OECDpublishing

Please cite this paper as:

Joly, P. *et al.* (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

OECD Food, Agriculture and Fisheries Papers No. 98

Agricultural research impact assessment

ISSUES, METHODS AND CHALLENGES

Pierre-Benoit Joly, Laurence Colinet, Ariane Gaunand, Stéphane Lemarié, Mireille Matt

JEL Classification: O31, O38, Q16

OECD

OECD FOOD, AGRICULTURE AND FISHERIES PAPERS

Comments are welcome and may be sent to tad.contact@oecd.org.

Abstract

AGRICULTURAL RESEARCH IMPACT ASSESSMENT: ISSUES, METHODS AND CHALLENGES

The Research Impact Assessment (RIA) is expected to increase the efficiency with which public funds are used, and to improve more broadly the functioning of the research and innovation system and its contribution to address a wide range of socio-economic and environmental issues. Both-standard economic approaches, which aim to estimate the economic benefits of research investments, and case-study approaches, which aim to analyse the processes of impact generation, have been applied to agricultural research in practice. Standard economic approaches generally focus on public research as information on private efforts in agricultural research is limited, and on economic impacts such as productivity growth. Case studies provide richer information, through a narrative, and highlight the complex relationships among the various variables, events and actors, but it is difficult to standardise results and scale them up. The challenge for RIA is to take into account broader impacts that go beyond science and economic impacts, and to improve knowledge on impact-generating mechanisms. This has become more difficult as agricultural research and innovation systems are increasingly open and complex, and changing quickly. Observation of practices applied to agricultural research in five selected organisations confirms the difference found in RIA between academic research and in practice. In both, the assessment systems pursue the same objectives: 1) Learning: enhance the know-how to produce an environment conducive to socio-economic impact; 2) Capacity building: spread the culture of socio-economic impact to its researchers; and 3) Reporting to stakeholders: from accountability purposes to advocacy targeted to various audiences. The accountability objective, including estimating returns on the financial investment, poses complex challenges and is in tension with the learning and capacity building objectives. The future of RIA will depend on the capacity to improve estimation methods and gather quality information (which also takes into account non-economic impacts) and the sharing of good practices.

Keywords: Agricultural research, research impact evaluation.

JEL Classification: O31, O38, Q16.

Acknowledgements

This consultant report was prepared by Pierre-Benoit Joly, Laurence Colinet, Ariane Gaunand, Stéphane Lemarié and Mireille Matt from the French Agricultural Research Institute INRA. It contributes to OECD work on innovation in food and agriculture.

Table of contents

Tables

Executive summary

Agricultural research and innovation systems are expected to address a wide range of socio-economic and environmental issues and to contribute to global challenges, while improving the efficiency with which public funds are used. Among other changes aimed to improve the functioning of the systems, this prompted renewed interest for Research Impact Assessment (RIA) methods and practice in the research community and governments.

The challenge for RIA is to take into account broader impacts, beyond those on science and economic performance and improve knowledge on impact generating mechanisms. It is difficult to link efforts and impact as research and innovation systems are increasingly open and complex; and are changing at a quick pace. Furthermore, in complex systems, the impact is not additive but depends on productive interactions.

Two sets of methodologies are discussed in this report — “standard economic approaches” and “approaches based on case studies”. Standard economic approaches aim to estimate economic benefits of research investments in order to calculate economic indicators of the impact of research such as internal rate of return or cost-benefit ratio. They generally focus on public research as information on private efforts is limited, and on economic impacts such as productivity growth. Case-study is a useful method to analyse how and why a phenomenon occurs. Case studies provide richer information, through a narrative, about the element which is evaluated and highlights the complex relationships among the various variables, events and actors, but it is difficult to standardise results and scale them up.

These methods are complementary and it is crucial to develop approaches that match quantitative and qualitative analysis to reinforce the credibility of RIA. For both sets of methods, adequate databases and metrologies need to be developed, in particular to take into account non-economic impacts.

Observation of RIA practices in five selected organisations confirms the gap between academic research and practices. However, in some organisations interactions between research and practice is organised in a systematic way and involve economists, agronomists and social scientists in the design of approaches. RIA is high on the agenda of all the organisations, as building credibility is expected to secure funding. However, limited resources devoted to RIA constrain the implementation of systematic and comprehensive evaluation in many cases. Evaluation systems vary across organisations in scope and level, depending on programming design.

The assessment systems pursue the same three objectives:

- Learning: enhance the know-how to produce an environment conducive to socio-economic impact.
- Capacity building: spread the culture of socio-economic impact to its researchers.
- Reporting to stakeholders: from accountability purposes to advocacy targeted to various audiences.

These objectives are difficult to pursue simultaneously. The accountability objective, including for the purposes of return on the financial investment, poses particularly complex challenges, and there are tensions between this objective and learning and capacity building objectives. Some experts suggest it may be necessary to make choices between these objectives, and determine whether evaluation is for internal purpose or external control. It is argued that multidimensional evaluation to foster internal learning would be more efficient at improving research impact than external evaluation for control. RIA would thus nurture a culture of impact and become a central tool for strategic intelligence.

Finally, the future of RIA depends on both the improvement of methods and information, which would be fostered by the commitment of institutions and the constitution of a community of professionals interacting globally.

Background

Research Impact Assessment (RIA) is not a new issue. Since the 1950s, the economic returns to research investment have been analysed repeatedly. Agriculture is the single economic activity that concentrates most analysis of economic impact of R&D investment, starting with the seminal contribution of Zvi Griliches (1958). Alston (2010) contains a comprehensive survey of these analyses. In addition, major programmes based on case studies (such as TRACES and HINDSIGHT) have focused on analysis of the non-academic impact of research.

RIA is receiving renewed attention in light of increased expectations about the ability of research to deliver a wider range of socio-economic impacts. The Lisbon Agenda (2000)[1] is one of the landmarks in this evolution, and the organisation of research towards major challenges has extended this logic. At the same time, many countries face budget constraints which reinforce the need to account for the impact of public funding. This context is promoting a revival of interest in RIA methodologies, and has been the motivation for a number of projects such as: Assessments of the impacts of the Advanced Technology Program (ATP) (Ruegg and Feller, 2003), Public Value Mapping (Bozeman, 2003), the Payback Framework (Donovan and Hanney, 2011), and the Social Impact Assessment Method (SIAMPI) (Spaapen and Van Drooge, 2011).

Various institutions have designed, and are experimenting with, new ways to assess the impacts of their research. Public sector Research Organisations (PROs) dedicated to agriculture are contributing to this rich field of experimentation including the Consultative Group for International Agricultural Research (CGIAR) (Walker et al., 2008), the Brazilian corporation of agricultural research (EMBRAPA, 2015), the Economic Research Service (ERS) of the US Department of Agriculture (USDA) (Heisey et al., 2010), and the Commonwealth Scientific and Industrial Research Organization for Australian research (CSIRO) (Acil Tasman Pty Ltd, 2010).

This report is based on an extensive survey of the literature of Agricultural Research Impact Assessment (ARIA) encompassing peer reviewed articles, publications and institutional reviews. It focuses on the links between academic research on RIA and actual implementation of approaches, an issue that is generally blind-spotted. Hence, the goal is to identify the state of the art and the current challenges of ARIA in the perspective of actual implementation.

Section 1 outlines the implications for ARIA of main societal challenges for agricultural R&D and developments in agricultural innovation systems. Section 2 looks at the two types of methodologies used for *ex post* research impact assessment, and Section 3 presents practices of ARIA in selected countries and research organisations. The final section proposes ways forward to improve research impact assessment and accountability.

1. See the strategic goals of the Lisbon agenda in the Presidency Conclusions of the Lisbon European Council 23 and 24 March 2000 at: www.consilium.europa.eu/en/uedocs/cms_data/docs/pressdata/en/ec/00100-r1.en0.htm.

1. Current challenges of agricultural R&D and issues for Agricultural Research Impact Assessment (ARIA)

Current challenges of agricultural R&D have some crucial implications for Agricultural Research Impact Assessment (ARIA). To sum up, the new agenda for ARIA must address the three following issues:

- take into account the diversity of dimensions of impacts of research;
- contribute to the credibility as well as to the improvement of the capacity of Agricultural Research and Innovation System (ARIS) to produce impact; and
- adapt to the growing complexity of ARIS.

The main societal challenges for agricultural research

As many inter-governmental and governmental organisations have pointed out, global agriculture will face multiple challenges over the coming decades. It must produce more food to feed an increasingly affluent and growing world population that will demand a more diverse diet. By 2050, world population is expected to swell to 9 billion people. The United Nations' Food and Agriculture Organization predicts that in that time global food production will need to increase by 70% in order to prevent massive famine.

Global agriculture must also contribute to overall development and poverty alleviation in many developing countries, confront increased competition for alternative uses of finite land and water resources, adapt to climate change, and contribute to preserving biodiversity and restoring fragile ecosystems (Interagency Report, 2012).

Improving agricultural productivity, while conserving and enhancing natural resources, is an essential requirement for farmers to increase global food supplies on a sustainable basis. While agricultural productivity represents a worldwide goal for agriculture research, other research objectives addressing societal challenges are becoming central in the agricultural research agenda. These include:

- Dealing with environmental issues:
  - address the shortage of natural resources, from fossil fuels to water to phosphorus,
  - contribute to biodiversity conservation,
  - reduce greenhouse gas emissions and contribute to carbon sequestration,
  - enhance ecosystem services,
  - reduce soil erosion,
  - reduce dependency on pesticides.
- Improving health: safety and healthy food provision, safety working conditions.
- Enhancing the social value of agriculture: poverty alleviation, maintenance of viable rural areas, and quality of life in rural areas.
- Reducing food waste from spoilage to produce culled by retailers.

These challenges are very high on the agenda of OECD and key partner countries, as ministers indicated when they met on 7-8 April 2016 at the OECD in Paris.[2] Research and innovation are considered as the main solutions to address these challenges. For instance, the US White House Council of Advisors on Science and Technology identified seven challenges: i) managing new pests, pathogens, and invasive plants; ii) increasing the efficiency of water use; iii) reducing the environmental footprint of agriculture; iv) growing food in a changing climate; v) managing the production of bioenergy; vi) producing safe and nutritious food; vii) assisting with global food security and maintaining abundant yields (PCAST, 2012).

2. See Ministerial statements at: www.oecd.org/agriculture/ministerial/statements/.

In order to meet these expectations, agricultural research needs sustained public and private funding but also a focus on challenges, an impact orientation and an improved responsiveness. Scholars acknowledge that addressing societal challenges requires innovation in research and innovation policies (Foray, Mowery and Nelson, 2012).

In this context, Research Impacts Assessment (RIA) must take into account a diversity of dimensions beyond productivity gains: environmental impacts (natural resources, biodiversity, climate change, soil conservation, and pollution); social impacts (viable rural areas, increased revenue of smallholders); and impacts on food safety and occupational health. As for other sectors, there is a strong need of RIA methodologies that take into account broader impacts (Bozeman and Sarewitz, 2011).

Looking back: Some shared concerns on the future ability of ARIS to address these challenges

It is widely acknowledged that the achievements in agricultural innovation over the past century have been impressive, supporting large increases in agricultural yields. Most of the increase in global agricultural production over the past 50 years has come from raising crop and livestock yields rather than through area expansion. This growth in productivity is attributed largely to investments in research and innovation (Wright and Shih, 2011). As a result, the economic impact of agricultural research since World War II (WW II) is very high (Alston, 2010).

The capacity of the ARIS to support transformations capable of addressing current global challenges is hotly debated. The World Bank Development Report 2008 identified a halving of the growth rate in grain yields in developing countries between 1970-1989 and 1990-2005 (Burch et al., 2007). More recent studies confirm these trends in yields, but bring more nuances between crops (Grassini et al., 2013; Ray et al., 2012). Some recent estimates suggest that total factor productivity (TFP), the most comprehensive measure of productivity reflecting the efficiency to turn all inputs into outputs, grew at an average rate of around 2% per year since 2000 across major world regions (Fuglie, 2010).

Other studies, in particular those using partial factor productivity indicators such as land and labour productivity, give a more pessimistic global picture, in particular when the People's Republic of China's (hereafter "China") performance is taken out of the calculation of the world average (Alston, 2010). The most popular indicator of land productivity is crop yield. The average global rates of growth in yield of most of the major cereals are declining. Since the 1980s, growth in wheat and rice yields fell from 2.5-3% to around 1%. Maize yields showed growth of slightly less than 2% over the last decade. While there is no evidence of a productivity slowdown, there is a clear decline in global cereal yield from close to 2.84% per year in the 1960s, to 2.31% between 1970 and 1989 and 1.35% between 1990 and 2007 (Fuglie, 2010: 85).

In addition to concerns about the recent contribution of ARIS, three elements further obscure the horizon. First, climate change may severely affect crop yields. Some studies estimate that global warming could result in a 6% reduction in global agricultural production by 2080 (Wilson, 2012). According to the same estimates, a 3°C to 4°C rise in temperatures would result in yield losses of 18% for wheat in northern Africa and 22% for maize in southern Africa. Parts of Africa and India are projected to suffer a 30% decline in food production under climate change.

Second, it is very likely that a growing part of the R&D effort is devoted to maintenance research. As discussed in Section 2, maintaining high yielding production based on limited biological and natural resources requires recurrent investments (related to pest management, soil conservation, etc.). This reduces the part of R&D devoted to the challenges identified.

Third, unlike the experience of the Green Revolution which relied on the wide diffusion of genetically uniform high-yield varieties complemented by high levels of inputs, increasing agricultural productivity in today's context will require gains among a large number of smallholders in very different agro-ecological regions. Note that this objective does not substitute productivity increase in large farms and that the balance between both objectives is discussed. The need to adapt to a wide diversity of situations may reduce the economies of scale in research, thus affecting the impact of investments.

Against this background, global institutions consider that ARIS will have to be more impact oriented. This may lead to developing outcome-based approaches, integrating research into development processes, and identifying key interventions required to remove blockages and barriers to large-scale impact. However, such an orientation may contribute to narrow the horizon of research (Mowery et al., 2010). The underlying theory of impacts should be examined more closely.

One of the key implications of these new challenges and concerns for agricultural research is that methodologies of impact assessment should improve knowledge on impact generating mechanisms. It is not sufficient to measure the efficiency of R&D investment; it is also necessary for impact assessment to contribute to the efficiency. Research impact assessment should improve research management at different levels, from research projects or programmes to the governance of the global system.

The increasing complexity of the Agricultural Research and Innovation System (ARIS)

Concerns on the future of global agriculture and food create a new momentum for agricultural research. The Global Conference on Agricultural Research for Development (GCARD), organised by the Global Forum on Agricultural Research (GFAR), has developed a roadmap for agricultural research.[3] This roadmap calls for an increasing investment in research, strengthening of relations between agricultural research (mainly related to developed countries) and agricultural research for development (AR4D, mainly related to developing countries), and the development of public and private partnerships (Beintema et al., 2012; CGARD, 2011). A transformation of AR4D systems is needed to: i) focus collective research and knowledge sharing on key outcome-focused themes globally; and ii) transform and strengthen agricultural innovation systems in developing countries.

As part of this new momentum, a worldwide system dedicated to Agricultural Science and Technology Indicators (ASTI) has been set up. It reports that in 2008 global investment in agricultural R&D amounted to USD 41 billion, of which 79% public funding and 21% private, 51% from high income countries, 49% low and middle income countries (Beintema et al., 2012). From 2000 to 2008, R&D investment increased steadily (22% public, 26% private, inflation-adjusted). However, these changes were unevenly distributed. Most of this growth was driven by developing countries, since growth in high-income countries stalled. Spending growth in developing countries was largely driven by positive trends in a number of larger, more advanced middle-income countries — such as China +38%, and India +11% — masking negative trends in numerous smaller, poorer, and more technologically challenged countries (Beintema et al., 2012).

The problem of data quality should be considered. In many countries, statistical sources for R&D investment are very poor for public funds and in general do not exist for private research. Coverage of agriculture in the OECD research expenditure database is also unequal (OECD, 2013).[4] Therefore, data on global agricultural research are estimates and they may be very different according to the sources.[5] The poor quality of private R&D sources raises major difficulties when measuring the returns to R&D (Fuglie et al., 2012).[6]

Globally the ARIS are in constant transformation. Recent analyses of European and OECD countries have pointed out the complexity and – especially for some countries - the fragmentation of the ARIS (EU SCAR, 2015; OECD, 2013; Moreddu and Poppe, 2013). In most countries, there is a relative increase of private R&D which results from the strengthening of property rights and a stagnation (or even decrease) of public R&D. Governments have encouraged the development of public private partnerships (Moreddu, 2016). This raises some questions on the efficiency of the research and innovation system and its ability to address global challenges that have in many cases public goods characteristics (Fuglie et al., 2012; Wright and Shih, 2011).

In this changing context, the assessment of research impact is increasingly difficult. Given the complexity of ARIS, it is very difficult to achieve impact attribution to the various actors participating which may threaten the sustained public investment in agricultural research. Alternative approaches based on contribution are currently under consideration. They are based on contextual and procedural analysis that allow to describe the role of the different actors of the innovation network to elicit their respective contributions. Contribution analysis will be discussed further in Section 2.

3. For more information on GCARD, see GFAR web site at: www.gfar.net/about-gcard.

4. In a few cases, it is because the country does not use international standards to report the data.

5. For the estimates of the InSTePP R&D Series (www.instepp.umn.edu/), see Pardey et al. (2014).

6. The European project IMPRESA (The Impact of Research on EU Agriculture) has achieved an inventory of the statistical sources available in 19 EU countries and Switzerland. It concludes that official sources on agricultural R&D are generally poor, which constrains the development of effective evidence based policies by European governments (Research Brief, August 2015 available on: www.impresa-project.eu/).

2. Methodologies for (ex post) Research Impact Assessment (RIA)

The literature points out key elements that identify the current strategic importance of RIA. As suggested by Rip (2003) the era of massive public investment in science based on a general expectation of positive outcomes has passed. Political systems, funding agencies and public research organisations (PROs) must demonstrate results from the public funds used. According to Rip, RIA serves not only to give an *ex post* delayed account of the impact of research; it needs to be more strategic and anticipatory, to assist systematic improvements, and to better identify the full range of outcomes from R&D investment. This does not mean that *ex post* RIA is out of scope. Past experience is the main source of knowledge. This means that RIA should be designed and performed in such a way that it helps to improve impact generating mechanisms. Hence, RIA approaches must foster their credibility and, at the same time, adapt to different purposes and address different audiences. The objectives are not only related to accountability but also to advocacy and learning. A major challenge is therefore to better link evaluation approaches and evaluation strategies with learning and continuous improvement, as well as evaluation’s more conventional role of justification (Shapira and Kuhlmann, 2003).

A wide variety of methods for RIA are available. Based on the assessment of the US Advanced Technology Program — probably the most important RIA initiative ever — Ruegg and Feller (2003:17) present a set of different methods:

- Analytical/conceptual methods for modelling and informing underlying programme theory
- Survey method
- Case study: Descriptive
- Case study: Economic estimation
- Sociometric and social network analysis
- Bibliometrics: counting, citing, and analysing content of documents
- Historical tracing
- Expert judgment.

The list presented by Ruegg and Jordan (2007) in their overview of evaluation methods for R&D programmes is even more comprehensive and includes: econometrics, mission/impact mapping, foresighting, etc. Based on a survey of US public agencies, most of the organisations involved in R&D perform some kind of ‘case studies’.

A comprehensive presentation of the different methods is beyond the scope of this report. Other surveys are available and it would be burdensome to add a new serial presentation. Furthermore, the comprehensiveness would not help answer the questions of interest of this report: What are the gaps between theory and practice? How to match qualitative and quantitative approaches? How to learn from *ex post* studies to improve R&D management? Neither would it add value to the points raised in Section 1: How to take into account the different dimensions of impact? How to improve knowledge on impact generating mechanisms? How to overcome the attribution problem?

Therefore, for sake of efficiency, this review focuses on two sets of methods. The first set is presented under the heading "Standard economic approaches" and follows Heisey et al. (2010) who use the same label to present econometric approaches and economic surplus techniques. These methods have been extensively applied in agriculture where they were first used in the late 1950s. Although a recent survey undertaken for the OECD presents the main results (Alston, 2010), it is necessary to return to the basics of these methods and to discuss their current limitations and challenges. The second set of methods includes what is generally referred to as “case studies methods”. If they are widely used in practice, they are also very heterogeneous. This section focuses on some approaches that aim at developing comprehensive and standardised approaches in order to strengthen the robustness of case studies. It also discusses the limitations and challenges of these approaches and concludes by a short discussion of the potential complementarity of standard economic approaches and approaches based on case studies.

