, Innovation and Employment (2015). *National Statement of Science Investment: 2015-2025*. Wellington: Ministry of Business, Innovation and Employment.

Ministry of Business, Innovaton and Employment (2017). *The Impact of Science, Discussion Paper*. Wellington: Ministry of Business, Innovation and Employment.

Morgan Jones, M., Manville, C. & Chataway, J. (2017). Learning from the UK's research impact assessment exercise: a case study of a retrospective impact assessment exercise and questions for the future. *The Journal of Technology Transfer 10, 1-25*

Organisation for Economic Co-operation and Development (2010). *Performance-based funding for public research in tertiary education institutions: Workshop proceedings*. Paris: OECD Publishing.

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy, 39*, 751-760.

Parks, S., Rodriguez-Rincon, D., Parkinson, S., & Manville, C. (2019). The changing research landscape and reflections on national research assessment in the future. *Research England, 2019*. https://www.rand.org/pubs/research_reports/RR3200.html.

Penfield, T., Baker, M.J., Scoble, R., & Wykes, M.C. (2014). Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation, 23*, 21-32.

Percy, H., Turner, J.A., Botha, N., White, T., Small, B., Fear, A., & Cooper, A. (2015). Beyond results: Planning science for meaningful change. *Rural Extension and Innovation Systems Journal, 11*, 173.

Pollock, C. (2012). Repairing a fractured pipeline: Improving the effectiveness of agricultural R&D in the UK. *International Journal of Agricultural Management, 2*, 1-4.

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation, 29,* 443-459.

Sheng, Y., Gray, E.M., & Mullen, J.D. (2011). Public investment in R&D; and extension and productivity in Australian broadacre agriculture (No. 422-2016-26880). Australian Agricultural and Resource Economics Society (AARES), 9–11 February 2011, Melbourne, Victoria

Smit, J. P., & Hessels, L. K. (2021). The production of scientific and societal value in research evaluation: a review of societal impact assessment methods. Research Evaluation, 30(3), 323-335. doi:10.1093/reseval/rvab002

Spiertz, J.H.J., & Kropff, M.J. (2011). Adaptation of knowledge systems to changes in agriculture and society: The case of the Netherlands. *NJAS - Wageningen Journal of Life Sciences, 58*, 1-10.

Stone-Jovicich, S., Percy, H., McMillan, L., Turner, J., Chen, L., & White, T. (2019). Evaluating monitoring, evaluation and learning initiatives in the New Zealand and Australisan agricultural research and innovation systems: The MEL^2 framework. *Evaluation Journal of Australasia, 19(1)*, 8 - 21.

Social Policy Evaluation and Research Unit (Superu) (2017a). *Making sense of evaluation: A handbook for everyone*. Wellington, New Zealand. Available at: https://dpmc.govt.nz/sites/default/files/2018-03/Evaluation%20Handbook%20Dec%202017.pdf (accessed 6 August 2021).

Social Policy Evaluation and Research Unit (Superu) (2017b). *Getting Your Organisation Ready to Do Evaluations: Assessment Guide and Templates*. Wellington, New Zealand. Available at: https://thehub.swa.govt.nz/assets/documents/Toolkit-print-FINAL-03-17.pdf (accessed 6 August 2021).

Taylor-Powell, E., & Henert, E. (2008). Developing a logic model: Teaching and training guide. *Benefits, 3*, 22.

Teagasc (2017). Statement of Strategy 2017 -2020. *Supporting Science-based Innovation in Agriculture and Food by driving sustainable profit from productivity*. https://www.teagasc.ie/media/website/publications/2017/Statement-of-Strategy-2017-2020.pdf

Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R., & Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature, 418*, 671.

Turner, J.A., Klerkx, L., Rijswijk, K., Williams, T., & Barnard, T. (2016). Systemic problems affecting co-innovation in the New Zealand Agricultural Innovation System: Identification of blocking mechanisms and underlying institutional logics. *NJAS - Wageningen Journal of Life Sciences, 76*, 99-112.

Turner, J.A., Rijswijk, K., Williams, T., Barnard, T., & Klerkx, L. (2013). Challenges to effective interaction in the New Zealand agricultural research and extension system: An innovation systems analysis. *Extension Farm Systems Journal, 9*, 89-98.

UK Research and Innovation (2022). Case for support – impact funding. https://www.ukri.org/councils/stfc/guidance-for-applicants/what-to-include-in-your-proposal/case-for-support-impact-funding/ (Accessed 24th February, 2022)

Vasen F, Sierra M, Paruelo JM, Negro C, Nolla F, Lapetina J, Salvagno M. (2021). Evaluation of technical production in agricultural sciences: a new certification scheme in Uruguay. Agrociencia Uruguay [Internet]. 2021 25(2):e491. Available from: http://agrocienciauruguay.uy/ojs/index.php/agrociencia/article/view/491.

Wade, J., & Kallemeyn, L. (2020). Evaluation capacity building (ECB) interventions and the development of sustainable evaluation practice: An exploratory study. *Evaluation and Program Planning, 79*, 1-9.

Watts, J., Horton, D., Douthwaite, B., La Rovere, R., Thiele, G., Prasad, S., & Staver, C. (2008). Transforming impact assessment: beginning the quiet revolution of institutional learning and change. *Experimental Agriculture 44*, 21-35.

White, T., Percy, H., & Small, B. (2018). Creating an evaluation culture through capacity building: A new frontier in a science organisation. *Evaluation Matters – He Take Tō Te Aromatawai, 4*, 111-135.