Standard economic approaches

The economic impact of agricultural research has been analysed in a large body of the agricultural economic literature. The objective is to match the economic benefits with the research investment in order to calculate economic indicators of the impact of research such as internal rate of return or cost-benefit ratio. The original contribution by Griliches (1958) matched research benefits and investment at the level for a given innovation (hybrid corn). Such microeconomic analyses do not take spillovers correctly into account and most of the literature matches benefits and investments at a national or state level. Particular attention is paid to the (temporal) lag between the investments and the related benefits. This research does not address the mechanism through which research investment leads to some economic impact. Moreover the analysis focuses only on the economic impact of research and pays less attention to the other dimension of this impact. Hence these standard economic approaches are complementary to the case studies analysis presented in the next sub-section.

This sub-section describes the methodological framework used in standard economic approaches. It presents and discusses the main synthetic results from the literature based on several surveys and meta-analysis that have been published.

The methodological framework

Internal rates of return (or other economic indicators) are calculated from the comparison of research investment with economic impact. In the direct cost-benefit analyses, the calculation is made directly, assuming that all the observed economic impact is explained by the research investment considered. The alternative approach is to use an econometric approach in order to control for other factors that may explain the observed impact.

Direct calculation of the internal rate of return

Direct calculation is often made at micro-economic level, considering a single innovation. For example, Griliches (1958) examines the case of hybrid corn in the United States. The simplest way to compile the economic benefit is to estimate the economic gain for each user that adopts the innovation and then aggregate this gain over the set of adopters for the whole period where the innovation is used. In the hybrid corn case — as well as for many innovations at the farm level — the economic gain of each user is the product of the yield gain by the economic value of the production. This simple calculation of the economic gain is however imperfect because it ignores the impact of the adoption of the innovation on the market equilibrium within the agricultural sector. For example, an increase in yield leads to higher production, lower prices, and possible substitution between crops. More elaborated compilation of the economic gain can be developed, using partial equilibrium approaches to better represent the innovation impact on the market equilibrium and the distribution of the economic benefit from the innovator to the final consumer (with farmers as intermediates and possibly other intermediate actors).[7]

Based on the economic gain and research investments related to the innovation, the net present value of the research investment can be calculated as follows:

$$NPV = \sum_{t=0}^{T} \frac{Benef_t - Invt_t}{(1+r)^t}$$

The time horizon of *T* years covers the research period where the innovation is developed and the diffusion period. $Benef_t$ is the annual benefit and $Invt_t$ is the annual research investment. It is expected that research investment occurs at the beginning of the period and benefits occur after some delay until the end of the period. The internal rate of return is the value discount rate *r* such that the net present value is equal to zero. Other economic indicators such as the recovery period or cost-benefit ratio can also be calculated on the basis of streams of research investment and economic benefits.

This analysis at the level of an innovation can either be used *ex post*, *ex ante* or during the diffusion of the innovation. *Ex ante* analysis may be used to compare and select projects.[8] In practice, this is rarely the case because of the lack of data and the high uncertainty of research (Section 3).

7. See Alston et al. (1995) for a detailed presentation of the impact of various types of innovations in a partial equilibrium framework.

8. See Mutangadura and Norton (1999) for one example of application of such *ex ante* analysis.

One difficulty and limit of this approach at the micro-level is to define the set of research that leads to the innovation. Indeed research activities are not necessarily targeted towards a given innovation: they may be rather basic and/or have an impact on multiple innovations. On the one hand, if only the research that has been targeted to a given innovation is examined, then the impact of this research is over-estimated because the innovation would probably not have been possible without more basic research. The literature uses the concept of project fallacy to refer to this problem (Georghiou et al., 2002).

On the other hand, if all the research activities that have been contributing directly or indirectly are examined, then the impact of these activities is probably under-estimated because they have also contributed to other innovations. Hence there is a methodological limit to these micro-level calculations because the attribution of economic benefit to research investment is arbitrary. This attribution problem is related to the fact that research leads to the production of knowledge but this knowledge impacts multiple innovations and spillovers exist between the different sources of knowledge production.

This limit can partially be overcome by moving from a micro to a macro-level of analysis. Both economic benefit and research investment are measured at the aggregate level such as Nations or States. Economic benefit is measured by the economic surplus or productivity gain for the whole agricultural sector. Hence, the benefits capture the impact of all the innovation applied in the country. All the agricultural research investments in the country are assessed so that "internal" spillovers between research activities conducted are taken into account. However, due to difficulties collecting data, most of the analyses only evaluate public research investment and ignore private research investment. Andersen (2015) made one of the most recent direct calculations of rate of return (and other indicators) for all the US States. The precise approach used in this article is quite similar to the econometric approach presented below

In conclusion, direct calculation of research impact can be made both at the micro or macro level. Micro level analysis provides an interesting indication on the magnitude of the impact of various innovations. However there are some limits to these compilations depending on the way the economic benefit is calculated and also on the attribution of the benefit to some sets of research. Attribution problems are partially solved with macro-level analysis. However, by design, such an approach can only provide an indication of aggregated impact over large sets of innovations. One important limitation of such direct calculation (either micro or macro) is that the methodology does not allow controlling for other factors that may explain the economic impact.

Estimations based on econometric approaches

Econometric analysis has mainly been applied to analyse the impact of research at an aggregate level (e.g. country level). One interest of econometric analysis is to introduce different factors along with past research. One strategy is to estimate a production function that represents the agricultural sector. The annual production level is explained by different factors such as capital, labour, weather conditions and past research. Making a time series estimate of such a production function is problematic because of co-linearity among the explanatory variables. For this reason, agricultural economists generally estimate total factor productivity by weather conditions and past research.

This estimation requires a proxy that captures past research. The standard way is to define a knowledge stock variable (KS_t) that is the weighted sum of the past research investments:

$$KS_t = \sum_{t=0}^{T} b_k \cdot Invt_{t-k}$$

The knowledge stock at period *t* is the variable that explains the innovation level and hence productivity at period *t*. b_k is the lag between the research investment at some period and the level of the innovation *k* years after. Because of data and calculation limitations, it is not possible to estimate b_k. directly, hence the need to test different assumptions about the lag structure in order to retain the most suitable one. This lag structure is expected to have an inverse U shape. There is a minimum gestation period between when the research investment is made and when it leads to innovation. However, knowledge becomes obsolete after some time so that the research investment has no impact. Several forms of lag structure have been tested, and the most recent analyses retain a Gamma function[9] over 50 years (*T*=50). Two parameters are necessary to define this Gamma function and the most suitable forms are generally those with a peak at 25 years.

9. See Alston et al. (2009, chapters 8 and 9) for a recent and detailed discussion of research lag and the assumption on the functional forms of the lag structure.

To take spillovers into account, one generally distinguishes the local knowledge stock and the external knowledge stock. The local knowledge stock results from the research investment made in the zone (Nation or State) being studied. The external knowledge stock results from the research investment made elsewhere and that might impact productivity in the zone studied. In the analysis of the US case, Alston (Alston, 2010; Alston et al., 2011) considers State level research investment and the external stock of knowledge as the weighted sum of the investments made by the other States. The weights reflect the similarities of the agricultural production systems between the different States. In the analysis of the Australian case, Sheng et al. (2011) define the external stock of knowledge as US investment in agricultural research.

The estimated model explains productivity for a given year by the local stock of knowledge, the external stock of knowledge, some proxies that reflect the weather in the year, and possibly other control variables. The key parameter that captures the impact of research investment is the elasticity of the local stock of knowledge. This elasticity can then be translated in various economic indicators such as Internal Rate of Return or Cost-Benefit ratio. These economic indicators may differ from those estimated by direct calculation for two reasons: (i) econometric methods enable control for other factors, such as spillovers, that may explain productivity gains; (ii) econometric analysis estimates the marginal impact of the research investment while direct calculation estimates the average impact (Andersen, 2015).

Results and discussions

This section mainly focuses on results from econometric analysis at aggregate level. Many articles have been published covering various countries and periods as well as reviews of this literature (Alston et al., 2000; Alston, 2010; Evenson, 2001). The main lesson from these surveys is that the rate of return to agricultural research is high: Alston et al. (2000) synthesised more than 1 000 estimates and find the median Internal Rate of Return (IRR) is above 42%. The estimates are quite sensitive to the assumption on the lag structure. The most recent estimates have generally considered longer lags and this leads to lower values of the IRR. Note also that most of the estimates are made for northern and southern America as well as some Asian countries, but very few estimates are available for Europe, which may be due to limited data availability.[10]

Some of the recent literature challenges the relevance of using IRR as an indicator of the economic impact of agricultural research. Several authors suggest using a Modified Internal Rate of Return (MIRR) where it is assumed that only part of the economic benefit generated by the innovation is re-invested in research. Hurley et al. (2014) re-examined more than 2 000 estimates with a median IRR of 39% and calculated a median MIRR of 9.8%. This remains a high rate of return which the authors claim is more realistic.

Another important factor that has been discussed in the literature concerns the type of research. Several analyses distinguish public research and extension. It is shown that the impact of extension is generally lower than the impact of research and that spillovers are smaller for extension. The objective of the research should also be taken into account. In the United States, some agricultural economists make the distinction between productivity enhancing research and maintenance research. The objective of maintenance research is to keep production from decreasing. Research that targets pest resistance, invasive species, and adaptation to climate change (etc.) can be considered as maintenance research. Sparger et al. (2013) estimate that up to 40% of agricultural research is devoted to maintenance. Ignoring the distinction between these different types of research may lead to very different estimates of IRR or MIRR.

Another important issue is related to private research, especially in industrialised countries such as Northern America or Europe. Private research investment in these countries is higher than public research investment. However collecting data on private research investment is very difficult. As a result, private research is often omitted from the analysis, even for countries like the United States where the data sets are the more comprehensive (Fuglie et al., 2012; Fuglie and Toole, 2014).

In the literature there is little focus on the evolution of the nature and performance of agricultural research over time. Studies generally consider very long periods of about 50 to 60 years. This assumes that the lag structures and the elasticity of productivity with respect to the stock of knowledge are stable over time. However, it is clear from the observation of public research institutes, such as Land Grand Universities, the Agricultural Research Service (ARS) of the USDA, the French National Institute for Agricultural Research (INRA), and Wageningen University and Research (WUR) in the Netherlands that their position in the national research systems as well as their mission have been evolving over time. Since WW II, these public research

10. As part of the European project IMPRESA (The Impact of Research on EU Agriculture), different methods have been used to estimate econometrically the impact of research on agricultural productivity in European member states. Preliminary results were presented in Rome in November 2016 and final results will be made available at www.impresa-project.eu.

organisations have shifted from applied to more basic research and this evolution is related to the emergence of private research. Also increasing priority given to the environment and nutrition leads these organisations to devote more investment to these objectives, which correspond to the needs of society but do not necessarily create economic impact. Such changes challenge the standard economic approaches to evaluating the impact of agricultural research. One way to capture these broader objectives is to calculate the economic impact only on the share of research that is devoted to increasing agricultural productivity. Another and more ambitious strategy is to extend the analysis to the non-economic objectives, which requires at least having synthetic and robust indicators.

Approaches based on case studies

The use of case studies in impact analysis has been popular since the 1960s and 1970s when some government agencies wanted to understand the relation between R&D spending and economic growth (Bozeman and Kingsley, 1997). Case-studies are useful to analyse how and why a phenomenon occurs. They provide rich information, through a narrative, about the element which is evaluated (a programme, a public research organisation, a series of projects, or an innovation) and highlight the critical relationships that exist among the various variables, events, and actors. Case-study can apprehend the complexities inherent to the processes of impact generation. Case studies are also widely used to explore topics for which no strong theory exists. For Ruegg and Jordan (2007), another strength of case studies is their ability to put flesh on the bones of quantitative approaches, i.e. to provide elements of contexts and qualitative explanation about quantitative trends.

Case studies are subject to several limitations. Very often case studies are criticised for their lack of rigour, the provision of equivocal evidence, ambiguous views, and biased results rather than quantitative indications. Another problem is the ability of the methodology to assess in robust manner causality and to generalise results from cases studies. This critic may be addressed by multiple case analysis and cross-case comparison that would increase the external validity. Cross-case comparison can also be strengthened by developing an analytical framework to conduct the comparative analysis. Another recurrent critic is related to the time and resources needed to collect and analyse the data in a valid and reliable manner.

The remaining part of this section presents the more recent approaches using case studies to evaluate the societal impact of public research and outlines their main strengths and limitations.

A melting pot of approaches

Case studies constitute an alternative methodology of impact evaluation and encompass at least two main characteristics that make them depart from the standard economic approaches. First, case-study approaches include a more broadly conceived concept of impact and consider social, cultural, environmental, political and economic returns (Donovan, 2007, 2008, 2011). To capture the broader societal benefits, “metrics-only approaches are behind the time, and state-of-the-art evaluations of research impact combine narrative with relevant qualitative and quantitative indicators” (ibid. 176). The aim of these approaches is to provide a full picture of all types of societal impacts. The second characteristic of a case-study approach is to take account of the complexity involved in the process of impact generation from a wide range of academic research. One of their main assets is to show that impacts are generated by a network of actors that interact to create and use research results. These networks evolve over time in terms of types and number of actors involved, objectives, and commitments. These approaches are based on theoretical frameworks such as the system of innovation, evolutionary economics and the Triple Helix model, that provide dynamic insights to public research impact evaluation studies. They highlight impact generating mechanisms.

Various approaches evaluate the societal impact of public research

Public Value Mapping (Bozeman, 2003; Bozeman and Sarewitz, 2011) assesses the capacity of research to achieve social goals. It is not a classical evaluation method but a rather conceptual approach helping to understand the contribution of science within a network of knowledge producers generating societal impact. In this approach, scientific knowledge gains value through its use and not only through its commodification on a market. "Knowledge value collectives arise to generate, develop, and use scientific research” (Bozeman, 2003: 13). They involve government and private funding agents, end-users, wholesalers, equipment and other scientific resource vendors. Public Value Mapping considers outcomes such as environmental quality and environmental sustainability, health care, and provision of basic needs, e.g. housing and food. The following factors and mechanisms can be seen as analytical lenses and determine the social impact of research (Bozeman and Sarewitz, 2011): characteristics of knowledge produced by research, institutional arrangements and management affecting knowledge production and use (user-producer interaction, networking); and policy and political domains of knowledge production and use (political and legal context). Bozeman and Sarewitz (2011, p.1) argue that it is vital to have a deeper understanding of these factors to help science policy-makers in "making choices among competing paths to desired social outcomes". Public Value Mapping is a model encompassing a theoretical framework, a set of assumptions and procedures and is based on case studies. It has been applied in various domains such as climate science, nano-medicine and chemistry.

The Payback Framework was created to assess the outcomes of health research. It consists of two elements (Donovan and Hanney, 2011): a logic model consisting of stages and interfaces between the research system and the wider user environment, and various categories of research impacts. The logic model contributes to analysing the "story" of an innovation from topic identification, project specification, research process, and primary outputs of the research, to the various dissemination steps until the final outcomes. Various types of benefits are considered: academic benefits (publications, research reports, etc.), benefits for future research (development of research skills), benefits of policy and product development information, health sector benefits (improved health, improved equity in service delivery), and broader economic benefits. The dissemination and adoption phases highlight the role played by intermediaries and beneficiaries. Wooding et al. (2014) underline various factors associated with high and low impacts. For instance, researchers engaging with practitioners and patients to plan and organise their research projects are associated with projects with high scientific and broader impacts. Research which considers the pathways of translation and application of clinical research are associated with broader impacts. The way data is compiled facilitates comparative cross-section analyses essentially in terms of paybacks generated.

The SIAMPI (Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions between science and society) approach considers the "productive interactions" between researchers and stakeholders as central to creating research with any kind of impact (Spaapen and Van Drooge, 2011). SIAMPI focuses on the interaction process in order to identify the relevance of the research, and how it is adopted, diffused, and applied, or not. Productive interactions are exchanges between researchers and stakeholders (industry, public organisations, government, and the general public) involved in achieving societal impacts. The interaction becomes productive because stakeholders make efforts to use and apply the research results to generate impact. In this approach, interaction between actors is central (de Jong et al., 2014) and depicts the main mechanisms at stake in the impact generation process. Productive interactions might be direct (personal links between researchers and stakeholders may accelerate research uptake), indirect via information carriers (publications, patents) or financial (research contracts, funding). The process of interaction is complex and takes account of the evolution of the network structure, the diversity of actors, research fields and sectors. These interactions all influence the way societal impact is generated. The approach does not always make "a clear distinction between social impact and 'productive interactions' because the transition from interaction to impact is often gradual" (Spaapen and Van Drooge, 2011: 212). The case studies are compared on a cross-sectional analytical basis. The approach has been applied to Health, Information and Communication Technology (ICT), Nanosciences and social science and humanities (Molas-Gallart and Tang, 2011) and more recently to engineering, artificial intelligence and biomedical science (De Jong et al., 2014).

Methods developed to evaluate the societal impacts of public agricultural research

Most of the main Agricultural Public Research Organisations have developed methodologies based on case studies to evaluate the various types of impacts generated by their research results (see Chapter 3 for a focus on five organisations). Their results underline that the research conducted affects a wide range of stakeholders in terms not only of economic impact but also environmental, health, and policy impacts. In terms of quantification of broader impacts, EMBRAPA has developed an original method to evaluate environmental and social impacts (Rodrigues et al., 2010). Ambitec-Agro is a multi-attribute indicators system that allows calculating impact indexes for a given innovation. Their case studies are mainly narratives devoted to justify the various steps of the quantification procedures.

Among these various experiences, the analysis aiming at assessing the societal impact of research conducted by the CGIAR (Consultative Group on International Agricultural Research) research centres is interesting because it introduces the notion of "impact pathway" (Douthwaite et al., 2003; Walker et al., 2008). The Impact Pathway (IP) is a model based on identification of the different phases of impact generation, the actors involved, the flow of resources, and the progressive transformation of knowledge into outcomes and impacts. The model was designed as an applied assessment tool by consultants in the German Development Agency, GTZ (Kuby, 1999), and refined for inclusion in the international agricultural research framework to evaluate the research impact of the CGIAR. An impact pathway captures the different stages of R&D from the basic research inputs to the final impacts, including the different research outputs and outcomes for different types of users. Networks of stakeholders play dominant roles in the construction of research outputs and in the diffusion and adoption at multi-scale levels. Technological change is brought about by the formation and actions of networks of stakeholders in what essentially is a social process of communication and negotiation.

The methodology encompasses a diversity of impacts (economic productivity, social and distributional, environmental impacts). The concept of impact pathway has been used and adapted in several methods developed to evaluate the societal impacts of public agricultural research.

For example, INRA has developed the ASIRPA approach (Socio-economic Analysis of the diversity of Impacts of Public Agricultural Research) (Joly et al., 2015). It is based on an analytical framework which identifies the factors affecting the generation of impacts. The ASIRPA approach pays attention to the process of transformation that makes knowledge actionable and which allows to incorporate it in new products, new processes and new ways of doing things or governing. The analysis identifies the chains of translations that occur in the process. This vision is inspired by Actor Network Theory (Callon, 1986) that defines translation as a four-stage process: problematisation, interessement, enrolment and mobilisation.[11] ASIRPA is an approach based on standardised case studies. The analysis of each case unfolds a standardised outline and sketch three analytical tools: (i) the chronology allows identifying the main events, resources and actors, (ii) the impact pathway (based on an adapted version of the CGIAR model) helps understanding the role of INRA at each step of the process within the various networks of actors, and (iii) the vector of impacts is composed of a table and a radar exhibiting qualitative and quantitative evaluation of the various dimensions of impacts. These three tools allow reducing the inherent complexity of cases to underline the main determinants in a standardised schematisation. Standardisation allows systematic codification of the variables of the case studies and to perform cross-cutting analyses.

In his impact evaluation guide, the CSIRO (Commonwealth Scientific and Industrial Research Organisation) exposes its evaluation principles and process (CSIRO, 2015). The evaluators use an impact pathway to trace the causal relationship between inputs, outputs, outcomes and impacts. The main objective of the case studies is to expose a narrative aiming at calculating a cost-benefit ratio. CSIRO develops a more classical approach based on the definition of a counterfactual, an attribution coefficient and an uptake profile.

Impact Pathways have also been used in participatory analysis and labelled PIPA (Participatory Impact Pathway Analysis). The CGIAR has developed this method which supposes that project implementers, intermediary users, end-users and political actors meet during three days in a workshop to elaborate a common vision about how and when a project might generate various types of impacts (Douthwaite et al., 2007). This *ex ante* step should induce the various actors involved to be highly committed in a collective way to reach the impacts. A PIPA workshop usually produces a series of objects such as a statement of the problem, outputs, vision, network maps, a project timeline, a logic model and an impact narrative. PIPA is also a method used to monitor and evaluate impacts. It has been used in an *ex post* evaluation study by the CIRAD (French Agricultural Research Centre for International Development) (Triomphe et al., 2015). In this study, researchers used focus groups and workshops bringing together various actors involved in the evaluated project. Actors participate to the evaluation exercise at different stages of the study and with various intensities. Using a participatory method allows to collect rich data during various interviews, workshops and focus groups, and to develop a common vision of how the innovation process evolved. The results are also validated collectively during a final workshop taking place at the end of the study.

What lessons can be learned?

The above mentioned studies lead to the following major results. The delay between the beginning of the research and the first societal impacts are rather long and generally estimated to an average of 15 to 20 years (Bornmann, 2013; Joly et al., 2015).

Impacts are produced by a network of actors and this network evolves along the impact pathway. The network in downstream phases of the pathway is often different from the research network. Impacts are thus difficult to attribute to one isolated agent and neither is it the sum of actions deployed by each actor. Various approaches claim that it is necessary to shift from attribution to contribution analysis. Attribution is commonly used both to identify causal relations and to estimate quantitatively how much of an observed impact is due to the intervention of a given organisation (Avila et al., 2015). Attribution supposes that the different causes that produce a given effect are additive, which contradicts what is observed in complex ecosystems of innovation, namely the key importance of synergistic (non-additive) interactions. Therefore, attribution may usefully be replaced by a contribution approach (Joly et al., 2015). Detailed analysis of the roles of actors in the process of impact generation makes it possible to identify the contributions made. Focusing on productive interactions

11. According to Callon (1986), the translation process follows four stages: 1) Problematisation: an actor (in general a researcher) analyses a situation, defines the problem and proposes a solution; 2) Interessement: a series of processes by which the researcher seeks to lock the other actors into the roles that have been proposed for them in that programme; 3) Enrolment: a set of strategies in which the researcher seeks to define and interrelate the various roles allocated to others. A new network of interests is generated; and 4) Mobilisation: a set of methods to ensure that actors involved in the process are able to represent the collectives to which they belong.

(Spaapen and Van Drooge, 2011) or configuration, actor networks, the role of intermediaries, the focus can be shifted to the contribution of specific actors, and the exchange of knowledge and expertise by the various stakeholders.

The societal impacts are determined by a set of mechanisms specific to each method and dependent on the theoretical framework used. SIAMPI focuses on the interaction between actors, Public Value Mapping looks at the institutional and social arrangements and settings, while ASIRPA considers synchronic and diachronic translation mechanisms.

The distribution of impacts is highly skewed as pointed out by several authors (Cunningham et al., 2013; Georghiou, 1999; Maredia and Raitzer, 2006; Molas-Gallart et al., 2002; Scherer and Harhoff, 2000: 562), who note that "researchers who seek to assess the success of government technology programmes should focus most of their effort on measuring returns from the relatively few projects with clearly superior payoffs". This means that *ex post* assessment allows concentration on a limited number of cases.

What are the main methodological challenges?

A number of methodologies for measuring different dimensions of impact are available, but current approaches do not provide a universal metric for each of the main dimensions of impact, or the resources for producing it. Indeed, universal metrics are available for scientific and economic impact, and to a lesser extent health impact (Hanney et al., 2007; Kamenetzky, 2013). For the other dimensions (environment, public policy, social), *ad hoc* measures are to be relied on.

If all the above mentioned methods recognise a variety of impacts, it is not always easy to separate one type of impact from another. According to (Salter and Martin, 2001) economic and non-economic impacts might overlap. Some impacts might need more time than others to be generated. For instance, policy impacts might be considered as an intermediate impact that could generate other types of impacts once the policy is implemented. There is no general model about the interconnectedness of impacts.

The evaluation of societal impacts lacks an accepted and standardised framework with appropriate datasets, criteria, indicators and methods (Bornmann, 2013). Some techniques and methods have been developed but they still lack validity and robustness especially in terms of measurement. For Bozeman and Youtie (2015), this is mainly due to the youth of this research area as compared to economic impact evaluation methods that started more than 60 years ago and bibliometric approaches more than 40.

Standardisation is also an issue within each of the developed methods. The above mentioned approaches possess some degree of standardisation in the sense that they often present cross-case analysis. However, their low degree of standardisation does not allow conducting more aggregated analysis.

What are the main limitations?

If case-study approaches encompass advantages linked to their ability to deal with complexity and impact generating mechanisms, they are also subject to some critics. They often lack objectivity and quantification, and are expensive and time-consuming. Case-study approaches often provide a set of detailed stories, each representing a specific situation within a wider set of situations where a PRO operates. According to Bornmann (2013: 226), “Case studies do not permit generalisations to be made but they do provide in-depth insight into processes which resulted in societal impact and therefore lead to a better understanding of these processes (Rymer, 2011)”. The literature recognises a general problem of aggregating the richness of case studies. Some PROs (EMBRAPA, 2015) provide aggregated data, very often in the form of a single figure of one type of impact (e.g. economic) based on the summation of quantified impacts with a comparable unit. Such a figure is easy to communicate. The downside is that this communicability may hide the knowledge characterising the complexity of the various pathways to impact. Other approaches (Payback Framework and SIAMPI) develop indicators that allow cross-case comparisons. All these approaches hardly consider a reduced number of impact pathways at the level of an organisation, and do not generate data to highlight these impacts. Referring to the R&D Value Mapping approach, Kingsley et al. (1996) underline that the quantification of elements across cases can lead to generalisable data.

The ASIRPA method made a first step towards aggregating the data by developing a typology of impact pathways. The method consists in codifying the data of each standardised case study into a database. The data are organised around the different steps of the impact pathway. Each step is detailed by a set of variables characterising the role of actors, intermediary products, and impacts. To build the four ideal-types two discriminating independent variables were created, representing the roles of INRA and the stakeholders within research networks (productive configuration) and adoption networks (the diffusion process). For each type, a cross-cutting analysis is conducted to further characterise each ideal type impact pathway (specific translation mechanisms, critical points, research and adoption networks, research outputs, and impacts).

Towards integrated approaches?

Standard economic approaches are complementary to those based on case studies

Evaluating the various impacts generated by agricultural public research requires the mobilisation of different methodologies. Evaluation should be considered as a multifaceted exercise that should provide relevant information to various stakeholders who are involved in different types of decision processes. There is a need to design a system of evaluation based on different approaches that can be integrated in a comprehensive management system and political discourse. This follows the recommendation of Irwin Feller and colleagues:

> *"The standard for future action is not a single flawless study that satisfies all structures, but rather a succession of studies that critique and improve upon each other as they collectively advance toward norms of formal evaluation methodology". (Feller, Glasmeier and Mark, 1996: 318, quoted in (Feller, 2003: 26).*

The two sets of approaches presented in previous sub-sections should thus be considered as complementary.

Quantitative methods used to evaluate economic impacts are usually based on aggregated data (Section 2.2). Such approaches quantify the economic benefits, and are useful for justifying existing public R&D programmes at the level of a Nation, a region (or State) or an industry. However, they do not add to an understanding of the process of generating economic benefits. Approaches based on case studies are instrumental for understanding the impact generating mechanisms, the beneficiaries and co-innovators' roles, etc. Such approaches are useful for learning purposes and the lessons drawn from these analyses can be used to anticipate critical points and the way to overcome them in future impact pathways. But in general, they do not produce an overall assessment of the research economic efficiency.

Both types of approaches could complement each other rather than be opposed. Some assumptions of standard economic approaches such as the adequate lag structure could be empirically founded by case-study analysis. Concerning economic impacts, efforts could be devoted to render measurement coherent between the two approaches to facilitate aggregation exercises. Case studies stay ahead of quantitative approaches in terms of the variety of impacts considered but are usually rather weak in terms of measurement. As presented in Section 3, most of the organisations integrate cost-benefit analysis in case-study approaches. Hence, both types of approaches are complementary and it is crucial to develop comprehensive approaches in order to provide more robust and original measures. Such approaches must have the ability to assess research impact at various scales: project or programme, organisation or country level.

However, there are some important differences that should not be overlooked

Quantitative methods used to assess the economic impact or research can hardly escape from a simple, reduced and rather stable representation of the research processes. Moreover, they focus mainly on the impact on productivity and social welfare and generally neglect distributional issues or possible side effects that may lead to negative impact. They consider knowledge, resources and projects as additive with the objective to **attribute** the economic specific organisations, projects or geographic region. The main objective of these approaches is linked to accountability issues and budget allocation.

On the contrary, RIA methods recently developed to assess broader impacts hypothesise that innovation is complex, interactive and conducted in systemic contexts, that there is an evident shift from mode 1 to mode 2 production of knowledge (Gibbons et al., 1994), and there is a shift from a pure competitive frame to the need to address societal requirements. In these system-oriented approaches, actors **contribute** to generate societal impacts within complex and evolving productive configurations or networks. The main aim of these approaches is to understand the mechanisms and processes generating impact, and to support policy learning.

The design of integrated approaches should not overlook these differences. Indeed, these approaches draw on interdisciplinarity and epistemic pluralism, which should not lead to analytical complacency or methodological amateurism, but on the contrary to reinforced robustness through the possibility of performing assessment from different perspectives.

3. Practices of ARIA in some research organisations

There is an important gap between methods of RIA published in academic journals and those actually used in practice (Shapira and Kuhlmann, 2003). Such an assessment, however, is based on anecdotal evidence. So far, RIA in practice has not been systematically studied. This section contains original data on ARIA in practice in order to grasp the current situation and to see how the gap between s is being dealt with.

It focuses on the practices of five public research organisations:

- The US Department of Agriculture (USDA)
- The French National Institute for Agricultural Research (INRA)
- Commonwealth Scientific and Industrial Research Organization (CSIRO) in Australia
- The Brazilian agricultural research organization, EMBRAPA
- The Consultative Group on International Agricultural Research (CGIAR)

The study focuses on the impact assessment practices of these research organisations and not on the impact of agricultural innovation systems as such (OECD 2012, 2013). The contribution to impact of these organisations increasingly depends on interactions with other actors such as: universities, other public research organisations, extension services, private companies, etc. Research impact assessment approaches allow these organisations to identify the respective contribution of the different actors.

The selection of these organisations is based on previous information available from the literature. CSIRO and EMBRAPA were chosen because of their long-established practices in impact assessment. USDA and CGIAR were selected because of their focus on evaluating programmes. INRA is a newcomer in research impact assessment. These organisations present special features and interests for monitoring and impact evaluation. Because of the limits of this study, the focus is on these organisations although other institutions also play an important role in agricultural research at national or international level. This study is not expected to be representative of how other institutions function. But because these organisations are very different, it allows wide range of diversity of practices to be observed.

This section draws on two main information sources: descriptions on the organisations' web pages, and additional information from one or two sources for each of the selected organisations. The five selected organisations are briefly introduced and their RIA experience presented. The second section is devoted to a cross-cutting analysis of RIA in practice.

Detailed descriptions of the approaches implemented in these five organisations, along with some insights into the challenges they are facing in terms of methodologies and implementation, are available in the Appendix.

An overview of selected organisations

US Department of Agriculture (USDA)

The Agricultural Research Service (ARS) is the US Department of Agriculture's chief research in-house agency. It employs more than 2 000 scientists in more than 90 laboratories throughout the country. ARS is therefore the main focus of this study. The Economic Research Service (ERS), which is only marginally concerned with research impact assessment, but provides studies on the productivity of agricultural research, will also be looked at briefly.

The main tool for evaluating all federal agencies, including research agencies, is the Government Performance and Results Act (GPRA). USDA is also a partner in the STAR METRICS consortium (Science and Technology for America's Reinvestment: Measuring the EffecT of Research on Innovation, Competitiveness and Science) between US federal science agencies and research institutions to document the return on investment, research impact, and social outcomes of federally-funded R&D.

ARS is established on the basis of a five-year strategic plan (presently strategic plan fiscal year (FY) 2012–17, revised in 2014). ARS research is organised into 17 National Programs. These programmes serve to bring co-ordination, communication, and empowerment to approximately 750 research projects. Programme performance against targets is monitored annually. Each National Program Team (NPT) prepares an annual report featuring the National Program major accomplishments. At the end of each national programme's five-year cycle, an accomplishment report is prepared by the NPT, and a retrospective Review by an external panel is carried out.

INRA

The French National Institute for Agricultural Research (INRA) employs 8 300 people (of which 1 800 researchers and 2 500 engineers) and is organised into 13 scientific divisions. INRA's scientific priorities are set every ten years in a strategic guidelines document (presently 2010-20). Before 2009, while science quality was monitored and evaluated there was no system in place to evaluate societal impact. Some reports on innovation were commissioned by senior managers.[12] The communication unit also maintained a database of salient facts that gathered narratives and data regarding more than a thousand innovations since 1996.

INRA's move in developing a comprehensive system for the assessment of its socio-economic impact dates from 2009. INRA managers funded a research project to design a methodology for the Analysis of the Socio-economic Impacts of the Public Agricultural Research (ASIRPA) that could be tested by INRA, the project was launched in 2011. In 2013-15, after a two-year pilot phase, half of INRA's scientific divisions were involved and tested the approach in real assessment conditions. The resulting impact assessment system was officially institutionalised in 2015. In parallel, in 2014, INRA's economists calculated the Internal Rate of Return of the French agricultural research, using standard economic methods described in Section 2 of this report.

CSIRO

In Australia, the Commonwealth Scientific and Industrial Research Organization (CSIRO) employs over 5 000 people, and is funded by the Department of Industry, Innovation and Science. CSIRO is organised into three Lines of Business (LoB)[13]: 1) Impact Science; 2) National Facilities and Collections; and 3) CSIRO Services. Within its Impact Science LoB, there are nine Business Units (BU, previously known as Flagships), focusing on the largest issues facing the nation across key sectors: Agriculture, Health and Biosecurity, Data 61, Food and Nutrition, Land and Water, Mineral Resources, Manufacturing, Energy, and Oceans and Atmosphere.

Since the 1990s and until 2011, the CSIRO was funded by the Federal Government through a quadrennial funding cycle. To prepare for each new funding phase, and in order to assess the performance achieved during the previous period, reviews were conducted by external consultants (the last review of this type was issued in 2014). These reviews drew on representative case studies to calculate the overall organisational economic value created by CSIRO during the period, calculating a total internal rate of return. The quadrennial funding cycle was replaced by a four-year rolling funding agreement process, which requires CSIRO to report yearly against key performance indicators linked to their strategic and corporate plans.

The impact reviewing procedure was renewed with the objective of developing a common framework for evaluating the different flagship programmes of CSIRO. Another objective is to link *ex post* assessment with monitoring of emerging impact. In 2010, it was decided to launch a consistent, organisation-wide approach to impact assessment and management, the Impact 2020 project. This impact-based system includes standardised case studies for *ex post* societal impact assessment and a monitoring database.

EMBRAPA

The Brazilian Agricultural Research organization (EMBRAPA) employs 9 800 people (of which 2 400 researchers). EMBRAPA is organised into 42 product-based, basic themes or eco-regional research centres. EMBRAPA's strategic plan sets large missions and goals over a period of 20 years. The first assessment of the impact of EMBRAPA's technologies started in the middle of the 1980s, as part of a national effort for impact assessment (Avila et al., 2015). Studies calculating economic surplus or econometric analyses were performed *ex post*, either at the level of commodities, grants, programmes, regions or the EMBRAPA as a whole. During the 1990s the majority of the econometric studies were performed at lower levels (research centre or commodity-oriented research) and based on local initiatives, training requirements (Ph.D. or M.Sc. thesis), or to meet the demand of large international funders (Inter-American Development Bank and World Bank).

The impact evaluation process was renewed in 1997. It led to the publication of an annual issue of an "EMBRAPA social balance" report tracking the dissemination and economic impact of 110 technologies and 220 cultivars. The report is based on case studies compiled in a yearly updated database. The aggregation of

12. See for example *Les chercheurs et l'innovation: regards sur les pratiques de l'INRA,* 1998, Paris : Quae.

13. www.csiro.au/en/About/Strategy-structure/Operating-model.

data allows calculating the Internal Rate of Return of EMBRAPA every year. Starting in 2000, this assessment and annual monitoring encompassed social and environmental impacts, calculated using the EMBRAPA's Ambitec method (Rodrigues et al., 2010).

CGIAR

The Consultative Group on International Agricultural Research (CGIAR) was created in 1971. It counts 15 independent research centres, with a total of 8 000 staff (researchers and technicians). In 2010, the CGIAR moved away from centre-based programming, to cross-centres CGIAR Research Programmes (CRP).

The CGIAR Consortium co-ordinates activities across research centres and is accountable for how the donors' funds are used. Since 2010 a Strategy and Results Framework (SRF) provides common goals to be jointly achieved by CGIAR centres through 16 CGIAR Research Programmes (CRP). Before receiving funding, CRPs set out their expected achievements and provide verifiable targets against which progress can be measured and monitored. The second CGIAR's 2016-30 strategic and result framework (SRF) was approved by CGIAR's Consortium Board in May 2015. At the system level CGIAR must contribute three goals (System Level Outcomes, SLO's) of the Sustainable Development Goals (SDGs) outlined by the United Nations.

There is a long history of evaluation in the CGIAR, with the main responsibility residing with the former Science Council (now the Independent Science and Partnership Council-ISPC) which organised the independent external review of CGIAR Centres. Many of these assessments were performed by university researchers, at the request of individual funders, which resulted in highly heterogeneous methods and quality. In parallel, independent reviews of the CGIAR as a whole were undertaken approximately every ten years. The independent Standard Panel for Impact Assessment (SPIA), a sub-group of the CGIAR ISPC, was created in 1995 to advise on donors' fund allocation by performing *ex post* meta-evaluation of the economic impact of the CGIAR's research. Historically, SPIA performed global modelling, using IFPRI models, to estimate Internal Rates of Return and addressed impact evaluation at a centre's level, or on broader thematic areas. Over the years, the CGIAR's research agenda expanded to address natural resource management and conservation issues. The Independent Evaluation Arrangement (IEA), created in 2012, has a central mandate for external evaluation of all parts of the reformed CGIAR system. The current structure for evaluations in CGIAR lays out a system of multi-level evaluations: CGIAR Research Program (CRP) and System (portfolio of CRPs) levels (e.g. the CGIAR as a whole).

RIA in practice: A cross-cutting analysis

This cross-cutting analysis presents RIA in practice in a rather linear way. It presents the purposes of RIA, the way RIA is designed, implemented, and the way its results are actually used. This presentation does not do justice to the complexity of the systems of evaluation. Tables 1 to 4 and the Appendix provide details on RIA practices in the selected organisations. This cross-cutting analysis outlines the similarities and differences and identifies styles of RIA.

Purposes of impact evaluation

All five public research organisations (PROs) assign multiple purposes to their impact assessment approach: upwards accountability (to funders, or the public at large), internal organisational learning (lessons to improve effectiveness in producing impact) and internal culture (engaging and building capacity of researchers). Accountability objectives usually require an external evaluation or validation, which can go against the objectives of involving internal staff and develop internal evaluative capacities, and can even prompt researchers to report overestimated data about their impact thereby limiting internal organisational learning. The value of the evaluation system can therefore be characterised according to the balance between these three — sometimes divergent — interests.

The assessment approach concerns multiple levels of the organisation: national or international research programmes (CGIAR, CSIRO, USDA), scientific divisions, centres, or business units (CSIRO, CGIAR, INRA), the organisation as a whole (CSIRO, INRA, EMBRAPA, USDA). The evaluation of single projects is generally not the objective of the assessment approach.

In terms of timing, the approaches are designed to be inserted in an external assessment calendar, even if internal learning is claimed to be a chief motivation. All evaluation systems (with the exception of INRA) are linked with a monitoring system. In our sample, the evaluation of programmes is linked with the end of a funding phase (USDA, CGIAR), and in these two cases the evaluation of the research *ex post* impact, which goes beyond the timeframe of a programme, is demand-driven and depends on the availability of external funds. Only INRA, CSIRO and EMBRAPA have a planning for reporting *ex post* impact. Only EMBRAPA regularly updates studies concerning the *ex post* impact of their research.

Evaluation designs

All PROs have set guidelines to standardise the way societal impact should be assessed (CSIRO being the most recent). In recent years they have explored new ways to evaluate their impact either by implementing research projects (INRA, USDA) or by reviewing their monitoring or evaluation systems (CGIAR, CSIRO, EMBRAPA). Some of these developments have not been followed through (example ERS guidelines for USDA), or are just being implemented (CSIRO). Recent monitoring systems (CGIAR, CSIRO, USDA) are integrated into a whole theory of change, creating a link between activities, outputs and outcomes, and trying to track progress towards societal impact at the end of each funding phase. *Ex ante* assessments (USDA, CGIAR, CSIRO and soon EMBRAPA) are part of a cycle linking monitoring and evaluation. *Ex post* assessment approaches are often goal-free, with no incentive to relate impact to previously set targets, contrary to monitoring incentives which increasingly account for the progress in achieving expected impacts. This “targeted” impact assessment approach faces a challenge related to time lag since the timespan imposed for the evaluation is often too short to observe time-distant societal impact.

All the organisations assess their economic impact, and most of them account for environmental and other broader social impact. EMBRAPA, INRA, and to a lesser extent CGIAR, developed their own methodologies to assess other non-economic impacts (institutional, political, sanitary, territorial).

In terms of methods, some PROs (INRA, EMBRAPA, and CGIAR) combine aggregated econometric approaches and case-study based approaches, with some attempts to complement each approach by the other. EMBRAPA demonstrated the consistency of this dual approach by comparing the sum of economic surplus of its hundred technology cases to the IRR calculated with aggregated data at EMBRAPA’s level. Econometrics is also performed at the case level: EMBRAPA calculates an internal rate of return by technology, CSIRO calculates cost benefit analysis of cases of varied sizes.

All organisations (with the exception of USDA) perform case studies although in different ways: some are relatively short illustrative narratives (EMBRAPA) or quantification of the various steps of the impact pathway (CSIRO) surrounding the quantification of impacts or the calculation of cost-benefit analysis, while other case studies encompass rich qualitative narrative including network analysis, changes of context and practices, in addition to the characterisation of impact (INRA). The approach implemented identifies the contribution of all the actors involved in the different phases of the impact pathways. PROs attempt to quantify impacts, often through monetisation (CSIRO, CGIAR). There are some attempts to quantify the impact with physical indicators specific to each impact dimension (INRA, EMBRAPA, CSIRO) and to try to align the desire for metrics and the need for meaningful analysis of achievements.

The quantification of impact is often assorted with a strong focus on attributing a share to the PROs research effort (CSIRO, EMBRAPA, CGIAR). Examples gathered from this study show that the rules to decide on the attribution shares among a network of research and diffusion partners are unclear (CSIRO, CGIAR, USDA-ARS). Furthermore the objective to achieve a quantitative attribution rate can lead to a bias in the case selection. To facilitate the calculation of the attribution rate, PROs are tempted to select recent cases for which memory is fresh and data are more easily available. They may also select smaller cases in which few external stakeholders have taken part rather than the much larger returns from the provision of international public goods, for lack of a credible scenario of attribution to the PRO supported-research (CSIRO, CGIAR).

The tendency towards centralised monitoring and evaluation systems (CGIAR), or standardised guidelines for evaluation (INRA, CSIRO, EMBRAPA), answers the managers’ and funders’ desire to increase the value of local evaluations and produce information on global impact. Building an overarching analysis of cases (aggregation, or cross-analysis), may consist in summing up the impacts yielded by a set of cases with the view to deliver a message accounting for the intensity (EMBRAPA, CSIRO) and diversity (INRA, EMBRAPA) of the impact of the organisation. It may also consist in analysing the impact generation steps and mechanisms with the view to provide organisational learning (EMBRAPA, INRA). Aggregation is often performed at the expense of detailed meaningful information. For example, the process used by EMBRAPA to aggregate its environmental impact relies on calculating the average of environmental impact scores (technical local performance ranging from -15 to +15) of a hundred technologies. Yet this average does not account for the level of adoption of the technologies. Aggregation entails a standardised methodology to study comparable cases, which is lacking in highly decentralised implementers such as CGIAR Research Programmes. Aggregation also requires a rationale for selecting representative cases as compared to the objective pursued. The rationale for case selection is barely detailed in the information provided by the PROs considered.

Implementation

All PROs are very concerned with the external credibility of their monitoring and or evaluation system, and the impact they report. Measures to promote external credibility include open calls for proposals to develop methodological supports (CGIAR), call for tender for consultants to apply standard guidelines (CGIAR, CSIRO), external validation of the case reports (CSIRO, INRA). CSIRO organises workshops with stakeholders in order to collect data and set its attribution shares. USDA-ARS review panels made up of outside experts (academics, stakeholders, and government) provide feedback on the programmes. Still, participatory approaches are not reported for setting the terms of reference of the assessment.

Examples gathered from this study show a gap between the theoretical method designed by PROs to assess their impact, on which they communicate, and the approach they actually implement. This gap may be explained by budget constraints (this problem was mentioned for CSIRO, CGIAR), management issues (CGIAR), or time delays to implement recent management changes (CSIRO, CGIAR, INRA).

Despite the interest demonstrated by PROs to assess their societal impact, very limited means (in terms of staff and budget) have been invested so far to implement the approaches designed.

Budget constraints limit the number of internal staff dedicated to perform impact case studies (INRA), and the possibility to hire consultants to do so (CSIRO), limiting in turn the number of cases and their representativeness at the level of the PRO. Managers from CSIRO and CGIAR also reported a problem of timing in the allocation of budget for impact evaluation: considering average impact generation lags, the *ex post* impact assessment of a programme is to be performed several years after the programme has ended, while the programme budget is discontinued.

Skills are also lacking. CGIAR research programmes lack competent staff for establishing the baseline or counterfactual to implement routine result-based management. CSIRO claims to be struggling with finding skilful external consultants to perform robust cost-benefit analysis. Internal as well as external capacity building needs time, and some institutions are developing training courses on impact evaluation (CSIRO, INRA). Management issues also include coordination efforts, a crucial challenge for CGIAR to harmonise methods, incentives and capabilities across its programmes, ensure quality of data and consistency of reporting, and solve present heterogeneous assessments that are detrimental to credibility.

Producing robust information on impact is costly (EMBRAPA). In terms of management, a limited availability of evaluation staff implies a greater involvement of non-specialist researchers. The identification and investigation of a case study relies on their willingness to take part. Resistance based on insecurity, fear of being monitored, or unwillingness to be evaluated increases the selection bias. This in turn limits the number of cases studied, the potency of the case-study-based approach to account for the diversity of the missions of the PRO (CSIRO), and therefore the effectiveness of the evaluation system to teach useful learning for internal management improvement (CGIAR).

The implementation of evaluation models is also hampered by the lack of streamlined and workable systems for information gathering. Most PROs relate their difficulty to implement their assessment framework because collected data are of poor relevance (CSIRO, INRA) or not locally available. For example, the framework designed to evaluate international CGIAR programmes can conflict with the availability of the data produced at the level of associated national research centres. Along similar lines, confidentiality can also be a problem (INRA and CSIRO struggle for accessing food sector private data). Again data availability orients the selection of impact studies towards “doable” cases, thus increasing bias.

A last reason to explain the theory-practice gap may be related to the delays required for new knowledge to disseminate. CGIAR for instance claims to be lacking operational tools like proxies for outcomes or methods for attributing impact shares, while they consider that robust theories on impact pathways do exist at CGIAR. Since the implementation of a revised assessment approach was decided only a few years ago, too few case studies have been produced so far to allow for a scaling-up and generalisation of learning at the CGIAR system-wide level.

Use of the results of impact evaluation

To what extent is evaluation used to communicate on impacts, or as support for decision-making? The findings of this study show that research impact assessment affects PROs management practices in a variety of ways, even if paradoxically, they make only limited uses of the results yielded so far.

The main use that all the PROs effectively make is demonstration to stakeholders. While INRA, USDA and EMBRAPA seem to build accountability on the motivation and interest of a diversity of stakeholders, CSIRO and CGIAR prioritise reporting on the good use they make of public funds. Information is lacking on how CSIRO, CGIAR, or USDA ground their funding allocation decision on the basis of societal impact assessment. In some instances, experts reviewing research programmes suggest that the objectives of influencing internal budgetary choices and priorities may be poorly served by present monitoring and evaluation systems (CGIAR). Similarly, institutional learning based on *ex-post* assessment seems limited, which is not surprising since the case-study methods used in the PROs considered do not go very far in terms of understanding of the impact generation mechanisms (except INRA). Learning objectives seem to be chiefly achieved through monitoring approaches, and may concern low-level tactical topics, rather than strategic higher level issues.

Conclusions

The fact that all five organisations of this study have recently made serious attempts to improve the way they evaluate their societal impact shows this has become an important consideration. Credibility is important to PROs and the five organisations have looked for methods that would combine excellence validated by their scientific peers and effectiveness in expressing outcomes or impacts for a specific audience. PROs have multiple ambitions for impact assessment approach: upwards accountability (to funders, or the public at large), internal organisational learning (lessons to improve effectiveness in producing impact) and internal culture (engaging and researchers' capacity building). However there is a gap between PROs ambitions for their RIA systems, and what they actually apply, which can only be partly attributed to a shortage of funds or staff capacity.

Accountability to funders (Treasury or Donors) is clearly an important driver for institutionalising impact evaluation and monitoring systems. One area where funders exert influence is their demand for quantitative targets and metrics for outputs, outcomes and impacts (USDA, CGIAR, CSIRO, EMBRAPA). The study suggests that in such instances (CGIAR, EMBRAPA) this objective may be served better than institutional learning on how to produce impact or capacity-building benefits. RIA systems build organically on the PRO organisational structure and operate in rhythm with existing processes; they will differ if funding comes through centres or programmes. This study suggests that organisations funding programmes tend to focus on collecting information at the various levels of change-activities-outputs-outcomes, more than joining the links to understand the dynamics for impact generation that are important for institutional learning.

The belief that learning and accountability goals can be mutually accommodated is widespread, but this study shows that it is not always so evident. From the interviews it appears that learning, and capacity building, are two important motivations for PROs' staff (CGIAR, CSIRO, INRA), however it is also evident that the accountability objective is of more universal interest for funders. Rhetoric is also widespread about the importance for an organisation of discerning what is not working. But the question arises that, for an organisation, exposing flaws or weaknesses could come at a cost in terms of reputation or future funding (CGIAR, USDA). This is especially the case when there is a sense that decisions in terms of funding flows would be immediately associated with such information, which can be a common case when funds come through programmes. In this situation immediate operational and tactical learning will be more readily addressed. Systems that introduce more distance between evaluation and funding decisions (EMBRAPA, INRA) leave more opportunities for strategic learning.

There is a tendency to construct RIA systems around accountability to the funders' viewpoint, with the objective of institutional and staff learning coming as an afterthought. This in turn limits the usefulness of RIA, as evaluations may be more directed towards considering how existing programmes and strategies can be delivered than calling this strategy into question. It is important to consider other dimensions, such as analysing the PROs contribution to impacts within the innovation network of actors, in order to inform strategic decisions.

Table 1. Purposes of impact evaluation

Organisation	Intended use of impact assessment (and Audience targeted) *Accountability to donors and the public, organisational learning, internal capacity building and culture of impact*	What level is intended to be evaluated? *Single project or innovation, Programme, Organisation, National research system*	Impact evaluation planning: *Timing of evaluation (regularity), internal or external, relation with monitoring*
CGIAR	Accountability to donors through monitoring and the assessment of CRPs between two funding phases. Accountability to a wider audience through the drafting of impact briefs. Organisational learning for CGIAR centres, programmes, donors through methodological development by SPIA and cross-CPR evaluations by IEA. Internal capacity building by IEA.	Multilevel: CGIAR system wide (impact), CGIAR multicentre programmes (CRPs monitoring and evaluations), research centres efficiency, specific themes or issues (demand –driven impact study).	CGIAR system-wide evaluation (every ten years). CGIAR multicentre programmes (CRPs monitoring and evaluation), in relation with CGIAR Strategy and Result Framework (SRF), and programmes funding cycles. Organised through IEA: four-year Rolling Evaluation Work Plan. Other topics and issues: demand-driven (funded by donors) through SPIA or other. Programme management: *ex ante*, monitoring and end-of-funding phase evaluations for CRPs.
USDA-ERS	Accountability to partners and Federal Government regarding the benefits of investments in US agricultural research. Organisational learning: for USDA-ARS some work on methodological issues regarding impact evaluation.	National innovation research system	Isolated studies.
USDA-ARS	Accountability to partners and Federal Government regarding the benefits of ARS-funded research systems programmes, through monitoring and evaluation.	17 national programmes (NPs). ARS through the contribution of its NPs to USDA objectives.	Evaluation is linked with a five year planning and evaluation cycle for NPs. NPs Annual monitoring and end of funding review provide information for new funding cycle and accomplishments at the level of ARS.
CSIRO	Accountability: value for money for Treasury. Accountability: demonstrating impacts to stakeholders and wider audience Organisational learning for CSIRO managers through monitoring to track impact. Combining internal capacity building and accountability objectives for CSIRO researchers and managers: seeking a cultural shift to account for societal impact in the management of research.	Business Units (B.U) through their programmes: impact statement to monitor future impact. CSIRO as a whole through case studies on different objects (from a research field to individual projects).	Continuous process of studying cases, integrated into annual reporting of CSIRO and external assessment of Business Units (every four years). Monitoring system for impact statement achievement of Business Units.
EMBRAPA	Accountability for Congress and international donors through performance report. Advocacy: Annual social report for stakeholders. Organisational learning: EMBRAPA's researchers: feedback on case studies. Decision-making: monitoring and expected impacts may influence researchers' careers and funding allocation choices.	Case studies on successful technologies and cultivars (3 technologies monitored for each research centre). EMBRAPA as a whole through the aggregation of cases and the calculation of EMBRAPA's IRR.	Annual update and monitoring of diffusion data.
INRA ASIRPA	Advocacy and learning (government). Accountability and advocacy: demonstrating impact size and diversity of INRA research to stakeholders and a wider audience. Organisational learning and capacity building (INRA managers and researchers).	National agricultural research productivity through the calculation of Internal Rate of return. Innovations (very diverse). INRA Scientific divisions. INRA institute: Aggregation of cases.	IRR: single study (2015). Linked with five-year external evaluation cycle (only for volunteering scientific departments) No relation with monitoring at this stage.

Table 2. Evaluation design, methods and methodology issues

Organisation	Existence of guidelines at the level of the organisation	Timing of assessment	Impact dimensions assessed	Theoretical framework (Goal-free evaluation vs expected impact, qualification of pathways and networks, attribution of impacts)	Methodologies : econometric, case studies, standardisation of cases, narrative	Measures , monetisation, indexes, counterfactual Aggregation
CGIAR: CGIAR system-wide	Counselling on methodologies by SPIA but not unified.	*Ex post*.	Mostly economic, a few environment, social and health.	Evaluation of expected impacts, as compared to programme objectives. Programme theory, impact pathway, theory of changes.	Socio-Econometric (Net Present Value and Benefit Cost Ratio). Case study with cost-benefit ratios for environmental impact.	Monetised, search for counterfactual. Cost-benefit ratios + specific indicators in development. Calculation of IRR of large innovations in comparison with CGIAR funding.
CRP programmes	Evaluation guidelines by IEA. Harmonised monitoring and reporting framework planned for 2017.	Monitoring, end of funding cycle evaluation.	CRPs: progress towards programmes objectives (linked to overarching expected impacts).	Evaluation of expected impact. Programme result-based management, impact pathway, theories of change, attribution.	Depending on the proposal from the external evaluator responsible for the CRP assessment. Some requisites in guidelines and terms of reference by IEA.	Programme based quantitative targets. Physical indicators by target. Sum of the quantitative contribution of each CRP to the System Level is to estimate CGIAR Outcomes.
USDA- ARS	No public guidelines. Some harmonisation in the format of NPs accomplishment reports.	Monitoring and end of funding cycle evaluation.	NPs: progress towards programmes objectives and targets (linked to USDA overarching expected achievements). ARS : An Action Plan Scorecard measures NPs outputs and outcomes, using narratives from the reports to provide evidence for impact	Evaluation of expected impact. Programme result-based management.	Mixed review panels (academics, stakeholders and government).	Achievement of programme-based quantitative targets; science quality, client satisfaction, diffusion of scientific output beyond academia.
CSIRO	Yes, guidelines for studying cases released in Nov 2015.	*Ex post, in itinere* and monitoring.	Achieved economic, environmental and social impact.	Goal-free evaluation for *ex post* assessment. Progress towards impact statement for monitoring (=pathways to achieving future impacts). Impact pathway, Attribution analysis, counterfactual.	Short case narrative, focused on quantification of the steps of the impact pathway, with quantification of impact (allows for qualitative methods).	Mostly quantitative and monetised measures (cost-benefit analysis or non-market valuation) Incentive to sum monetised impact of cases but no aggregation of environment or social impact.

Table 2. Evaluation design, methods and methodology issues (*cont.*)

Organisation	Existence of guidelines at the level of the organisation	Timing of assessment	Impact dimensions assessed	Theoretical framework (Goal-free evaluation vs expected impact, qualification of pathways and networks, attribution of impacts)	Methodologies : econometric, case studies, standardisation of cases, narrative	Measures , monetisation, indexes, counterfactual Aggregation
EMBRAPA	yes	*Ex post* (pending *ex ante*)	Economic, social, environmental, institutional (pending: political, food safety)	Goal-free evaluation, but in the process of changing toward contribution to EMBRAPA's impact axes.	Case-study of technologies. Narrative, standardised through a set template. Calculation of attribution shares.	Quantitative assessment of economic surplus at technology level. Internal rate of return at technology and EMBRAPA's level Quantitative assessment of multidimensional impacts: Ambitec method: impacts scored from -15 to +15. Aggregation of impact by type of technology: calculation of average scores of all cases by dimension.
INRA	ASIRPA : yes	*Ex post*	All innovations studied are screened for economic, social, environmental, political, and sanitary impact.	Goal-free evaluation (Not dependent on innovation objectives). Case studies	Case studies, with contribution and network analysis. Issue of standardised case report: template and three analytical tools (impact pathway characterised with qualitative and quantitative data, chronology, vector of impact.	Qualification and quantification of multidimensional impacts. Scored from 1 to 5/5. Economic impact: economic surplus calculation. Other dimensions: metric designed by expert panels based on local descriptors expressed by stakeholders.

Table 3. Implementation

Organisation	Selection of evaluators (external or internal)	Internal taskforce and links with stakeholders (participatory evaluation)	Sources of data (panels, surveys, national data, interviews) ways and means to access impact data Data collection methods (interviews, experimental design)	Number of studies undertaken Database built for impact assessment
CGIAR: CGIAR system-wide	*Ex post*: external researchers selected after a call for proposals by SPIA.	SPIA: small secretariat with 5 non-permanent academic researchers and consultants often hired.	Depending on evaluator proposals.	SPIA: 30 case studies currently investigated.
CRP programmes	CRPs: external call for tenders by IEA.	IEA supports evaluations by CRPs.	CRP's: data collected by centres in their country using their own methods	CRPs : by end 2016, 15 CRP evaluations + a few cross cutting evaluations related to Capacity Strengthening, Gender and Partnerships.
USDA- ARS	NPs: External review panel facilitated by National programme team (NPT) report.	Stakeholders in external review panel (selection criteria unknown).	Data for annual monitoring are provided by the research projects. Five-year accomplishment reports prepared by NPTs, discussion with external review panels, who in turn prepare the NP evaluation report	All NPs are evaluated at the end of each funding cycle.
CSIRO	Usually: call for tenders and self-evaluation with external validation. For early feedback on risky projects: totally internal self-assessment	Performance and Evaluation unit: 1.5 FTE dedicated to impact assessment. Quality assurance by internal senior economist. Several workshops by case with stakeholders	Depending on evaluator and stakeholders' proposals. Data sets created by the stakeholders or the research team, surveys, focus groups…	13 case studies completed and 9 under study. Database created to collect data related to monitoring of future impact
EMBRAPA	Internal evaluation	Dedicated team at the headquarters and representative in each of the 42 centres	Centres' staff collects data for *ex post* evaluation Annual updating (pending *ex ante*) National statistics data + 10 farm surveys/case	Database of 110 technologies and 220 cultivars.
INRA	Internal evaluation with external validation	Small unit (1 FTE and 4 non-permanent researchers) for capacity building on impact evaluation. Stakeholders provide some local impact descriptors	6 to 10 interviews with stakeholders provide some impact indicators, and validation of impact data.	41 case studies, codified along a hundred variables related to steps of the impact pathway, including impact. Database of > 3 000 salient facts of INRA reported yearly since 1996

Table 4. Utilisation of evaluation's results

Organisation	Use of results for demonstrating impact (accountability, advocacy, communication)	Use for organisational learning	Capacity training carried out	Use for decision-making,
CGIAR	*Ex post* impact evaluations and Impact briefs on CGIAR website. CRPs evaluation reports published on CGIAR website.	IEA cross-analysis of Phase1 evaluation of 17 CRPs (pending).	n.a.	The reviews of CRPs and their accomplishment during the 2010-15 strategic plan is made available to funders.
USDA-ARS	All NPs post their Action Plans, Annual Reports, and Five Year Accomplishment Reports, and the Executive Summaries of the reviews by External Assessment Panels. ARS post impact briefs on their web.	Through the NPs next Action Plan	n.a.	The reviews by external panels provide insight as to the future direction of the research, programme areas or focus, serving management purposes.
CSIRO	Plan to use the case studies to document the annual report of CSIRO. Use by units for self-assessment report to international panel.	Not enough case studies yet. Plan to use monitoring for identifying what are the keys to generate impacts and tracking opportunities for business and partnership development.	Some tools being designed: on line and face-to-face course for senior researchers to plan or monitor their project impact (MOOC).	
EMBRAPA	Annual report featuring internal rate of return posted on the web.	n.a.	n.a.	No (planned in 2016)
INRA	Used for the evaluation of Research thematic divisions. Used for the evaluation of institute.	Some insights provided by the cross analysis of cases, and a typology of impact generating mechanisms.	Used by 50% of INRA's scientific divisions. Training sessions planned in 2016 (in-house and external).	No

n.a.: not available.

4. Challenges of ARIA: Promoting improved practices?

Since 2000, there has been a new momentum for agricultural research and accordingly a renewed interest for Research Impact Assessment (RIA). This is visible both when surveying the academic literature and when studying RIA in practice in some of the major Public Research Organisations (PROs).

Some of the main lessons from the survey of existing methods:

- In a context where research and innovation has to address big societal challenges, it is necessary for RIA to deal with two main issues: (i) take into account broader impacts, beyond impacts on science and economic impacts; and (ii) improve knowledge on impact generating mechanisms.
- RIA methods face increasing difficulties due to key characteristics of research and innovation systems: (i) research and innovation systems are increasingly open and complex; and (ii) they are changing at a quick pace. Complexity and instability do not enable implementation of quantitative methods that allow solving the attribution problem in an appropriate way. Furthermore, in complex systems, the impact is not additive but depends on productive interactions. To face this situation, some scholars suggest shifting from attribution to the analysis of contribution of a variety of actors.
- The two sets of methods presented here ("standard economic approaches" and "approaches based on case studies") are complementary. It is crucial to develop approaches that match quantitative and qualitative analysis. This will reinforce the credibility of RIA.
- These two sets of methods still need improvement. For standard economic approaches, the two limitations are: (i) the weak ability to take into account private research and (ii) the (quasi-exclusive) focus on economic impacts. For case studies, the main limitation is related to the (still) low level of standardisation. For both sets of methods, it is crucial to develop adequate databases and metrologies to take into account non-economic impacts.

Concerning RIA in practice in the five organisations selected:

- The findings of this study confirm the gap between academic research and practices of RIA. However, it also shows that in some organisations interactions between research and practice is organised in a systematic way. In these cases social scientists are involved in the design of approaches.
- RIA is high on the agenda of all the organisations. They are all keen to build credibility of RIA for securing funding. However, evidence shows a gap between external communication on RIA and practices. This is partly related to budget constraints that restrain the resources devoted to RIA in almost all the organisations.
- Evaluation systems vary across organisations. One of the main differences is related to the scope of assessment that depends on the internal organisation. Organisations that have adopted a programme (or flagship) structure assess the programmes (and sometime the projects related to the programmes). Organisations that do not have this programme structure have designed approaches based on case studies related to technologies and aggregate these technologies according to different organisational scales (divisions, centres, the whole organisation).

The improvement of methods will be fostered by the engagement of institutional and the constitution of a community of professionals interacting globally. This is crucial for the future of RIA.

Overall, the analysis contained in this report allows identifying tensions between different styles of evaluation. The assessment systems pursue the same three objectives:

- Learning: enhance the know-how to produce an environment conducive to socio-economic impact;
- Capacity building: spread the culture of socio-economic impact to its researchers;
- Reporting to stakeholders: from accountability purposes to advocacy targeted to various audiences.

Yet none of the five systems manages to adequately meet these three objectives. The accountability objective, including for the purposes of return on the financial investment, poses particularly complex challenges, and conflicts with learning and capacity building objectives.

It may be necessary to make choices between these objectives, and adopt the corresponding type of evaluation. Power (1994) identifies two contrasted ideal-types of evaluation (type 1/type 2) characterised by a set of dichotomies. According to him, evaluation may be oriented toward external control or internal learning; unidimensional or multidimensional; evaluation process may assume low trust or high trust between evaluators and evaluated; evaluation may be performed by external experts or selected insiders; etc.

According to the characteristics of research and to the importance of learning and capacity building in research organisations, one should pay attention to foster styles of evaluation of type 2. Research organisations are learning organisations based on distributed intelligence; they are more networks than hierarchies. Hence, as suggested by Kuhlmann (2003), it is necessary to consider RIA as a central tool for strategic intelligence. RIA methods should not be implemented in a command and control logic but give a sense of ownership to the members of the organisation, in order to nurture a culture of impact.

References

Alston, J.M., C. Chan-Kang, M.C. Marra, P.G. Pardey, and T.J. Wyatt (2000), A *Meta-Analysis of Rates of Return to Agricultural R&D: Ex Pede Herculem?*, IFPRI Research Report No. 113, IFPRI. http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/125334.

Alston, J., G. Norton, and P. Pardey (1995). *Science Under Scarcity: Principles and Practice for Agricultural Research Evaluation and Priority Setting*, Ithaca, NY, US: Cornell University Press.

Andersen, M.A. (2015), "Public investment in U.S. agricultural R&D and the economic benefits", *Food Policy* Vol. 51, pp. 38–43. DOI: http://dx.doi.org/10.1016/j.foodpol.2014.12.005.

Avila, A.F., G.S. Rodrigues, G.L. Vedovoto, R. de Camargo Penteado Filho, and W. Corrêa da Fonseca Junior (2015), EMBRAPA's experience on the impact assessment of agricultural R&D: 15 years using a multidimensional approach, presented at the ImpAR Conference, INRA, Paris, France. www.alice.cnptia.embrapa.br/alice/bitstream/doc/1036444/1/EmbrapaExperienceontheimpact.pdf.

Beintema, N.M., G.-J. Stads, K.O. Fuglie, and P.-W. Heisey (2012), A*STI global assessment of agricultural R&D spending — Developing Countries Accelerate Investment*, International Food Policy Report, Agricultural Science and Technology Indicators Initiative and IFPRI, Washington D.C. www.ifpri.org/publication/asti-global-assessment-agricultural-rd-spending-0.

Bornmann, L. (2013), "What is societal impact of research and how can it be assessed? a literature survey", *Journal of the American Society for Information Science and Technology* Vol. 64, pp. 217–233. DOI: http://dx.doi.org/10.1002/asi.22803

Bozeman, B. (2003), *Public value mapping of science outcomes: theory and method. Knowledge flows and knowledge collectives: Understanding the role of science and technology policies in development,* 2,3-48. http://cord.asu.edu/home/scientific-and-technical-human-capital/public-value-mapping-of-science-outcomes-theory-and-method/.

Bozeman, B., and G. Kingsley (1997), "R&D value mapping: A new approach to case study-based evaluation", *J Technol Transfer* Vol. 22, pp. 33–41. DOI: http://dx.doi.org/10.1007/BF02509643.

Bozeman, B., and D. Sarewitz (2011), "Public Value Mapping and Science Policy Evaluation", *Minerva* Vol. 49, pp. 1–23. DOI: http://dx.doi.org/10.1007/s11024-011-9161-7.

Bozeman, B., and J. Youtie (2015), Socio-Economic Impacts and Public Value of Government-Funded Research: Lessons From Four US National Science Foundation Initiatives, presented at the ImpAR Conference, INRA, Paris, France. https://colloque.inra.fr/impar/Program-Material.

Burch, D., G. Lawrence, G.P. Green, K. Ichijo, I. Nonaka, M. Pimentel, J. Bower, C. Gilbert, V. Couto Filho, and L. Flavio (2007), *World Development Report 2008: Agriculture for development*, The World Bank, Washington D.C. http://siteresources.worldbank.org/INTWDR2008/Resources/WDR_00_book.pdf.

Callon, M. (1986), "The sociology of an actor-network", in: *Mapping the Dynamics of Science and Technology*, M. Callon, J. Law, and A. Rip, London.

CGARD (2011), *The GCARD RoadMap — Transforming Agricultural Research for Development Systems for Global Impact*, Rome: FAO, CGARD, Rome, Italy. www.fao.org/docs/eims/upload/294891/GCARD%20Road%20Map.pdf.

CSIRO (2015), Impact Evaluation Guide, CSIRO, Performance & Evaluation Unit.

Cunningham, P., A. Gök, and P. Larédo (2013), *The Impact of Direct Support to R&D and Innovation in Firms*, Compendium of Evidence on the Effectiveness of Innovation Policy, London: NESTA. www.nesta.org.uk/publications/impact-direct-support-rd-and-innovation-firms.

de Jong, S., K. Barker, D. Cox, T. Sveinsdottir, and P. van den Besselaar (2014), "Understanding societal impact through productive interactions: ICT research as a case", *Research Evaluation* Vol. 23, pp. 89–102. DOI: http://dx.doi.org/10.1093/reseval/rvu001.

Donovan, C. (2011), "State of the art of assessing research impact: introduction to a special issue", *Research Evaluation* Vol. 20, pp. 175–179. DOI: http://dx.doi.org/10.3152/095820211X13118583635918.

Donovan, C. (2008), T*he Australian Research Quality Framework: A live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research*, New Directions for Evaluation 2008, pp. 47–60. DOI: http://dx.doi.org/10.1002/ev.260.

Donovan, C. (2007), "The qualitative future of research evaluation", *Science and Public Policy* Vol. 34, pp. 585–597. DOI: http://dx.doi.org/10.3152/030234207X256538.

Donovan, C., and S. Hanney (2011), "The “Payback Framework” explained", *Research Evaluation* Vol. 20, pp. 181–183. DOI: http://dx.doi.org/10.3152/095820211X13118583635756

Douthwaite, B., S. Alvarez, S. Cook, R. Davies, P. George, J. Howell, R. Mackay, and J. Rubiano (2007), "Participatory impact pathways analysis: a practical application of program theory in research-for-development", T*he Canadian Journal of Program Evaluation* Vol. 22, p. 127. www.evaluationcanada.ca/secure/22-2-127.pdf.

Douthwaite, B., T. Kuby, E. van de Fliert, and S. Schulz (2003), "Impact pathway evaluation: an approach for achieving and attributing impact in complex systems", *Agricultural Systems* Vol. 78, pp. 243–265. DOI: http://dx.doi.org/10.1016/S0308-521X(03)00128-8.

EMBRAPA (2015), *SOCIAL REPORT - Balanço Social 2014,* EMBRAPA: Communication Secretariat, Strategic, Management Secretariat, Brasilia, Distrito Federal.

EU SCAR (2015), *Agricultural Knowledge and Innovation Systems Towards the Future – a Foresight Paper*, EU SCAR, Brussels, Belgium. https://ec.europa.eu/research/scar/pdf/akis-3_end_report.pdf.

Evenson, E. (2001), "Economic impacts of agricultural research and extension", in: In Gardner, R.E. and G. Rausser, *Handbook of Agricultural Economics*, Vol. 1A. pp. 573–628.

Feller, I. (2003), "The academic policy analyst as reporter: The who, what and how of evaluating science and technology programmes", in: *Learning from Science and Technology Policy Evaluation*, Edward Elgar, pp. 18–31.

Foray, D., D.C. Mowery and R.R. Nelson (2012), “Public R&D and social challenges: What lessons from mission R&D programs?”, *Research Policy*, Vol. 41, pp. 1697-1702.

Fuglie, K.O. (2010), "Total factor productivity in the global agricultural economy: Evidence from FAO data", in: *The Shifting Patterns of Agricultural Production and Productivity Worldwide*. Julian Alston, Bruce Babcock, Philip Pardey, Ames, Iowa, pp. 63–91.

Fuglie, K., P. Heisey, J. King, C.E. Pray, and D. Schimmelpfennig (2012), "The contribution of private industry to agricultural innovation", *Science, Policy Forum* Vol. 338, pp. 1031–1032. DOI: http://dx.doi.org/10.1126/science.1226294.

Fuglie, K.O., and A.A.Toole (2014), "The evolving institutional structure of public and private agricultural research", *American Journal of Agricultural Economics* Vol. 96, pp. 862–883. DOI: http://dx.doi.org/10.1093/ajae/aat107.

Georghiou, L. (1999), "Socio-economic effects of collaborative R&D—European experiences", *The Journal of Technology Transfer* Vol. 24, pp. 69–79. DOI: http://dx.doi.org/10.1023/A:1007724804288.

Georghiou, L., J. Rigby, H. Cameron, N. Vonortas, G. Prastacos, Y. Spanos, S. Kuhlmann, and T. Heinze (2002), *Assessing the Socio-economic Impacts of the Framework Programme*, PREST, Manchester. https://ec.europa.eu/research/evaluations/pdf/archive/other_reports_studies_and_documents/assessing_the_socio_economic_impacts_of_the_framework_programme_2002.pdf.

Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, and M. Trow (1994), *The new production of knowledge: Dynamics of Science and Research in Contemporary Societies*, SAGE Publications Ltd. ed. London.

Griliches, Z. (1958), "Research costs and social returns: hybrid corn and related innovations", *The Journal of Political Econnomy* Vol. 66, pp. 419–431. www.jstor.org/stable/1826669.

Grassini, P., K.M. Eskridge, and K.G. Cassman (2013), “Distinguishing between yield advances and yield plateaus in historical crop production trends”, *Nature communications*, Nature Publishing Group, Vol. 4, p. 2918. DOI: http://dx.doi.org/10.1038/ncomms3918.

Hanney, S., M. Buxton, C. Green, D. Coulson, and J. Raftery (2007), "An assessment of the impact of the NHS Health Technology Assessment Programme", *Health Technology Assessment* Vol. 11. DOI: http://dx.doi.org/10.3310/hta11530.

Heisey, P., J.L. King, K. Rubenstein, D.A. Bucks, and R. Welsh (2010), *Assessing the Benefits of Public Research Within an Economic Framework — The Case of USDA’s Agricultural Research Service*, Economic Research Report No. 95, USDA Economic Research Service. www.ers.usda.gov/publications/err-economic-research-report/err95.aspx.

Hurley, T.M., X. Rao, and P.G. Pardey (2014), "Re-examining the Reported Rates of Return to Food and Agricultural Research and Development", *American Journal of Agricultural Economics* Vol. 96, pp. 1492–1504. DOI: http://dx.doi.org/10.1093/ajae/aau047.

Interagency Report (2012), *Sustainable agricultural productivity growth and bridging the gap for small-family farms*, Interagency Report to the Mexican G20 Presidency, with contributions by Bioversity, CGIAR Consortium, FAO, IFAD, IFPRI, IICA, OECD, UNCTAD, Coordination team of UN High Level Task Force on the Food Security Crisis, WFP, World Bank, and WTO, 12 June. Available at: www.oecd.org/tad/agriculturalpoliciesandsupport/50544691.pdf.

Joly, P.-B., A. Gaunand, L. Colinet, P. Larédo, S. Lemarié, and M. Matt (2015), "ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization", *Research Evaluation* Vol. 24, pp. 440–453. DOI: http://dx.doi.org/10.1093/reseval/rvv015.

Kamenetzky, J.R. (2013), "Opportunities for impact: Statistical analysis of the National Science Foundation’s broader impacts criterion", *Science and Public Policy* Vol. 40, pp. 72–84. DOI: http://dx.doi.org/10.1093/scipol/scs059.

Kingsley, G., B. Bozeman, and K. Coker (1996), "Technology transfer and absorption: An “R & D value-mapping” approach to evaluation", *Research Policy* Vol. 25, pp. 967–995. DOI: http://dx.doi.org/10.1016/0048-7333(96)00890-6.

Kuby, T. (1999), Innovation as a Social Process: What Does this Mean for Impact Assessment in Agricultural Research, presented at the CIAT workshop, Costa Rica.

Kuhlmann, S. (2003), "Evaluation as a source of ’strategic intelligence’", in: *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*, Shapira, P., Kuhlmann, S, pp. 352–379.

Maredia, M.K., and D.A. Raitzer (2006), *CGIAR and NARS partner research in sub-Saharan Africa: evidence of impact to date*, Science Council Secretariat, Rome, Italy. http://impact.cgiar.org/sites/default/files/pdf/Maredia-Raitzer_October2006.pdf.

Molas-Gallart, J., A. Salter, P. Patel, A. Scott, and X. Duran (2002), Measuring third stream activities (Final Report to the Russell Group of Universities), Science and Policy Research Unit, University of Sussex.

Molas-Gallart, and P. Tang (2011), "Tracing “productive interactions” to identify social impacts: an example from the social sciences", *Research Evaluation* Vol. 20, pp. 219–226. DOI: http://dx.doi.org/10.3152/095820211X12941371876706.

Moreddu, C. (2016), "Public-Private Partnerships for Agricultural Innovation: Lessons From Recent Experiences", *OECD Food, Agriculture and Fisheries Papers*, No. 92, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jm55j9p9rmx-en.

Moreddu, C. and K.J. Poppe (2013), "Agricultural Research and Innovation Systems in Transition", *EuroChoices* Special Issue: Innovation in Agri-food, Vol. 12, Issue 1, pp. 15–20, April. http://onlinelibrary.wiley.com/doi/10.1111/1746-692X.12014/epdf.

Mowery, D.C., R.R. Nelson, and B.R. Martin (2010), "Technology policy and global warming: Why new policy models are needed (or why putting new wine in old bottles won’t work)", *Research Policy* Vol. 39, pp. 1011–1023. DOI: http://dx.doi.org/10.1016/j.respol.2010.05.008.

Mutangadura, G. and G.W. Norton (1999), "Agricultural research priority setting under multiple objectives: an example from Zimbabwe", Agricultural Economics, Vol. 20, pp. 277–286. http://dx.doi.org/10.1016/S0169-5150(99)00012-2.

OECD (2013), *Agricultural Innovation Systems: A Framework for Analysing the Role of the Government*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264200593-en.

OECD (2012), *Improving Agricultural Knowledge and Innovation Systems: OECD Conference Proceedings*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264167445-en.

Pardey, P.G., C. Chan-Kang, S. Dehmer, J.M. Beddow, T.M. Hurley, X. Rao and J.Alston (2014), "Investments in and the Economic Returns to Agricultural and Food R&D Worldwide", In *Encyclopedia of Agriculture and Food Systems*, ed. Neal K. Van Alfen, (Oxford, United Kingdom: Academic Press, 2014), pp. 78-97. http://dx.doi.org/10.1016/B978-0-444-52512-3.00045-0.

PCAST (2012), Report to the President on Agricultural Preparedness & the United States Agricultural Research Enterprise, President's Council of Advisors on Science and Technology (PCAST). www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_agriculture_20121207.pdf.

Power, M. (1994), The audit explosion, Demos.

Ray, D.K., N. Ramankutty, N.D. Mueller, P.C. West, and J.A. Foley. (2012), “Recent patterns of crop yield growth and stagnation”, *Nature communications*, Nature Publishing Group, Vol. 3, p. 1293. DOI: http://dx.doi.org/10.1038/ncomms2296.

Rip, A. (2003), "Societal challenges for R&D evaluation", in: *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*, Shapira, P., Kuhlmann, S, Chetltenham, UK, pp. 32–53.

Rodrigues, G.S., C.C. de Almeida Buschinelli, and A.F. Dias Avila (2010), "An environmental impact assessment system for agricultural research and development ii: institutional learning experience at Embrapa", J*ournal of technology management & innovation* Vol. 5, pp. 38–56.

Ruegg, R., and I. Feller (2003), A Toolkit for Evaluating Public R&D Investment: Models, Methods, and Findings from ATP’s First Decade. (Grant/Contract Report), National Institute of Standards and Technology, Gaithersburg.

Ruegg, R., and G. Jordan (2007), Overview of evaluation methods for R&D programmes: A Directory of Evaluation Methods Relevant to Technology Development Programmes, US Department of Energy, The National academies Press, Washington, DC. www.eere.energy.gov/analysis/pdfs/evaluation_methods_r_and_d.pdf.

Rymer, L. (2011), *Measuring the impact of research: the context for metric development, Group of eight Backgrounder No. 23, September*. www.snowballmetrics.com/wp-content/uploads/go8backgrounder23_measimpactresearch.pdf.

Salter, A.J., and B.R. Martin (2001), "The economic benefits of publicly funded basic research: a critical review", *Research Policy* Vol. 30, pp. 509–532. DOI: http://dx.doi.org/10.1016/S0048-7333(00)00091-3.

Scherer, F., and D. Harhoff (2000), "Technology policy for a world of skew-distributed outcomes", *Research Policy* Vol. 29, pp. 559–566. DOI: http://dx.doi.org/10.1016/S0048-7333(99)00089-X.

Shapira, P., and S. Kuhlmann (2003), *Learning from Science and Technology Policy Evaluation*, Edward Elgar. ed. Shapira, P., Kuhlmann, S, Chetltenham, UK.

Sheng, Y., E.M. Gray, J.D. Mullen, and A. Davidson (2011), *Public investment in agricultural R&D and extension: an analysis of the static and dynamic effects on Australian broadacre productivity*, ABARES Research Report No. 117.

Spaapen, J.M., and L. Van Drooge (2011), "Introducing “productive interactions” in social assessment", *Research Evaluation* Vol. 20, pp. 211–218. DOI: http://dx.doi.org/10.3152/095820211X12941371876742.

Sparger, J.A., G.W. Norton, P.W. Heisey, and J. Alwang (2013), "Is the share of agricultural maintenance research rising in the United States?", Food Policy *Vol. 38*, pp. 126–135. DOI: http://dx.doi.org/10.1016/j.foodpol.2012.11.004.

Triomphe, B., D. Barret, D. Clavel, M.H. Dabat, A. Devaux-Sparatakis, G. Faure, E. Hainzelin, S. Mathé, L. Temple, L. and A. Toillier (2015), Towards a generic, comprehensive and participatory approach for assessing the impact of agricultural research in developing countries, presented at the ImpAR conference, INRA, Paris, France. https://colloque.inra.fr/impar/Program-Material.

Walker, T., M. Maredia, T. Kelley, R.L. Rovere, D. Templeton, G. Thiele, and B. Douthwaite (2008), *Strategic Guidance for Ex Post Impact Assessment of Agricultural Research*, report prepared for the Standing Panel on Impact Assessment, CGIAR Science Council, Science Council Secretariat: Rome, Italy. www.fao.org/docrep/011/i0276e/i0276e00.htm.

Wilson, K. (2012), *One Billion Hungry: Can We Feed the World? Facts and Figure*, Agriculture for Impact, Cornstock Pub Associates, Ithaca, New York.

Wooding, S., S.R. Hanney, A. Pollitt, J. Grant, and M.J. Buxton on behalf of the Project Retrosight Team (2014), "Understanding factors associated with the translation of cardiovascular research: a multinational case study approach", *Implementation Science* Vol. 9, p. 47. DOI: http://dx.doi.org/10.1186/1748-5908-9-47.

Wright, B.D., and T.M. Shih (2011), "Agricultural innovation", *in: Accelerating Energy Innovation: Insights from Multiple Sectors*, edited by R.M. Henderson and R.G. Newell, National Bureau of Economic Research, University of Chicago Press, pp. 49–85.

Appendix

Practices of RIA in Five International Public Agricultural Research Organisations

USDA-ARS (and ERS)

In the United States in 2013, federal, state, and private institutions funded and performed roughly USD 16.3 billion worth of R&D for food and agriculture. Of this total, the majority was funded and performed by the private sector. The Federal government funded approximately USD 2.8 billion of R&D, of which the US Department of Agriculture (USDA) accounts for about 20-25%.

The main tool for evaluating all federal agencies, including research agencies, is the Government Performance and Results Act (GPRA).[14] USDA is also a partner in the STAR METRICS consortium[15] (Science and Technology for America's Reinvestment: Measuring the EffecT of Research on Innovation, Competitiveness and Science) between US federal science agencies and research institutions to document the return on investment, research impact, and social outcomes of federally-funded R&D.

Research performed at USDA is heavily oriented toward agriculture, but also includes research on natural resources, food and nutrition, economics and statistics, and rural development. The Agricultural Research Service (ARS) is the USDA's chief research in-house agency. It operates 17 National Programmes and employs more than 2 000 scientists in more than 90 laboratories throughout the country. This appendix will focus on USDA's methods for the evaluation of agricultural research programmes at ARS, as well as the Economic Research Service (ERS), which is only marginally concerned with research, but provides insights on agricultural research productivity.

ERS

The ERS's chief mission is to inform and enhance public and private decision making on economic and policy issues related to agriculture, food, the environment, and rural development. Only a few reports concern research.

ERS reports address agricultural productivity and investigate the direction and efficiency of the public and private sectors in enhancing the stock of agricultural knowledge and in developing new technologies with the objective to inform key decision makers in USDA, federal, state and local government agencies, and all groups interested in public policy issues.

ERS has published 38 studies regarding agricultural science policy since 1996,[16] concerning agricultural productivity or the performance of the public and private sectors in the US agricultural research system, including but not restricted to federal funding.

ERS has prepared a recent summary of the studies of the rates of return to public agricultural R&D in the United States (OECD, 2016). There are very few studies that attempt to incorporate economic analysis of both public and private R&D impacts in the same analytical framework, and those that do are still limited in their analysis. Data do not allow distinguishing easily neither the private and public research economic impacts, nor the federal and state research. Furthermore agricultural research impact is not always separated from the influence of other factors. Therefore the calculation of the overall rate of return of research is mostly convincing for broad issues.

14. www.whitehouse.gov/omb/performance/gprm-act.

15. www.starmetrics.nih.gov/.

16. www.ers.usda.gov/publications.aspx?sortExpression=date&sortDirection=DESC&topicId=1793, consulted on 15 March 2016.

Besides, ERS is not charged with the evaluation responsibility of ARS research programmes. In 2010 ERS published a report on methodological issues in the economic impact assessment of agricultural R&D, with illustrations taken from three ARS programmes (Heisey et al., 2010).[17] ERS research also addresses the complementary roles of public and private R&D (see for example Fuglie and Toole, 2014).

ARS

The ARS annual budget is established on the basis of a five-year strategic plan (presently strategic plan FY 2012–2017 revised in 2014). ARS research is organised into 17 National Programmes (NPs) which serve to bring coordination, communication, and empowerment to approximately 750 research projects carried out by ARS.

ARS NPs are divided into four major broad categories:

- Nutrition, Food Safety, and Quality
- Animal Production and Protection
- Natural Resources and Sustainable Agricultural Systems
- Crop Production and Protection

The primary tool of research evaluation for ARS is the five-year programme planning and evaluation cycle. There is neither a specific budget line nor a unit dedicated to impact evaluation. Fourteen performance measures describe specific measurable achievements, which indicate progress toward reaching USDA strategic goals and priorities. Baseline and performance targets to be reached by 2017 are established for NPs and projects within the programmes, to align them with performance measures and ARS vision for agricultural research. Most targets established for programmes are on providing research outputs, but a few targets concern research outcomes or impacts. Funds for the monitoring and evaluation of targets are planned within the programme budget.

Programme performance against targets is monitored annually. Each NP Team (NPT) prepares an annual report featuring the NP's major accomplishments. At the end of each NP's five-year cycle, an accomplishment report is prepared by the NPT, and a retrospective Review by an external panel is convened. The NPT puts together an accomplishment report, selecting research outputs to illustrate accomplishments in the impact areas identified at the outset of the five-year period. There is no unified procedure for the terms of reference or content of retrospective reviews, as they are dependent on the national programme objectives. An outside group of experts (made up of academics, stakeholders, and government) give their feedback on the programme. Criteria used by the review panels include achievement of goals, client satisfaction, and impact of scientific outputs. While experts can observe outputs or scientific accomplishments, and discuss research relevance, they sometimes state that they lack data to observe the impacts of recent funding on society.[18]

An Action Plan Scorecard measures NPs outputs and outcomes, using narratives from the reports to provide evidence for impact.

There are, occasionally, simple economic evaluations driven by individual grants and interests, at the level of an entire research structure. For example the Dairy forage centre (one centre of ARS) made a simple calculation exercise of financial impacts of some of its research programmes, seemingly with the assistance of an outside consulting firm (USDA-ARS, 2015a).

The results of ARS evaluations are used for communication, accountability and management purposes.

ARS impact Report (USDA-ARS, 2015b) presents narratives of recent achievements of ARS research in crop and animal production, disease and pest protection, bioenergy, natural resources, food safety, and human nutrition. All NPs post their Action Plans, Annual Reports, Five Year Accomplishment Reports, and the Executive Summaries of the reviews by External Assessment Panels.[19]

17. This report applies qualitative economic reasoning to the evaluation of three case studies of ARS research offerings (Bovine Quantitative Genetics and Genomics, Water Quality and Watersheds, Nutrient Data Laboratory).

18. See for example National Program 304 Crop Protection and Quarantine External Review Assessment, December 2013, National Program 216 or: Agricultural System Competitiveness and Sustainability Executive Summary, 22 December 2011

19. www.ars.usda.gov/research/programs.htm.

The purpose of external reviews is two-fold: they ensure that the research is being conducted as indicated in the Action Plan (summative); and they provide insight as to the future direction of the research, programme areas or focus, serving management purposes (formative).

Key references

In addition to an interview with an ERS researcher, some key bibliographic documents were studied:

Fuglie, K.O., and A.A. Toole (2014), "The evolving institutional structure of public and private agricultural research", *American Journal of Agricultural Economics* Vol. 96, pp. 862–883. DOI: http://dx.doi.org/10.1093/ajae/aat107.

Heisey, P., J.L. King, K. Rubenstein, D.A. Bucks, and R. Welsh (2010), *Assessing the Benefits of Public Research Within an Economic Framework. The Case of USDA's Agricultural Research Service*, Economic Research Report No. 95, USDA Economic Research Service. May. www.ers.usda.gov/webdocs/publications/err95/7547_err95_1_.pdf.

OECD (2016), *Innovation, Agricultural Productivity and Sustainability in the United States*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264264120-en.

USDA-ARS (2015a), *Improving the Economic and Environmental Sustainability of Dairy Forage Farm Systems. 8 examples showing the $1.5 billion impact of research at the U.S. Dairy Forage Research Center.* www.ars.usda.gov/ARSUserFiles/50901500/pdf's/impact%20statement%20(8)%20v1.pdf.

USDA-ARS (2015b), *Impacts—Selected accomplishments under research, education, and economics mission area goals*, USDA-ARS Office of Communication. www.ars.usda.gov/is/np/ARSImpacts/ARSImpactsIntro.htm.

USDA-ARS (2012), *ARS Strategic Plan for FY 2012-2017*, The Service, Washington, D.C. www.ars.usda.gov/ARSUserFiles/00000000/StrategicPlan/USDAARSFY2012-2017StrategicPlan.pdf.

Link to ARS 17 national programmes (NPs): www.ars.usda.gov/research/programmes.htm

- NP 211, Water Availability and Watershed Management (action plan, accomplishment report, assessment executive summary)
- NP 216, Agricultural System Competitiveness and Sustainability (action plan, accomplishment report, assessment executive summary)
- NP 101, Food Animal Production (action plan, accomplishment report, assessment executive summary)
- NP 107, Human Nutrition (action plan, accomplishment report, assessment executive summary).

INRA

The French National Institute for Agricultural Research (INRA) employs 8 300 people (of which 1 800 researchers and 2 500 engineers), for an annual budget of EUR 880 million (in 2014). INRA is organised into 13 scientific divisions. INRA's missions have been set by the laws on research in 1982 and 2006, recently revised in 2015, and its scientific priorities are set every ten years in a strategic guidelines document (presently 2010-20).

History of impact assessment

Before 2009, while science quality was monitored and evaluated, there was no system in place to evaluate societal impact in INRA. Some reports were commissioned, for example in 1995 two external research teams from École des Mines Paris reported on 12 *ex-post* case studies of innovation successes and failures of INRA.[20] Since 1996 the communication unit has maintained a database of salient facts, the "Zoom" database, to illustrate the annual report and other communication needs. This database gathers narratives and data regarding more than a thousand innovations.

Since 2009, INRA, like other French Public Research Organisations, is assessed every five years by an external agency called HCERES (formerly the AERES). The first external assessment report in 2009 recommended INRA to go beyond the evaluation of the scientific quality, and address the socio-economic impact of its research. As there was no universal standard for doing so, it was suggested that a comprehensive system for impact assessment should be developed. The AERES experts argued that it would lead to a greater legitimacy and increase INRA's role in mediating science-society debate on growing scientific issues.

Following these recommendations, a team of seven INRA researchers in economy and sociology was appointed to design a methodology that could be implemented to perform an Analysis of the Socio-economic Impacts of the Public Agricultural Research, and be tested on INRA: the ASIRPA project was launched in 2011. After a pilot phase in 2011-13 (Colinet et al 2014), several scientific divisions of INRA tested the approach in real assessment conditions in 2013-15, and a resulting impact assessment system was formally institutionalised in INRA in 2015. In parallel, in 2013, INRA's economists were asked to calculate the Internal Rate of Return of the French agricultural research (Butault et al., 2015a, 2015b). They used standard methods described in Section 2 of this report.

Purpose of impact evaluation and evaluation design

The ASIRPA approach is based on the use of standardised *ex post* case studies (Colinet et al., 2013; Joly et al., 2015). It chiefly serves three purposes: to facilitate a culture of impact that will improve research management practices; to stimulate the understanding of the context, processes and the mechanisms through which impact is generated (i.e. the chains of operations which 'translate' research knowledge into a format which can be put into practice by users beyond the academic sphere); to demonstrate the value of the Institute's research to stakeholders and the public at large, taking into consideration that these stakeholders often have different, even contradictory, values and priorities.

Based on the information gathered in the Zoom database of salient facts, research outputs, beneficiaries, and impacts, the information on a thousand INRA innovations (gathered from 1995 to 2012) has been codified. The analysis produced a limited number of impact patterns (Gaunand et al., 2015), which facilitates the selection of a representative sample of cases. The selection of cases is completed by interviews with heads of scientific divisions.

ASIRPA is an *ex post* and backwards approach, which means that the analysis starts with identifying the impact observed and working back the impact pathway to qualify the contribution made by research and that of other partners. The impact for all cases is investigated, characterised and quantified along five dimensions: economic, political, social-territorial, environmental, and health. The ASIRPA standard for case studies comprises a report and three analytical and visual tools: a chronology which defines the beginning and end of each case, along with the principal events taking place between these dates; an impact pathway, which drafts the different phases and actors involved in the impact generation chain. It enables identification of the specific contribution of INRA in the innovation network, analysing the role of contextual factors and identifying the critical mechanisms which underpin impact-generating actions; an impact vector summarised in a table and illustrated with a radar chart rates the intensity for each of the five dimensions of impact.

20. *Les chercheurs et l'innovation: regards sur les pratiques de l'INRA,* 1998, Paris : Quae.

In terms of metrics, impact is assessed using local descriptors collected in interviews with stakeholders and beneficiaries, for each "professional adhocracy field" related to a case study, and for each dimension of impact (economic, environmental, political, human health, social). To compare impact across cases, and above all to give an overall picture of the diversity of INRA's impact, the impact intensity is quantified on an ordinal scale from 1 (weak impacts) to 5 (strong impacts). A scoring metric template is established for each impact dimension based on an inter-case comparison. The translation of quantified values into rankings from 1 to 5 is based on a scoring range built by ASIRPA using expert panels. The work with expert panels, required to rate impacts of all standardised cases in a comparable manner, has been completed for economic, political, and environmental impact; a similar approach is planned for health, social and territorial impact. Stakeholders' opinions are taken into consideration to validate the impact pathway and characterise the impacts. In this regard, the approach developed by the French Agricultural Research Centre for International Development (CIRAD),[21] which is comparable to ASIRPA, uses a more participatory method by organising workshops where stakeholders also have a say in the terms of reference for the assessment.

Regarding the aggregation of cases, the process rests on the design of a database, built by encoding the qualitative information collected in the standardised case studies. The variables are related to each step of the impact pathway, and were designed during the project based on the first pilot cases studied.

Implementation

The whole impact assessment system of INRA is co-ordinated and methodologically supported by the ASIRPA team, comprising of one full-time equivalent (FTE) engineer at the Delegation for Evaluation, with the support of five researchers.

The ASIRPA project started with the study of 14 cases designed to build and standardise the methodological tools. The ASIRPA research team selected the cases for their representativeness of INRA and the methodological issues they raised. The team conducted the interviews and drafted the reports.

Once the standard was stabilised, INRA top-level management prompted the scientific divisions to study cases and document their impact in the self-assessment report that is submitted every five years for review by an international panel. Seven scientific divisions requested the support of the ASIRPA team and produced 27 cases. Cases were selected by the head of division's team, who further co-ordinated the instruction and drafting of case studies by the principal investigators of each case, in close relation with the ASIRPA team, who was responsible for the enforcement of the standard methodology. Forty-one case studies have been produced so far.

The principal investigator of each division team identifies the relevant stakeholders and conducts semi-directed interviews to collect data related to each step of the impact pathway. Impact data must originate from sources that are external to INRA. Desk research, database analysis (contracts, patents, etc.) also complete the information.

To scale-up from the sample of single cases, the first 33 case studies were encoded along a hundred of variables identified for performing a cross-cutting analysis. Some descriptive statistics were also performed to build a typology of four categories of pathways, with corresponding impact pathways, and the main mechanisms and critical factors which underpin impact-generating actions.

The next objective for INRA is to increase the number of cases produced in order to deepen the representativeness of cases, increase the robustness of the typology and better characterise the societal impact of INRA's scientific divisions. Given the limited resources of the ASIRPA team, generalising the implementation of the ASIRPA methodology requires a system to build internal capacity and increase the delegation of support delivery to principal investigators and head of divisions' teams. Each research division will designate an "impact champion" to assist the researchers who want to develop a case study. An important objective of ASIPRA's next phase will also be to investigate ways to extract rich information and knowledge from the typology and *ex post* studies, in order to improve INRA's methods of producing impact. Another challenge is the necessity to update the cases, with new data on their impacts, changes in the networks involved and the external context.

Utilisation of evaluation results

The results yielded by the implementation of the evaluation approach enabled capacity building, knowledge and accounting, and may assist managers in decision-making processes.

21. See impact evaluation methodology toolbox, ImpresS, developed by CIRAD at: http://impress-impact-recherche.cirad.fr/impress/a-five-step-method.

The ASIRPA case study reports are posted on the ASIRPA project website[22] and are used by INRA scientific divisions in the self-assessment reports distributed to the international review panels appointed to assess their performance. Most panels reported they appreciated the information received. At the level of INRA, case studies and results from the cross-analysis and typology have fed the section dedicated to impact in the 2015 self-assessment report provided to the HCERES.

In terms of learning and capacity building, ASIRPA's approach, results, and the findings on the Internal Rate of Return of French agricultural research have been communicated both internally and externally. Externally, INRA organised several conferences dedicated to agricultural research assessment: international conferences in 2012, 2015, 2016[23] that contributed to building a research community on these topics, and two national conferences (2012, 2015) more specifically intended for INRA's researchers, French stakeholders and partner organisations. INRA's internal culture on socio-economic impact evaluation spread, since 7 of the 13 scientific divisions have implemented the ASIRPA approach with about 120 INRA staff directly involved in analysing case studies. A collective training intended for INRA volunteering impact champions, along with representatives of partner organisations took place in the second half of 2016 to nurture the impact culture.

The typology of impact pathways and results from the 33 case studies led to robust results on INRA's societal impact and the identification of the general characteristics of impact pathways, their mechanisms, and the roles of INRA within the innovation network of actors (Matt et al., 2016). Thanks to the typology, some characteristics of INRA's general impact can be described in terms of size and diversity. This typology also enabled identification of some conditions required for making a diversity of impact pathways productive.

This project needs further developments, but the knowledge generated can assist managers and researchers in taking effective measures (structures, partnerships etc.) that have been proven to increase the odds of generating positive societal impact.

By implementing ASIRPA, INRA has built new links between high-level managers and researchers to improve research management. INRA's division for Partnership Transfer and Innovation, recently renewed, has gained some insights from the project results on impact-generating mechanisms.

Key references

Butault, J.-P., S. Lemarié, A. Musolesi, F. Huard, M. Simoni, and B.Schmitt (2015a), *L'impact de la recherche agronomique sur la productivité agricole française. Une approche par le taux de rentabilité interne (TRI) des dépenses publiques affectées à la recherche agronomique en France*, INRA Sciences Sociales, No. 1/2015 - Septembre 2015. INRA Sciences Sociales 1. http://ageconsearch.umn.edu/bitstream/210178/2/iss15-1.pdf.

Butault, J.-P., S. Lemarié, A. Musolesi, and B.Schmitt (2015b), "The impact of agronomic research on French agricultural productivity: an estimation of the internal rate of return", Presented at the ImpAR Conference, INRA, Paris.

Colinet L., P-B. Joly, A. Gaunand, M. Matt, P. Larédo and S. Lemarié (2014), *ASIRPA – Analyse des Impacts de la Recherche Publique Agronomique*, Rapport final. Rapport préparé pour l'Inra. Paris, France. https://inra-dam-front-resources-cdn.brainsonic.com/ressources/afile/259143-5d854-resource-le-rapport-asirpa.html.

Colinet, L., A. Gaunand, P.-B. Joly, S. Lemarié, M. Matt, P. Larédo, and A. Hocdé (2013), "Une approche multidimensionnelle de la mesure des effets de la recherche publique agronomique : le cas de l'INRA", in: *Penser La Valeur D'usage Des Sciences*, Olivier Glassey, Jean-Philippe Leresche et Olivier Moeschler, Université de Lausanne.

Gaunand, A., A. Hocdé, S. Lemarié, M. Matt, and E. De Turckheim (2015), "How does public agricultural research impact society? A characterization of various patterns", *Research Policy* Vol. 44, pp. 849–861. DOI: http://dx.doi.org/10.1016/j.respol.2015.01.009.

Joly, P.-B., A. Gaunand, L. Colinet, P. Larédo, S. Lemarié, and M. Matt (2015), "ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization", *Research Evaluation* Vol. 24, pp. 440–453. DOI: http://dx.doi.org/10.1093/reseval/rvv015.

Matt, M., A. Gaunand, A., P.-B. Joly, and L. Colinet. (2016), "Opening the black box of impact – Ideal-type impact pathways in a public agricultural research organization", *Research Policy,* in press, available on-line. DOI: http://dx.doi.org/10.1016/j.respol.2016.09.016.

22. www6.inra.fr/asirpa_eng/Method-and-Cases/Case-studies.

23. https://colloque.inra.fr/impar.

CSIRO

In Australia, the Commonwealth Scientific and Industrial Research Organisation (CSIRO) was established in 1916. CSIRO employs over 5 000 people, and is funded by the Department of Industry, Innovation and Science with a 2014-15 budget of AUD 1.2 billion (CSIRO Operational Plan 2014–15, "Achieving positive impact together") (CSIRO, 2014). CSIRO is organised into three Lines of Business (LoB):[24] 1) Impact Science; 2) National Facilities and Collections; and 3) CSIRO Services. Within its Impact Science LoB, there are nine Business Units (BU, previously known as Flagships), that focus on the biggest challenges facing the nation across key sectors: Agriculture, Food and Nutrition, Health and Biosecurity, Data 61, Land and Water, Mineral Resources, Manufacturing, Energy, and Oceans and Atmosphere.

History of impact assessment

Since the 1990s and until 2011, CSIRO was funded by the Federal Government through a quadrennial funding cycle. To prepare for each new funding phase, and in order to assess the performance achieved during the previous period, reviews were conducted by external consultants (the last review of this type was issued in 2014 by ACIL Allen Consulting, 2014). These reviews drew on representative case studies to calculate the overall organisational economic value of CSIRO during the period, calculating a total Internal Rate of Return. The quadrennial funding cycle was replaced with a four-year rolling funding agreement process, which requires CSIRO to provide ongoing insight into business units' performance on an annual basis by reporting on key performance indicators linked to their strategic and corporate plans.

Purpose of impact evaluation and evaluation design

Two key drivers recently led to the renewal of the impact reviewing procedure. First, findings from a Deloitte assessment report called for a common framework and procedure for evaluating the different flagship programmes of CSIRO. Second, CSIRO was increasingly interested in linking *ex post* assessment to monitoring of impact. In 2010, this combination of factors led to the development of a consistent, organisation-wide approach to impact assessment and management. This decision launched the Impact 2020 project, which investigated and designed an impact based system that replaced the temporary external assessment process. The 2013 Public Governance, Performance and Accountability Act[25] further supports that shift, by imposing that government agencies use more mixed methods (than internal rate of return only) to assess their broader impact, consistent with the diversity of CSIRO's missions, and that they focus on impacts instead of outputs. In addition, the new CSIRO Strategy 2020,[26] released in 2015, continues to maintain CSIRO's core mission which is to deliver triple-bottom-line impact to the nation through its research.[27]

The main goals of the Impact 2020 project,[28] as well as the Strategy 2020, were to underline CSIRO's commitment to achieving societal impacts in compliance with the new act, to account for defensible and robust evidence of impact, to develop guidelines for assessing impacts across the different lines of business, to internally increase the culture of impact, and to evaluate impact outcomes. The key audience of these documents was the government, industry, other R&D organisations, and universities. Through its impact assessment approach, CSIRO aims to improve: advocacy (increased capacity to articulate future and delivered impact); accountability (the ability to provide defensible, robust evidence of impact); analysis (the opportunity to better understand and maximise research impact through continuous improvement; and funding allocation (better informed decision making). The Impact 2020 initiative asserts that impact assessment not only informs investment decisions, but also inspires improved research management practices, and particularly collaborating practices. This organisational learning is particularly targeted through monitoring of impact. The Impact 2020 initiative also aims at changing the attitude of scientists to partners.

An important output of the Impact 2020 project is the development of a common framework for all impact assessments within CSIRO, to be assessed using a case study approach. This framework is described in the Impact Evaluation Guide (CSIRO, 2015) and is to be implemented by whoever is performing *ex post, ex ante* or *in itinere* impact assessments of CSIRO. These guidelines notably develop a programme logic based on input-outcome-impact model. It is suggested to broaden the dimensions of impacts considered, including

24. www.csiro.au/en/About/Strategy-structure/Operating-model.

25. www.comlaw.gov.au/Details/C2013A00123.

26. www.csiro.au/strategy/.

27. **Triple bottom line** (or otherwise noted as **TBL** or **3BL**) is an accounting framework with three parts: social, environmental (or ecological) and financial

28. www.csiro.au/en/About/Our-impact/Our-impact-model.

economy, environment and society, and to use a mix of methods for investigating the impact, including qualitative, cost-benefit analysis and option values (to account for the externalities). Great efforts have been provided in an attempt to define each of the three impact categories. Measurement is to be done on a case-by-case basis, using quantitative or qualitative methods. However, few tools are provided to assess non-economic impact without monetisation.

Regarding monitoring, within CSIRO the future intended impact is investigated at the BU programme level. These future impacts are described in terms of impact pathways, captured in "*Impact Statements" (*pathways to achieve future intended impacts),[29] *which are aggregated to "Impact Areas" and linked to each BU's goal.*[30]

In terms of method, given the diversity of sectors tackled by CSIRO, it is a challenge to identify relevant indicators of impacts for each of the sectors. Nevertheless, CSIRO has clearly defined the subcategories of environmental, social and economic impacts their research tends to achieve (see Table 2 of the *Impact Evaluation Guide*[31]).

Implementation

The Performance and Evaluation team of CSIRO, allocates 1.5 FTE in order to implement the impact assessment approach. The team identifies potential case studies, leads the case selection process, assists researchers in drafting reports (notably for analysing economic impact), ensures the consistency of the cases by promoting the implementation of the guidelines, and mediates the case studies' participants (consultants, stakeholders, etc.). The team also assists in the development and maintenance of the impact statements, and provides capability building courses. Three different ways to study *ex post* impact are used (whatever the model, a case requires around 8 weeks to be completed):

- External consultants are hired for 90% of the cases and draft reports following CSIRO's guidelines. This model costs AUD 20 000 to AUD 45 000 per case.
- CSIRO's Performance and Innovation team performs the economic analysis and drafts the report which is validated by an external third party. This model costs AUD 5 000 to AUD 8 000 per case.
- CSIRO's Performance and Innovation team carries out the study totally internally. This model is only used to provide very early feedback on risky research, since independent evaluation process is a core value of the approach.

Two internal calendars incite CSIRO to regularly perform case studies: its annual report and the external assessment of its BUs. CSIRO Impact Science BUs are evaluated every four years by an international evaluation panel to evaluate their impacts. Previously, BUs presented case studies but these were inconsistent, assessing either intended or delivered impacts, and using different methodologies. Guidelines are now implemented to study at least one case, each year across each BU, requiring support from the BUs as well as the CSIRO's reporting effort. Still, the total number of cases studied per year is dictated by budget and internal resources.

In terms of results yielded by the implementation of the evaluation approach, between 2011 and 2015, thirteen case studies have been undertaken in the different research areas of CSIRO. These case studies were used to pilot test the Impact Evaluation Guide and refine the assessment process. An additional nine cases are being conducted in the 2015-16 period. All twenty-two cases followed the guidelines, which were publically released in November 2015.

Cases are selected according to a specific rationale: for each case, a strong counterfactual and a high share of attribution back to CSIRO must be ensured. Also, maintaining relationships with the stakeholders involved eases the access to evidence and impact data. A cross section of cases should be ensured between long investments of CSIRO in traditional research areas involving fundamental science and more recent innovative research but no balance of case number is sought between programmes or BUs. The cases must cover the three lines of business of CSIRO (Science, National Facilities and Collections and Services). A CSIRO case study can be variable in size (either one project, or one field of research comprising over 30 projects, or one patent or one license). BUs are actively involved in the case selection process through proposing a shortlist of three case studies each to the Performance and Evaluation Team. Each case is

29. See the Energy BUs Impact Statements: www.csiro.au/en/Research/EF/Areas/Our-impact-strategy.

30. www.csiro.au/en/About/Our-impact/Planning-and-monitoring-our-impact.

31 . www.csiro.au/en/About/Our-impact/Evaluating-our-impact.

described in a template, including the timespan, the stakeholders involved, the counterfactual and the contribution.

In order to collect data for *ex post* assessment, two to three workshops are organised with stakeholders (partner universities, industries or spin-offs) for each case study. The first workshop provides an inventory of the data available, decides on the data to be collected and the measurement method to be used, raises analytical questions possibly amending the terms of reference of the study, and makes hypotheses regarding the potential benefits. The next workshops are mainly dedicated to co-creating the data sets, and discussing and deciding on impact attribution to be assigned to CSIRO. The Performance and Evaluation Team ends up processing these data to characterise the economic, environmental and social impact of each case, using either cost-benefit analysis, or non-market valuation, or non-monetary quantification or narrative.

The data collection for monitoring relies on the involvement of programme managers. CSIRO's BU programmes have defined between one and five impact statements each (for instance, one impact statement of the Energy BU is "National transition pathways to decentralised, low carbon electricity systems").[32] Programmes are responsible for developing and managing the impact pathways (or statements) for the portfolio of projects they manage. Monitoring data related to Impact Statements (i.e. future impacts) are stored in a central database capable of producing simple reports and graphics to assist assessors. This monitoring system, already in place, enables identification of opportunities for business development and partnership development, and of the keys to generating impacts. This system also shows what data are already available in order to decide on studying a case, and provides an archive of research activities at CSIRO.

In terms of capacity building, CSIRO's Performance and Evaluation team is designing two courses (partly online and partly face to face) aimed at enabling all staff to plan and monitor ongoing impact pathways of their projects or assess their impacts *ex post*.

Ex post case studies' aggregation relies on calculating a total benefit-cost ratio indicator derived from the cost-benefit analysis. For cases for which no cost-benefit analysis has been performed, it is suggested to provide the full range of relevant and measurable – monetary and non-monetary – costs and benefits of the work programme.

Regarding the monitoring database in place, it ensures consistency of tracking of impacts and enables identification of the key targets when it comes to impact at the level of the whole enterprise. The information in that database allows for the creation of CSIRO's Impact Map,[33] a communication tool which describes the major impacts being progressed and delivered by the organisation.

The shift toward assessing societal impact, described in Strategy 2020, is still recent at CSIRO and several implementation issues remain to be addressed, particularly regarding the selection of cases, the way their investigation is conducted and the impact assessed, and the aggregation at the level of CSIRO.

Regarding the identification of case studies, so far most of the cases proposed for impact assessment by BUs are related to commercially-based outputs from industry-oriented research, and few cases are related to policies, land use, or climate change (i.e. public good or environmentally based research). That bias prevents methodological improvements on the related impact dimensions. Another cultural challenge that may influence number and types of cases proposed for investigation is related to the understanding of impact pathways. The lack of impact planning and monitoring is affecting CSIRO's ability to conduct assessments but this is a key area under focus for capacity development and support. Besides, the methodological choices related to the attribution issue leads to case selection bias. Indeed, accounting for the pre-existing stock of knowledge while trying to attribute a share of the impact being assessed is a tricky task. Thus, CSIRO tends to select recent cases or cases with specific achievements (as opposed to climate change or land change impacts), where attribution share is easier.

Implementation issues are also a concern in the case investigation steps, since keeping the study within eight weeks is difficult while accounting for other commitments of the stakeholders and researchers involved and issues with data collection. Besides, dealing with the competing objectives of producing credible reports resulting from an external validation, while developing an internal culture of impact, remains a "golden egg" perspective for CSIRO.

32. See example impact statements for the Energy BU: www.csiro.au/en/Research/EF/Areas/Our-impact-strategy.

33. Figure 3 at www.csiro.au/en/About/Our-impact/Our-impact-model/Ensuring-we-deliver-impact.

In terms of impact characterisation, access to data may be hindered by lack of data history for long-term cases or by confidentiality matters when competitive advantage or commercial confidences are at stake. Regarding quantification, CSIRO is struggling finding skilful external consultants, with experience in impact assessment accounting for triple bottom line benefits, which extends beyond performing standard high-quality cost-benefit analysis.

A large number of case studies are required in order to deliver a relevant impact message on the different sectors where CSIRO is involved. Increasing the number of cases studied annually is thus an important pre-condition. Budget and staff capacity being limited, this increase relies on a greater involvement and capacity building of researchers. A cultural shift is required in order to arouse more proposals for cases from researchers, as well as a significant increase in funding to the BUs and the Performance and Evaluation team to resource the aspired increase in the number of case studies.

Despite being mentioned in the guidelines released in November 2015, *ex ante* impact assessment is not performed yet and its implementation at CSIRO remains to be planned.

Utilisation of evaluation results

Impact case studies are used for communication on impacts and reporting. They intake the form of 12-15 page summaries which feed into other performance reports such as the CSIRO's annual report[34] and the BUs self-assessment reports, and published on CSIRO's website.[35] Other infographics or communication tools are developed based on these reports and used by each business unit for its own communication.

Key references

In addition to two interviews conducted with senior managers of CSIRO, some key bibliographic documents were studied:

Acil Allen Consulting (2014), *CSIRO's Impact and Value. An Independent Evaluation*, Melbourne. www.acilallen.com.au/cms_files/ACILAllen_CSIROAssessment_2014.pdf.

CSIRO (2015), Impact Evaluation Guide, CSIRO, Performance & Evaluation Unit. www.csiro.au/impact.

CSIRO (2014), Operational Plan 2014-2015: Achieving positive impact together. www.csiro.au/impact.

EMBRAPA

The Brazilian Agricultural Research organization, EMBRAPA (*Empresa Brasileira de Pesquisa Agropecuária*), employs 9 800 people (of which 2 400 researchers), for an annual budget of BRL 2.6 billion (USD 670 million). EMBRAPA is organised into 42 product-based, basic theme or ecoregional research centres. Each centre comprises research or service units. The EMBRAPA strategic plan sets large missions and goals over a period of 20 years.

History of impact assessment

The first assessment of the impact of EMBRAPA's technologies started in the middle of the 1980s, as part of a national effort for impact assessment (Avila et al., 2015). Studies calculating economic surplus or econometric analyses were performed *ex post*, either at the level of commodities, grants, programmes, regions or the EMBRAPA as a whole. During the 1990s the majority of the econometric studies were related to smaller objects (research centre or commodity-oriented research) and based on local initiatives, training requirements (Ph.D. or M.Sc. thesis), or at the demand of large international funders (Inter-American Development Bank and World Bank).

Purpose of impact evaluation and evaluation design

The impact evaluation process was renewed in 1997, with the annual issue of an "EMBRAPA social balance" report (adapted from a model suggested by the Brazilian Institute for Social and Economic Analysis, Ibase). This process chiefly serves accountability purposes, demonstrating the impact of donors' investments, thus easing the negotiation of the annual budget with the Government, Congress, and access to international loans. Internal learning regarding the impact generation mechanisms is also sought, with regular feedback to

34. www.csiro.au/en/About/Our-impact/Reporting-our-impact.

35. www.csiro.au/en/About/Our-impact/Our-impact-in-action.

the researchers. The main effort concerns *ex post* impact evaluation. Econometrics was developed in the 2000s in collaboration with international researchers. These studies used a variety of models and approaches (including TFP) and demonstrated the decisive roles played by investment in agricultural research in Brazil, notably those of EMBRAPA. A large initiative was launched to monitor and assess the dissemination and impact over time (and until they disappear from the market) of 110 technologies and 220 cultivars. In 2000, the Impact Assessment process of EMBRAPA shifted from a one-dimensional approach (economic), to a multidimensional approach, accounting for economic, social, and environmental. Other societal impact dimensions have later been accounted for: institutional impact, intangible impact, and impact on consumption. More recently, the *Brazilian* Office of the Comptroller General (CGU), which commonly serves as an external adviser, checking internal finances of EMBRAPA, and ensuring no bribery is at stake, requested EMBRAPA to account for its results and how it intends to achieve the goals set in its strategic plan. This led EMBRAPA to launch an initiative to assess *ex ante* impacts.

Each impact dimension assessed relies on specific *ex post* metrics. Economic impact is calculated as an economic surplus, and takes into consideration the adoption rate of the technology and the share of impact attributed to EMBRAPA’s contribution. The attribution share to EMBRAPA cannot exceed 70%. Since FTEs are also recorded, ROI is calculated for each technology. It appears that economic impacts are highly skewed among technologies. For the assessment of the non-economic impact of each technology, an *ad hoc* method has been designed, called Ambitec. The Ambitec method was first employed by EMBRAPA’s researchers in ecology, geography, and later joined by economists. Ambitec is a "multi-attribute indicators system", comprising 24 criteria and 125 sustainability indicators, which enable environmental and social impacts of technologies (impacts on consumption/food safety are being integrated into Ambitec) to be identified. The indicators have been selected from prior experience and field trials (Rodrigues et al., 2010). For each technology, impact data are collected by centres through surveys and interviews with farmers/administrators to obtain change coefficients related to a given technology or rural activity effects observed. Ten farmers are surveyed for each technology, and average change coefficients are calculated. Ambitec assigns relevant coefficients to the different impact criteria, according to its relevance toward effecting socio-environmental impacts and its scale of occurrence. This results in an aggregated index of socio-environmental impact ranging from -15 to +15. Ambitec accounts for specific evaluation contexts since it allows for emphasizing relevant local aspects or evaluation objectives, or excluding non-applicable indicators. Institutional impact of EMBRAPA is defined as the impact of research on external actors and on EMBRAPA’s organisation itself. It encompasses issues related to knowledge advancement, capacity building and use in public policies. This impact dimension is assessed through an internal survey. The intangible impacts of technologies are related to knowledge, training and other political and institutional impacts. They are assessed at EMBRAPA since 2006 using the ESAC methodology that has been designed by Geopi/Unicamp (Brazil).

Regarding the characterisation of *ex ante* impact assessment, the EMBRAPA’s 2014-34 strategic plan (EMBRAPA, 2015) set five impact targets: sustainability, insertion in bioeconomy, contribution to public policies, poverty reduction and positioning at the frontier of knowledge. The *ex ante* impact assessment criteria of projects are derived from the *ex post* Ambitec method (type of impact, technologies expected and target aimed). *Ex ante* impact of research projects is to be assessed on a scale running from very negative to very positive impact, and along criteria related to the dimensions of impact affected (economic, social, environmental, institutional, food safety) and the impact targets.

EMBRAPA’s methodological choices still encompass challenges. In terms of method, the rules to decide on attribution shares among EMBRAPA and stakeholders remain unclear. Only public research organisations (no private research facilities) are considered for a potential contribution to EMBRAPA’s impact. Another important objective for EMBRAPA’s current work is its policy impact. Policies that EMBRAPA’s researchers have investigated have been reported (the 2014 survey reported 60 policies). Two different approaches (quantitative using aggregated data, or qualitative inspired by expert panel work by ASIRPA) are being considered for assessing and monitoring this impact. Regarding *ex ante* impact assessment, the tools to enable researchers to assess and argue on their credible expected impacts remain to be developed.

Implementation

The whole impact assessment and monitoring system of EMBRAPA is co-ordinated and methodologically supported by an impact assessment team located at headquarters, under the guidance of the Secretariat of Management and Institutional Development. In each of the 40 regional centre, appointed dedicated socio-environmental researchers who are specifically trained for impact assessment, and are supported by the headquarter team, are in charge of prospective and annual monitoring of three technologies. Case studies are carried out by the centres' “impact team” along with the principal investigator in charge of each technology assessed. A budgetary allowance is made available through the Secretariat for Management and Institutional Development (SGI) for this task.

For each technology, *ex post* economic impact data are collected through national services, rural extension contacts and Embrapa or private surveys. *Ex post* non-economic impacts data are collected and processed by centres by implementing the Ambitec method. Ambitec has been designed as a user-friendly device, with an integrated software platform. It incorporates formalised and standardised guidelines to assist the implementation of Ambitec, where three steps are defined: perimeter definition, field survey including indicator's scaling checklists, and reporting in template. Ambitec is a practical, expeditious, low cost, and reproducible socio-environmental impact assessment procedure relevant for the wide range of agricultural technologies and rural activities concerned in Embrapa's research programme. Each technology leads to an annual case study report that is prepared by the regional centres, and details the measurement of impacts by the scientific board of each centre. This annual assessment is part of the performance measurement system of the centres, implemented since 2001.

At the level of EMBRAPA, the *ex post* economic impact is calculated by adding all the economic surplus of the individual technologies monitored. The productivity of research is estimated by dividing this total surplus by the annual budget of EMBRAPA, thus producing a single figure: "each BRL invested generated BRL 8.53 to Brazilian society" in 2014. At the level of EMBRAPA, the environmental and social impact is aggregated by calculating average indexes of impact by type of technology (plant varieties, animal production, software, processing technologies). That method does not account for different adoption rates among the portfolio of technologies.

In 2016, EMBRAPA will implement a global management plan related to *ex ante* impact assessment. According to the draft plan, employees will be linked to projects and actions aligned to the five impact targets defined in the strategic plan. *Ex ante* impact will be assessed by each project leader before each funding request is submitted. The Secretariat of intelligence and macro strategy of EMBRAPA will then select the projects to be funded, notably accounting for its expected impact. The coming implementation of the strategic plan of EMBRAPA 2014-34 is a challenge for the headquarter impact assessment team. The team is to reorganise the existing institutional process to evaluate the *ex post* impacts of the EMBRAPA technologies to align with the five impact axes newly established in the 2014-34 strategic plan and to support the R&D selection process of new projects with *ex ante* impact assessment.

Utilisation of evaluation results

So far, the evaluation's results are used for advocacy and accountability. *Ex post* impact assessment is annually reported in a standardised Social Balance Report which is largely diffused to stakeholders and funders. A website[36] is dedicated to the Social Balance of EMBRAPA, and diffuses the databases of social actions of EMBRAPA.

Ex ante impact assessment of technologies is expected to notably influence funding of new research projects (with a highly selective rate, selection being done by external experts). *Ex ante* impact claims could also be considered for the annual negotiations of salaries and benefit sharing.

Key references

In addition to an interview conducted with a senior manager of EMBRAPA's headquarters, some key bibliographic documents were studied:

Avila, A.F., G.S. Rodrigues, G.L. Vedovoto, R. de Camargo Penteado Filho, and W. Corrêa da Fonseca Junior (2015), EMBRAPA's experience on the impact assessment of agricultural R&D: 15 years using a multidimensional approach, presented at the ImpAR Conference, INRA, Paris, France. www.alice.cnptia.embrapa.br/alice/bitstream/doc/1036444/1/EmbrapaExperienceontheimpact.pdf.

EMBRAPA (2015), *VI Plano Director da Embrapa 2014-2034*, EMBRAPA, MBRAPA, Secretaria de Gestão e Desenvolvimento Institucional, Brasilia. www.infoteca.cnptia.embrapa.br/handle/doc/1025506.

Rodrigues, G.S., C.C. de Almeida Buschinelli, and A.F. Dias Avila (2010), "An environmental impact assessment system for agricultural research and development ii: institutional learning experience at Embrapa", *Journal of technology management & innovation* Vol. 5, pp. 38–56. www.jotmi.org/index.php/GT/article/view/art173.

36. http://bs.sede.embrapa.br/.

CGIAR

The *Consultative* Group on International Agricultural Research (CGIAR) was created in 1971. It has a network of 15 independent research centres, with a total of 8 000 staff (researchers and technicians). In 2010, the CGIAR moved away from centre-based programming, to cross-centres programme-based research implementation. The aim of the reform was to address stagnating funding, and integrate core competencies and appropriate partnerships across CGIAR centres.

The CGIAR Consortium develops and carries out research programmes to address complex development issues related to agriculture. Since 2010 a Strategy and Results Framework (SRF) provides common goals, to be jointly achieved by CGIAR centres through 16 CGIAR Research Programmes (CRPs). Donors fund strategic research programmes contributing to the SRF by contributing either to the CGIAR Fund, or through bilateral funding (for a large share). The CGIAR Consortium co-ordinates the activities across research centres and is accountable for how the funds are used. Before receiving funding, CRPs set out their expected achievements and provide verifiable targets against which progress can be measured and monitored. The 2010–2015 SRF identified four over-arching objectives. The second CGIAR's 2016–2030 SRF (CGIAR, 2015), approved by CGIAR's Consortium Board in May 2015, will contribute to the achievement of the Sustainable Development Goals (SDGs) outlined by the United Nations. At the system level CGIAR has three goals or System Level Outcomes (SLOs):

- Reduce poverty: help 100 million people, of which 50% are women, get out of poverty.
- Improve food and nutrition security for health: ensure that 150 million people, of which 50% are women, meet minimum dietary energy requirements.
- Improve natural resource systems and ecosystem services: restore 190 million hectares of degraded land by 2030.

History of impact assessment

There is a long history of evaluation in the CGIAR, with independent evaluations (External Program and Management Reviews) carried out since the late 1970s. In parallel, independent system-wide reviews of CGIAR were undertaken approximately every ten years. Attention to impact started in 1990. CGIAR usually assessed its *ex post* impact through calculating an internal rate of return, based on the adoption rate of improved varieties. Recently, many impact assessments were carried out at donors' request by CGIAR centres or university researchers, which resulted in studies with very variable focus, time-frame, methods and quality.

The independent Standard Panel for Impact Assessment (SPIA), a sub-group of the CGIAR Independent Science and Partnership Council (ISPC), was created in 1995 to advise donors' fund allocation by performing *ex post* meta-evaluation of the economic impact of the CGIAR's research (Renkow and Byerlee, 2010). Historically, SPIA performed global modelling, using IFPRI models, to estimate Internal Rates of Return and addressed impact evaluation at a centre's level, or on broader thematic areas (de Janvry et al 2011). Over the years, the CGIAR's research agenda expanded (Kelley et al 2008; Raitzer and Kelley, 2008), notably to address natural resource management and conservation issues (CGIAR, 2012). Recent work on methodological developments was carried out by SPIA to reflect CGIAR's expanded agenda.

With the creation of the CGIAR Strategy and Results Framework in 2010, and the funding of a portfolio of research programmes cutting across the centres, new modalities for evaluation were developed. The CGIAR Policy for Independent External Evaluation approved in 2012 sets out the mandate, scope and proposed implementation arrangements for evaluation in the reformed CGIAR.

Purpose of impact evaluation and evaluation design

Impact assessment and *ex post* assessment studies chiefly serve accountability motives (demonstrating impact and value for money to CGIAR donors and stakeholders). They also support CGIAR organisational learning, and develop internal impact-oriented culture and capacity.

The current structure for evaluations in CGIAR lays out a system of multi-level and cascading evaluations at different levels: CGIAR Research Programme (CRP), themes (portfolio of CRPs), institutions (centres), and the CGIAR as a whole.

The independent Evaluation Assessment (IEA), created in 2012, reports to the CGIAR Fund Council and has a central mandate for external evaluation of all parts of the reformed CGIAR system. It assesses, among other things, a CRP's *ex ante* and *ex post* impact assessment strategy and implementation, as well as all ex *post* impact assessment studies and publications related to a particular CRP (IEA, 2015a; IEA, 2015b; IEA evaluation reports of CRPs and guiding documents). IEA commissioned evaluations are conducted by independent and external teams. They have a large scope, including science quality and potential for future development impact. The evaluation agenda is set by a four-year Rolling Evaluation Work Plan (REWP). The creation of the IEA is closely linked with the donor's need to monitor their funding efficiency and effectiveness towards achieving CGIAR's three SLOs through the development of result-based management (RBM). The IEA also has a role in promoting good practices and capacity building in evaluation.

The IEA was also created because, as the ISPC provided scientific and programmatic guidance, the SPIA, which is a sub-group of the ISPC, could not evaluate performance arising from this advice. In the new structure, SPIA's role in the development of *ex post* assessment methodologies for broader societal impacts has however been maintained and sustained.

In terms of guidelines for *ex post* evaluation of broader impact, SPIA *ex post* assessment relies on case studies that use a diversity of analytical frameworks since these are proposed by the external researchers who respond to thematic calls for proposals (SPIA, 2015). Most SPIA case studies, however, rely on quantitative econometric analysis, based on cost-benefit analysis and the design of counterfactuals. For environmental impact assessment, the SPIA recommended the extension of the standard cost-benefit analysis to include revealed and stated preferences to capture non-market effects. In the past, SPIA's mission included the improvement of assessment methodologies and tackling transversal questions across the CGIAR research portfolio. This way, SPIA performed state of the art analysis regarding impact assessment methodologies, designed guidelines, and implemented them on a set of pilot cases. SPIA did this work for policy-oriented impacts in 2006 and for environmental impact in 2008. In 2013, SPIA received a grant called SIAC (Strengthening Impact Assessment in the CGIAR) for methodological developments. They include efforts to have more accurate estimates of adoption rates, or cross-country impact assessment. Methodologies to assess the full range of CGIAR societal impact were also targeted, like the effects on poverty and health, natural resources management, or under-researched areas, like the impact of livestock research or social science. An open call for proposals was launched in 2014 to gather case studies on these topics, with researchers offering their own methodologies (mostly quantitative ones). It resulted in 30 case-studies being selected (and currently investigated).

In terms of programme monitoring, the new SRF will be implemented over several phases, starting with the so-called Phase II (2017–2022) portfolios of CRPs. The CGIAR Fund Council asked the CGIAR Consortium to establish a clear link between the research carried out by the CRPs and the three CGIAR System Level Outcomes (SLOs). SLOs are broken down into 10 Intermediate Development Outcomes (IDOs) and 30 sub-IDOs, for which quantitative targets have been set. A result-based management policy is under development. Its implementation, planned for 2017 will be driven by the CRPs, the ISPC and the System Office.

Each CRP must derive its own targeted quantitative contribution to CGIAR global targets. CRPs define their intended impact by developing impact pathways (outputs-to-outcomes-to-impacts) and related theories-of-change (explicit assumptions about how to get from research to development results), which combine qualitative and quantitative components. CRPs are expected to design their activities and monitor their contribution to the sub-IDO level which is linked to the IDOS through programme theory and then up to the SLOs. Methodologies are currently developed by each CRP, with the support of an inter-CRP Monitoring, Evaluation and Learning Community of Practice. Sharing of these bottom-up initiatives has not been organised yet at the level of the CGIAR.

There are some issues with establishing targets. Time lag is clearly an issue as the CGIAR SLO's timeline for targets is too near (2022), and will be achieved through contributing research performed 20-30 years ago, before the CRPs existed. CRP recent external evaluations of Phase I funding have also found that in many cases targets set to CRPs were unrealistic. CRPs' evaluations have emphasised the need to use the theories of change as dynamic concepts where the assumptions may become hypothesis to be tested through research. Given the risky and protracted nature of research, result-based management that involves adaptive management on the basis of lessons would have to focus on progress and results relatively closer to research.

Researchers operating CRPs expressed their interest for proxies that can link CRPs' research outputs and outcomes with CGIAR SLO targets and system-wide indicators tracking impacts.

Implementation

Data collection is an issue for the implementation of *ex post* impact assessment at different levels of the CGIAR system, as well as the implementation of monitoring approaches to CRPs' progress towards impacts.

One issue regarding *ex post* impact assessment is that data sets for regression are not readily available. Either researchers bring or collect their own measures and data, or they work on existing publically available data, which can be inconsistent. Not all measures are monetary (notably for health and environment impacts). For nutrition for instance, indicators deal with diverse scores of food supply and blood test measures. SPIA gives methodological support for specific studies. For example, some recommendations were made to improve the robustness of data in plant genetic research: research showed that photos or DNA fingerprinting gave far better results for identifying improved varieties, and were far more robust, when compared to traditional methods for collecting data (surveys, interviews, field observations). A similar recommendation is tested for livestock genetics, but it is a costly investment. In terms of aggregation of *ex post* assessment, some studies have calculated an Internal Rate of Return for CGIAR research. Renkow and Byerlee (2010) for example have reported an aggregated *ex post* benefit cost ratio of all CGIAR research investments of 17.26%. These estimates are however based heavily on assumptions regarding adoption rates and come with very large margins. One respondent to this study considers that robust SPIA impact assessment designs could also be discredited in some instances by inflated figures from centres' self-assessment, or inconsistent data. SPIA suggested that it "could provide a label of impact assessment quality" on a voluntary basis, based on a peer review system of impact studies carried out by the centres.

As far as the monitoring of CRPs is concerned, in January 2015 the IEA produced a series of six guidance documents concerning the external evaluation of CRPs. Funding for monitoring and evaluation is available through the "programme management" budget of CRPs. CRPs are expected to commission evaluation of components of the CRP as an input to the IEA's evaluations. The problem of data availability is complicated by issues related to the levels of assessment: most of the data are currently generated at the centre level, but the assessment is to be performed at the CRP level. Each centre still develops its own evaluation method for establishing the baseline of its accomplishments or the counterfactual for the evaluation of a CRP's results in a particular geography. For example, IRRI calculates baseline values by conducting household surveys, which provide rich data sets, but prove to be quite expensive, and may not be applicable to all CGIAR centres. Nevertheless, no methodology and monitoring system has so far been shared among CRPs to decide on results' attribution versus contribution, outcome calculation methods, etc. In preparation for the second funding phase (2017-2022), CRPs propose their own method and a critical need for comparability and/or harmonisation has been identified.

Regarding the implementation of the monitoring initiative, CGIAR plans to enforce an "annual reporting or programme progress with financial reporting, and performance assessment". As a principle of the call for Phase II (2017-2022), all CRPs and their lead centres are to follow a harmonised and homogeneous monitoring and reporting framework. Standardisation of minimum requirements, consistency and alignment of reporting are key to this framework and demand interoperability of platforms. This push for harmonisation in the short term represents an enormous workload for CRPs' managers, which calls for capacity building and budget. A crucial issue is the continuous monitoring of development outcomes, impacts and targets achievement at CRP and System levels, notably after the CRPs have ended. In particular, CRPs' managers ask for agreed upon, standard proxies for development outcomes and impact, to limit investments in monitoring.

Utilisation of evaluation results

SPIA impact studies are used for communication purposes. Impact briefs[37] about case studies are regularly posted on the CGIAR website. They result from SPIA *ex post* evaluations, other external *ex post* impact assessments commissioned by donors (e.g. ACIAR's report on wheat improvement in Afghanistan) or by the CRPs themselves. There is an issue with too few case studies produced in contrast to CGIAR's widening agenda. For example SPIA will only produce 25-30 case-studies through the SIAC grant, which is insufficient to provide useful lessons for donors at the level of CGIAR. Some centres and programmes also communicate figures to impress donors, which could affect CGIAR's reputation.

In terms of evaluation methodology, ISPC published a white paper[38] with recommendations regarding planning, domains for research and target groups, trade-offs and theory of change for CRPs' design and monitoring. Emphasis was placed on having feasible and realistic intermediate outcome objectives where

37. http://impact.cgiar.org/impact-briefs.

38. www.sciencecouncil.cgiar.org/sites/default/files/ISPC_WhitePaper_Prioritization.pdf.

agricultural research can contribute. It will be important to see if these recommendations influence the monitoring and evaluation system.

In terms of results, between 2013 and 2015, the IEA supported the evaluation of all 15 CRPs. It also carried out a few cross-cutting thematic evaluations such as Capacity Strengthening, Gender and Partnerships.

CRPs' evaluation results are used for communication purposes (the evaluation documents are published on the CGIAR website[39]) but also for internal learning. With the completion of Phase I evaluation, the IEA is planning to conduct a synthesis to build on the evaluative evidence from the 15 CRP evaluation reports.

Evaluations are accompanied by a Management Response and an action plan. Thus CRPs are expected to implement the recommendations in an agreed timeline. Some CRP evaluation reports may contribute to decisions regarding fund allocation, or structural changes in CRPs (Bennett, 2009).

Key references

In addition to an interview conducted with members of SPIA, IEA and a CRP, some key bibliographic documents were studied:

Bennett, J. (2009), Advancing *ex-post* impact assessment of environmental and social impacts of CGIAR research, paper submitted to the Standing Panel on Impact Assessment of the CGIAR Science Council. http://impact.cgiar.org/sites/default/files/docs/Bennett-2009.pdf.

CGIAR (2015), *CGIAR Strategy and Results Framework 2016-2030.* www.cgiar.org/our-strategy/.

CGIAR Independent Science and Partnership Council (2012), A Stripe Review of Natural Resources Management Research in the CGIAR, Independent Science and Partnership Council Secretariat, Rome.

de Janvry, A., A. Dunstan, and E. Sadoulet (2011), Recent Advances in Impact Analysis Methods for Ex-post Impact Assessments of Agricultural Technology: Options for the CGIAR. Report prepared for the workshop: Increasing the rigor of ex-post impact assessment of agricultural research: A discussion on estimating treatment effects, organized by the CGIAR Standing Panel on Impact Assessment (SPIA), 2 October 2010, Berkeley, California, USA. Independent Science and Partnership Council Secretariat: Rome, Italy. http://impact.cgiar.org/sites/default/files/docs/deJanvryetal2011.pdf.

IEA (2015a), Guidance for Managing the Independent External Evaluation of CGIAR Research Programmes (CRPs) (IEA Guidance Note G1), IEA CGIAR. http://iea.cgiar.org/sites/default/files/G1.pdf.

IEA (2015b), CGIAR standards for Independent external Evaluation, IEA CGIAR. http://iea.cgiar.org/publication/cgiar-standards-independent-external-evaluations.

IEA evaluation reports of CRPs available on http://iea.cgiar.org/evaluations.

IEA guiding documents for external evaluation available on http://iea.cgiar.org/publications.

Kelley, T., J. Ryan, and H. Gregersen (2008), "Enhancing *ex-post* impact assessment of agricultural research: the CGIAR experience", *Research Evaluation* Vol. 17, p. 201. DOI: http://dx.doi.org/10.3152/095820208X331711.

Raitzer, D.A., and T.G. Kelley (2008), "Assessing the contribution of impact assessment to donor decisions for international agricultural research", *Research Evaluation* Vol. 17, pp. 187–199. DOI: http://dx.doi.org/ DOI: http://dx.doi.org/10.3152/095820208X331702.

Renkow, M., and D. Byerlee (2010), "The impacts of CGIAR research: A review of recent evidence", *Food Policy* Vol. 35, pp. 391–402. DOI: http://dx.doi.org/10.1016/j.foodpol.2010.04.006.

SPIA (2015), SIAC Program Progress Report, SPIA, CGIAR, Rome. http://impact.cgiar.org/sites/default/files/docs/SIACProgramProgressReport_20Feb2015.pdf.

39. http://iea.cgiar.org/evaluations.